

SI EDITION

SECOND EDITION



MECHATRONICS SYSTEM DESIGN

DEVDAS SHETTY
RICHARD A. KOLK

CONVERSIONS BETWEEN U.S. CUSTOMARY UNITS AND SI UNITS

U.S. Customary unit	Times conversion factor		Equals SI unit		
	Accurate	Practical			
Acceleration (linear) foot per second squared inch per second squared	ft/s ² in./s ²	0.3048* 0.0254*	0.305 0.0254 meter per second squared meter per second squared	m/s ² m/s ²	
Area circular mil square foot square inch	cmil ft ² in. ²	0.0005067 0.09290304* 645.16*	0.0005 0.0929 645	square millimeter square meter square millimeter	mm ² m ² mm ²
Density (mass) slug per cubic foot	slug/ft ³	515.379	515	kilogram per cubic meter	kg/m ³
Density (weight) pound per cubic foot pound per cubic inch	lb/ft ³ lb/in. ³	157.087 271.447	157 271	newton per cubic meter kilonewton per cubic meter	N/m ³ kN/m ³
Energy; work foot-pound inch-pound kilowatt-hour British thermal unit	ft-lb in.-lb kWh Btu	1.35582 0.112985 3.6* 1055.06	1.36 0.113 3.6 1055	joule (N·m) joule megajoule joule	J J MJ J
Force pound kip (1000 pounds)	lb k	4.44822 4.44822	4.45 4.45	newton (kg·m/s ²) kilonewton	N kN
Force per unit length pound per foot pound per inch kip per foot kip per inch	lb/ft lb/in. k/ft k/in.	14.5939 175.127 14.5939 175.127	14.6 175 14.6 175	newton per meter newton per meter kilonewton per meter kilonewton per meter	N/m N/m kN/m kN/m
Length foot inch mile	ft in. mi	0.3048* 25.4* 1.609344*	0.305 25.4 1.61	meter millimeter kilometer	m mm km
Mass slug	lb-s ² /ft	14.5939	14.6	kilogram	kg
Moment of a force; torque pound-foot pound-inch kip-foot kip-inch	lb-ft lb-in. k-ft k-in.	1.35582 0.112985 1.35582 0.112985	1.36 0.113 1.36 0.113	newton meter newton meter kilonewton meter kilonewton meter	N·m N·m kN·m kN·m

SELECTED PHYSICAL PROPERTIES

Property	SI	USCS
Water (fresh) weight density mass density	9.81 kN/m ³ 1000 kg/m ³	62.4 lb/ft ³ 1.94 slugs/ft ³
Sea water weight density mass density	10.0 kN/m ³ 1020 kg/m ³	63.8 lb/ft ³ 1.98 slugs/ft ³
Aluminum (structural alloys) weight density mass density	28 kN/m ³ 2800 kg/m ³	175 lb/ft ³ 5.4 slugs/ft ³
Steel weight density mass density	77.0 kN/m ³ 7850 kg/m ³	490 lb/ft ³ 15.2 slugs/ft ³
Reinforced concrete weight density mass density	24 kN/m ³ 2400 kg/m ³	150 lb/ft ³ 4.7 slugs/ft ³
Atmospheric pressure (sea level) Recommended value Standard international value	101 kPa 101.325 kPa	14.7 psi 14.6959 psi
Acceleration of gravity (sea level, approx. 45° latitude) Recommended value Standard international value	9.81 m/s ² 9.80665 m/s ²	32.2 ft/s ² 32.1740 ft/s ²

SI PREFIXES

Prefix	Symbol	Multiplication factor	
tera	T	10^{12}	= 1 000 000 000 000
giga	G	10^9	= 1 000 000 000
mega	M	10^6	= 1 000 000
kilo	k	10^3	= 1 000
hecto	h	10^2	= 100
deka	da	10^1	= 10
deci	d	10^{-1}	= 0.1
centi	c	10^{-2}	= 0.01
milli	m	10^{-3}	= 0.001
micro	μ	10^{-6}	= 0.000 001
nano	n	10^{-9}	= 0.000 000 001
pico	p	10^{-12}	= 0.000 000 000 001

Note: The use of the prefixes hecto, deka, deci, and centi is not recommended in SI.

CONVERSIONS BETWEEN U.S. CUSTOMARY UNITS AND SI UNITS

U.S. Customary unit	Times conversion factor		Equals SI unit
	Accurate	Practical	
Moment of inertia (area) inch to fourth power	in. ⁴	416,231 0.416231×10^{-6}	millimeter to fourth power meter to fourth power
inch to fourth power	in. ⁴	$416,000$ 0.416×10^{-6}	mm^4 m^4
Moment of inertia (mass) slug foot squared	slug·ft ²	1.35582	kilogram meter squared
Power		1.36	kg·m ²
foot-pound per second	ft-lb/s	1.35582	watt (J/s or N·m/s)
foot-pound per minute	ft-lb/min	0.0225970	watt
horsepower (550 ft-lb/s)	hp	745.701	watt
Power		1.36 0.0226 746	W W W
Pressure; stress			
pound per square foot	psf	47.8803	pascal (N/m^2)
pound per square inch	psi	6894.76	Pa
kip per square foot	ksf	47.8803	pascal
kip per square inch	ksi	6.89476	kilopascal
			MPa
Pressure; stress			
Section modulus			
inch to third power	in. ³	16,387.1	millimeter to third power
inch to third power	in. ³	16.3871×10^{-6}	meter to third power
Velocity (linear)			
foot per second	ft/s	0.3048*	meter per second
inch per second	in./s	0.0254*	m/s
mile per hour	mph	0.44704*	meter per second
mile per hour	mph	1.609344*	m/s
Velocity (linear)			
Volume			
cubic foot	ft ³	0.0283168	kilometer per hour
cubic inch	in. ³	16.3871×10^{-6}	m/s
cubic inch	in. ³	16.3871	m/s
gallon (231 in. ³)	gal.	3.78541	m/s
gallon (231 in. ³)	gal.	0.00378541	km/h
Volume			

*An asterisk denotes an *exact* conversion factor

Note: To convert from SI units to USCS units, *divide* by the conversion factor

Temperature Conversion Formulas

$$T(\text{°C}) = \frac{5}{9}[T(\text{°F}) - 32] = T(\text{K}) - 273.15$$

$$T(\text{K}) = \frac{5}{9}[T(\text{°F}) - 32] + 273.15 = T(\text{°C}) + 273.15$$

$$T(\text{°F}) = \frac{9}{5}T(\text{°C}) + 32 = \frac{9}{5}T(\text{K}) - 459.67$$

MECHATRONICS SYSTEM DESIGN

SECOND EDITION, SI

Devdas Shetty, Ph.D., P.E.

Dean of Research and Professor of Mechanical Engineering
University of Hartford
West Hartford, Connecticut

Richard A. Kolk

Sr. Vice President—Technology
PaceControls
Philadelphia, Pennsylvania



Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

This is an electronic version of the print textbook. Due to electronic rights restrictions, some third party content may be suppressed. Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. The publisher reserves the right to remove content from this title at any time if subsequent rights restrictions require it. For valuable information on pricing, previous editions, changes to current editions, and alternate formats, please visit www.cengage.com/highered to search by ISBN#, author, title, or keyword for materials in your areas of interest.



**Mechatronics System Design,
Second Edition, SI
Devdas Shetty and Richard A. Kolk**

Publisher, Global Engineering:
Christopher M. Shortt
Senior Acquisitions Editor: Swati Merehishi
Senior Developmental Editor: Hilda Gowans
Editorial Assistant: Tanya Altieri
Team Assistant: Carly Rizzo
Marketing Manager: Lauren Betzos
Media Editor: Chris Valentine
Senior Content Project Manager:
Colleen Farmer
Production Service: RPK Editorial Services
Copyeditor: Shelly Gerger-Knecht
Proofreaders: Erin Wagner/Martha
McMaster
Indexer: Shelly Gerger-Knecht
Compositor: Integra Software Services
Senior Art Director: Michelle Kunkler
Cover Designer: Andrew Adams
Cover Images: © Yanir Tafov/Shutterstock
Permissions Account Manager: Mardell
Glisnksi Schultz
Text and Image Permissions Researcher:
Kristiina Paul
First Print Buyer: Arethea Thomas

© 2011, 1997 Cengage Learning

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher.

For product information and technology assistance,
contact us at **Cengage Learning Customer &
Sales Support, 1-800-354-9706.**

For permission to use material from this text or product,
submit all requests online at www.cengage.com/permissions.

Further permissions questions can be emailed to
permissionrequest@cengage.com

Library of Congress Control Number: 2010932699

International Student Edition
ISBN-13: 978-1-4390-6199-2
ISBN-10: 1-4390-6199-8

Cengage Learning
200 First Stamford Place, Suite 400
Stamford, CT 06902
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at: international.cengage.com/region.

Cengage Learning products are represented in Canada by Nelson Education Ltd.

For your course and learning solutions, visit
www.cengage.com/engineering.

Purchase any of our products at your local college store or at our preferred online store www.Cengagebrain.com.

LabVIEW is a registered trademark of National Instruments Corporation, 11500 N. Mopac Expressway, Austin TX.

MATLAB is a registered trademark of The MathWorks, 3 Apple Hill Road, Natick, MA.

VisSim is a trademark of Visual Solutions, Incorporated, 487 Groton Road, Westford, MA.

Printed in the United States of America

1 2 3 4 5 6 7 14 13 12 11 10

To my wife, Sandya, and sons, Jagat and Nandan, for their
love and support.

Devdas Shetty

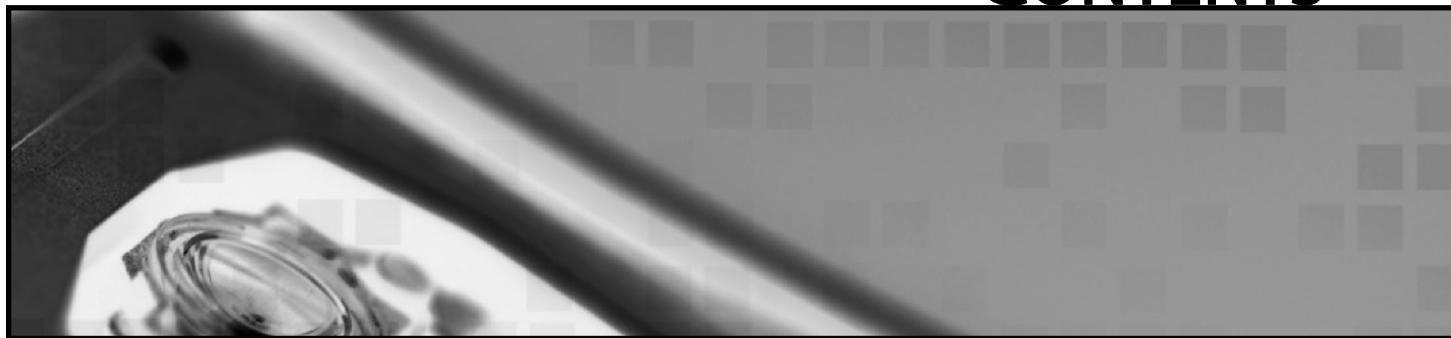
To my wife, Cathie; daughters, Emily and Elizabeth;
and E. Gloria MacKintosh for her encouragement

Ric Kolk

This page was intentionally left blank

This page was intentionally left blank

CONTENTS



1 MECHATRONICS SYSTEM DESIGN 1

- 1.1 What is Mechatronics 1
- 1.2 Integrated Design Issues in Mechatronics 4
- 1.3 The Mechatronics Design Process 6
- 1.4 Mechatronics Key Elements 10
- 1.5 Applications in Mechatronics 18
- 1.6 Summary 39
- References 39
- Problems 40

2 MODELING AND SIMULATION OF PHYSICAL SYSTEMS 41

- 2.1 Operator Notation and Transfer Functions 42
- 2.2 Block Diagrams, Manipulations, and Simulation 43
- 2.3 Block Diagram Modeling—Direct Method 51
- 2.4 Block Diagram Modeling—Analogy Approach 64
- 2.5 Electrical Systems 75
- 2.6 Mechanical Translational Systems 82
- 2.7 Mechanical Rotational Systems 90
- 2.8 Electrical–Mechanical Coupling 95
- 2.9 Fluid Systems 102
- 2.10 Summary 116
- References 117
- Problems 118
- Appendix to Chapter 2 123

3 *SENSORS AND TRANSDUCERS 131*

- 3.1 Introduction to Sensors and Transducers 132
- 3.2 Sensitivity Analysis—Influence of Component Variation 139
- 3.3 Sensors for Motion and Position Measurement 144
- 3.4 Digital Sensors for Motion Measurement 162
- 3.5 Force, Torque, and Tactile Sensors 168
- 3.6 Vibration—Acceleration Sensors 183
- 3.7 Sensors for Flow Measurement 195
- 3.8 Temperature Sensing Devices 210
- 3.9 Sensor Applications 216
- 3.10 Summary 246
- References 246
- Problems 247

4 *ACTUATING DEVICES 255*

- 4.1 Direct Current Motors 255
- 4.2 Permanent Magnet Stepper Motor 262
- 4.3 Fluid Power Actuation 269
- 4.4 Fluid Power Design Elements 274
- 4.5 Piezoelectric Actuators 287
- 4.6 Summary 289
- References 289
- Problems 289

5 *SYSTEM CONTROL—LOGIC METHODS 291*

- 5.1 Number Systems in Mechatronics 291
- 5.2 Binary Logic 297
- 5.3 Karnaugh Map Minimization 302
- 5.4 Programmable Logic Controllers 309
- 5.5 Summary 321
- References 321
- Problems 322

6 *SIGNALS, SYSTEMS, AND CONTROLS 329*

- 6.1 Introduction to Signals, Systems, and Controls 329
- 6.2 Laplace Transform Solution of Ordinary Differential Equations 332
- 6.3 System Representation 338
- 6.4 Linearization of Nonlinear Systems 343
- 6.5 Time Delays 346

6.6	Measures of System Performance	349
6.7	Root Locus	357
6.8	Bode Plots	370
6.9	Controller Design Using Pole Placement Method	378
6.10	Summary	383
	References	383
	Problems	383

7 SIGNAL CONDITIONING AND REAL TIME INTERFACING 387

7.1	Introduction	387
7.2	Elements of a Data Acquisition and Control System	388
7.3	Transducers and Signal Conditioning	392
7.4	Devices for Data Conversion	394
7.5	Data Conversion Process	402
7.6	Application Software	409
7.7	Summary	445
	References	445

8 CASE STUDIES 446

8.1	Comprehensive Case Studies	446
8.2	Data Acquisition Case Studies	466
8.3	Data Acquisition and Control Case Studies	476
8.4	Summary	489
	References	489
	Problems	490

APPENDIX 1 DATA ACQUISITION CARDS 491

INDEX 493

PREFACE TO THE SI EDITION

This edition of *Mechatronics System Design*, has been adapted to incorporate the International System of Units (*Le Système International d'Unités* or SI) throughout the book.

Le Système Internationales

The United States Customary System (USCS) of units uses FPS (foot-pound-second) units (also called English or Imperial units). SI units are primarily the units of the MKS (meter-kilogram-second) system. However, CGS (centimeter-gram-second) units are often accepted as SI units, especially in textbooks.

Using SI Units in this Book

In this book, we have used both MKS and CGS units. USCS units or FPS units used in the US Edition of the book have been converted to SI units throughout the text and problems. However, in case of data sourced from handbooks, government standards, and product manuals, it is not only extremely difficult to convert all values to SI, it also encroaches upon the intellectual property of the source. Some data in figures, tables, and references, therefore, remains in FPS units. For readers unfamiliar with the relationship between the FPS and the SI systems, a conversion table has been provided inside the front cover.

To solve problems that require the use of sourced data, the sourced values can be converted from FPS units to SI units just before they are to be used in a calculation. To obtain standardized quantities and manufacturers' data in SI units, the readers may contact the appropriate government agencies or authorities in their countries/regions.

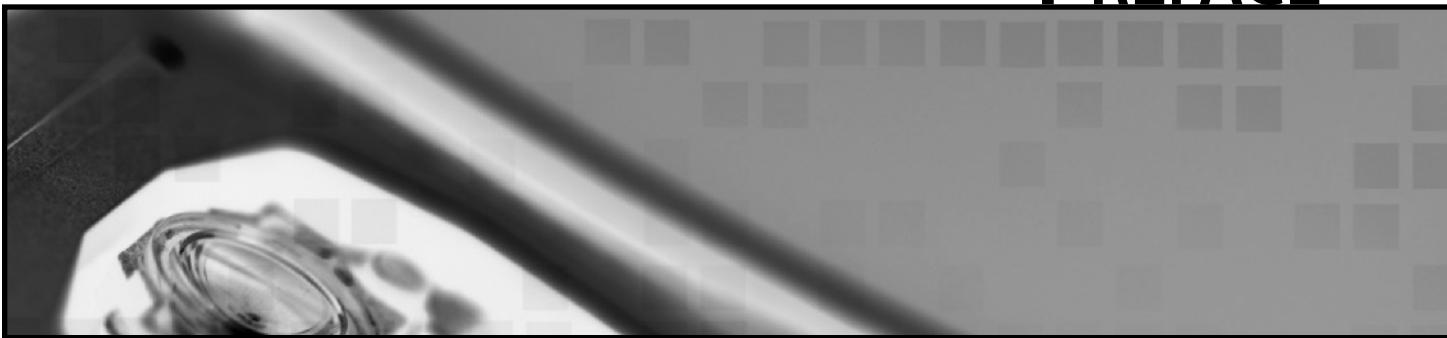
Instructor Resources

The Instructors' Solution Manual in SI units is available through your Sales Representative or online through the book website at www.cengage.com/engineering.

The readers' feedback on this SI Edition will be highly appreciated and will go a long way in helping us improve subsequent editions.

The Publishers

PREFACE



Competing in a globalized market requires the adaptation of modern technology to yield flexible, multifunctional products that are better, cheaper, and more intelligent than those currently on the shelf. The importance of mechatronics is evidenced by the myriad of smart products that we take for granted in our daily lives, from the cruise control feature in our cars to advanced flight control systems and from washing machines to multifunctional precision machines. The technological advances in digital engineering, simulation and modeling, electromechanical motion devices, power electronics, computers and informatics, MEMS, microprocessors, and DSPs have brought new challenges to industry and academia.

Mechatronics is the synergistic combination of mechanical and electrical engineering, computer science, and information technology, which includes the use of control systems as well as numerical methods to design products with built-in intelligence.

The field of mechatronics allows the engineer to integrate mechanical, electronics, control engineering and computer science into a product design process. Modeling, simulation, analysis, virtual prototyping and visualization are critical aspects of developing advanced mechatronics products. Mechatronics design focuses on systematic optimization to ensure that quality products are created in a timely fashion. Getting electromechanical design right the first time requires teamwork and coordination across multiple segments and disciplines of the engineering process. The integration is facilitated by the introduction of new software simulation tools that work in tandem with systems to create an efficient mechatronics pathway.

The first edition of this book was designed for the upper-level undergraduate or graduate student in mechanical, electrical, industrial, biomedical, computer, and of course, mechatronics engineering. The book was widely used in the United States and also in Canada, China, Europe, India, and South Korea. Following feedback from experts in this field and also from the faculty who used this text book, the second edition has been considerably extended and augmented with extra depth so that not only is it still relevant for its original users, but is also apt for other emerging programs.

Currently, there exists a trend to include mechatronics in the traditional curricula with the purpose of providing integrated design experience to graduating engineers. This experience is created by using measurement principles, sensors, actuators, electronics circuits, and real-time interfacing coupled with design, simulation, and modeling. Some of these courses end with case studies and a

unifying design project that integrates various disciplines into a successful design product that can be quickly assembled and analyzed in a laboratory environment.

This second edition has been updated throughout. The aim is to provide a comprehensive coverage of many areas so that the readers understand the range of engineering disciplines that come together to form the field of mechatronics. The interdisciplinary approach taken in this book provides the technical background needed in the design of mechatronics products.

The second edition is designed to serve as a text for the following:

- Stand-alone mechatronics courses.
- Modern instrumentation and measurement courses.
- Hybrid electrical and mechanical engineering course covering sensors, actuators, data-acquisition, and control.
- Interdisciplinary engineering courses dealing with modeling, simulation, and control.

Key Features

- Extensive coverage of sensors, actuators, system modeling, and classical control system design coupled with real-time computer interfacing.
- Industrial case studies.
- In-depth discussions on modeling and simulation of physical systems.
- Inclusion of block diagrams, modified analogy approach to modeling, and the use of state-of-the-art visual simulation software.
- Shows how interactive modeling created in a graphical environment with visual representation is crucial to the design process.
- Step-by-step mechatronics system design methodology.
- Illustration of how the design process can be done right the first time.

New to This Edition

- Numerous design examples and end-of-chapter problems added to help students understand the basic mechatronics methodology.
- A simple motion control example carried out throughout the eight chapters covering the different elements of mechatronics systems progressively.
- Simulation and real-time interfacing using LabVIEW® included in *addition* to VisSim™.
- Inclusion of current trends in mechatronics and smart manufacturing.
- Illustration of block diagram approach and emphasis on the comprehensive use of mathematical analysis, simulation and modeling, control and real-time interfacing in implementing case studies.
- Expanded coverage of sensors, real-time interfacing, and multiple input and multiple output systems.
- Design examples and problems drawn from situations encountered in everyday life.

- Illustration of synergistic aspects of mechatronics and its influence in design.
- Hardware-in-the-loop examples and illustration of optimum design.
- Control system analysis for multiple input and multiple output situations.
- Complete illustration of permanent magnet DC motor integrated with hall effect sensor, its mathematical analysis, and position control.
- Creation of virtual prototype of mechatronics systems.

Chapter 1 provides an in-depth discussion of the key issues in the mechatronics design process and examines emerging trends. In addition, this chapter addresses recent advances of mechatronics in smart manufacturing and discusses the improvements to conventional designs by using a mechatronics approach.

Chapter 2 is devoted entirely to system modeling and simulation. Students will learn to create accurate computer-based dynamic models from illustrations and other information using the modified analogy approach. The procedure for converting a transfer function to a block diagram model is presented in this section as a six-step process. This unique method combines the standard analogy approach to modeling with block diagrams, the major difference being the ability to incorporate nonlinearities directly without bringing in linearization. Chapter 2 addresses a variety of physical systems often found in mechatronics. Such systems include mechanical, electrical, thermal, fluid, and hydraulic components. Models and techniques developed in this chapter are used in subsequent chapters in the chronology of the mechatronics design process.

Chapter 3 presents the basic theoretical concepts of sensors and transducers. The topics include instrumentation principles, analog and digital sensors, sensors for position, force, and vibration, and sensors for temperature, flow, and range.

Chapter 4 discusses several types of actuating devices, including DC motors, stepper motors, fluid power devices and piezoelectric actuators.

Chapter 5 looks at system control and logic methods. This includes fundamental aspects of digital techniques, digital theory such as Boolean logic, analog and digital electronics, and programmable logic controllers.

Chapter 6 presents controls and their design for use in mechatronics systems. Special attention is paid to real-world constraints, including time delays and nonlinearities. The Root Locus and Bode Plot design methods are discussed in detail, along with several design procedures for common control structures, including PI, PD, PID, lag, lead, and pure gain.

Chapter 7 discusses the theoretical and practical aspects of real-time data acquisition. Signal processing and data interpretation are handled using the visual programming approach. Several examples using LabVIEW and VisSim are presented. A case study involving pulse width modulation of a PI controller output of the PM DC Gear Motor Position Control System is also presented.

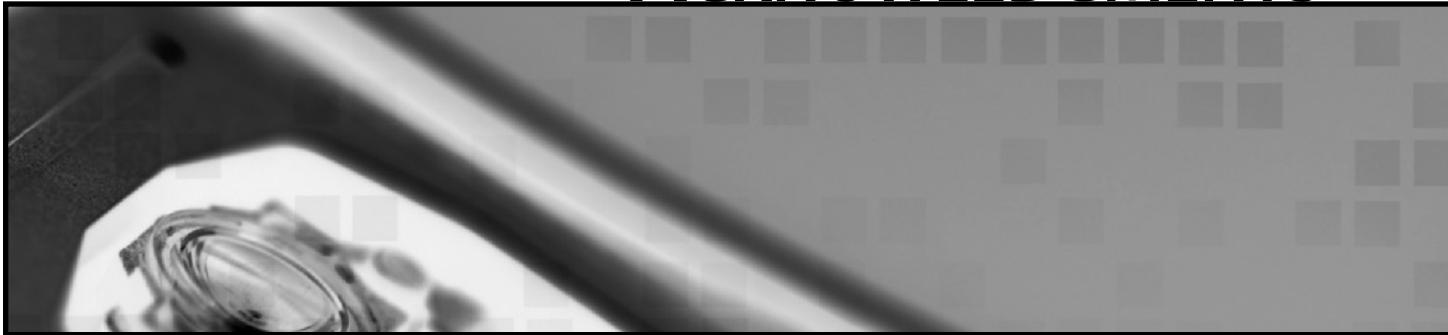
Chapter 8 presents a collection of case studies suitable for laboratory investigations. All case studies are implemented using a general purpose I/O board, visual simulation environment, and application software. The key aspect of the graphical environments is that the visual representation of system partitioning and interaction lends itself to mechatronics applications.

The combination of class discussions, simulation projects, and laboratory experimental design exposes the students to a practical platform of mechatronics. The real challenge in writing this book has been to connect complex and seemingly independent topics in a clear and concise manner, which is necessary for the understanding of mechatronics. The users of the book are requested to give feedback for further improvement of the text.

For students: Instructions for downloading the VisSim trial version can be found by visiting the textbook's student companion site. Please visit www.cengage.com/engineering/shetty for more information.

For instructors: Additional resources can be found on the textbook's instructor companion site. Please visit www.cengage.com/engineering/shetty for more information.

ACKNOWLEDGMENTS



The material presented in this book is a collection of many years of research and teaching by the authors at the University of Hartford, Cooper Union, and Lawrence Technological University as well as the insight gained from working closely with industry affiliates such as United Technologies, McDonnell Douglas, and many others.

Many have contributed greatly, in reviewing the manuscript. We wish to acknowledge the hundreds of students from the classes in which we have tested the teaching material. We are grateful to a number of professors whose comments and suggestions at various stages of this project were helpful in revising the manuscript. We would like to acknowledge Prof. Claudio Campana of University of Hartford, Prof. Ridha Ben Mrad of University of Toronto, Prof. M.K. Ramasubramanian of North Carolina State University, and George Thomas of Lawrence Technological University.

Special thanks to Dr. Walter Harrison, President of the University of Hartford; Dr. Lewis Walker, President of Lawrence Technological University; Dr. Donna Randall, President of the Albion College; Dr. Maria Vaz, Provost of Lawrence Technological University; and Dean Lou Manzione and Dr. Ivana Milanovic of the University of Hartford for their encouragement. We thank Visual Solutions, Inc. and National Instruments Inc. for their assistance with the real-time interfacing portion of the text.

Funding from the National Science Foundation and United Technologies Mechatronics Grant is gratefully acknowledged. The tremendous support and encouragement that we have received from our colleagues has been invaluable.

*Devdas Shetty
Richard Kolk*

MECHATRONICS SYSTEM DESIGN

SECOND EDITION, SI

CHAPTER 1

MECHATRONICS SYSTEM DESIGN

- 
- 1.1 What is Mechatronics
 - 1.2 Integrated Design Issues in Mechatronics
 - 1.3 The Mechatronics Design Process
 - 1.3.1 Important Features
 - 1.3.2 Hardware in the Loop Simulation
 - 1.4 Mechatronics Key Elements
 - 1.4.1 Information Systems
 - 1.4.2 Mechanical Systems
 - 1.4.3 Electrical Systems
 - 1.4.4 Sensors and Actuators
 - 1.4.5 Real-Time Interfacing
 - 1.5 Applications in Mechatronics
 - 1.5.1 Condition Monitoring
 - 1.5.2 Monitoring On-Line
 - 1.5.3 Model-Based Manufacturing
 - 1.5.4 Supervisory Control Structure
 - 1.5.5 Open Architecture Matters with Mechatronic Models: Speed and Complexity
 - 1.5.6 Interactive Modeling
 - 1.5.7 Right First Time—Virtual Machine Prototyping
 - 1.5.8 Evaluating Trade Off
 - 1.5.9 Embedded Sensors and Actuators
 - 1.5.10 Rapid Prototyping of a Mechatronic Product
 - 1.5.11 Optomechatronics
 - 1.5.12 E-Manufacturing
 - 1.5.13 Mechatronic Systems in Use
 - 1.6 Summary
 - References
 - Problems

This chapter provides the student with an overview of the mechatronic design process and a general description of the technologies employed in the mechatronic approach. This chapter begins by introducing the key elements, techniques, and design processes used for the mechatronics system design. Following a definition of mechatronics and a discussion of several important design issues, the mechatronic key elements of information systems, electrical systems, mechanical systems, computer systems, sensors, actuators, and real-time interfacing are introduced. Characteristics pertinent to mechatronics are developed from these first principles. Although experience in any of the supporting technologies is helpful, it is not necessary. The chapter closes with a description of the mechatronics design process and a discussion of some emerging trends in simulation, modeling, and smart manufacturing.

1.1 What is Mechatronics

Mechatronics is a methodology used for the optimal design of electromechanical products.

A *methodology* is a collection of practices, procedures, and rules used by those who work in a particular branch of knowledge or *discipline*. Familiar technological disciplines include thermodynamics, electrical engineering, computer science, and mechanical engineering, to name several. Instead

of one, the mechatronic system is *multidisciplinary*, embodying four fundamental disciplines: *electrical, mechanical, computer science, and information technology*.

The F-35, a U.S. Department of defense joint strike fighter plane developed by Lockheed Martin Corporation, is an example of mechatronic technology in action. The design metric emphasizes reliability, maintainability, performance, and cost. Multidisciplinary functions, including the on-board prognostics for zero downtime and cockpit technology, are being designed into the aircraft starting at the preliminary design stage.

Multidisciplinary systems are not new. They have been successfully designed and used for many years. One of the most common is the *electromechanical* system, which often uses a computer algorithm to modify the behavior of a mechanical system. Electronics are used to transduce information between the computer science and mechanical disciplines.

The difference between a mechatronic system and a multidisciplinary system is not the constituents, but rather *the order in which they are designed*. Historically, multidisciplinary system design employed a sequential *design-by-discipline* approach. For example, the design of an electromechanical system is often accomplished in three steps, beginning with the mechanical design. When the mechanical design is complete, the power and microelectronics are designed, followed by the control algorithm design and implementation. The major drawback of the design-by-discipline approach is that, by fixing the design at various points in the sequence, new constraints are created and passed on to the next discipline. Many control system engineers are familiar with the quip:

Design and build the mechanical system, then bring in the painters to paint it and the control system engineers to install the controls.

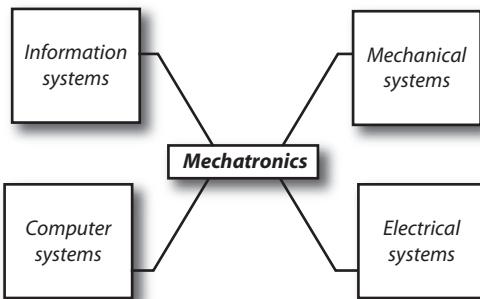
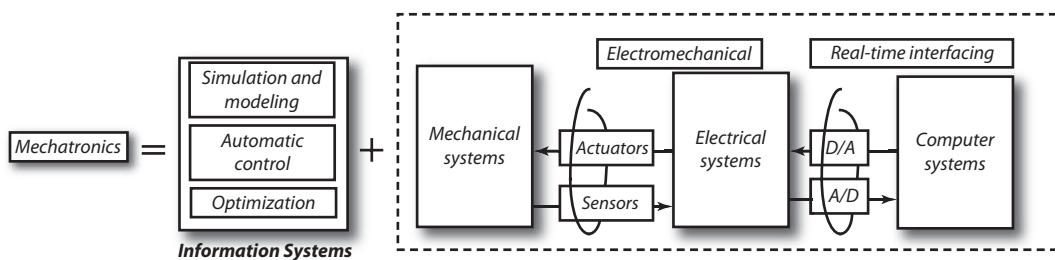
Control designs often are not efficient because of these additional constraints. For example, cost reduction is a major factor in most systems. Trade offs made during the mechanical and electrical design stages often involve sensors and actuators. Lowering the sensor–actuator count, using less accurate sensors, or using less powerful actuators, are some of the standard methods for achieving cost savings.

The mechatronic design methodology is based on a concurrent (instead of sequential) approach to discipline design, resulting in products with more synergy.

The branch of engineering called *systems engineering* uses a concurrent approach for *preliminary design*. In a way, mechatronics is an extension of the system engineering approach, but it is *supplemented* with information systems to *guide* the design and is applied at *all* stages of design—not just the preliminary design step—making it more *comprehensive*. There is a synergy in the integration of mechanical, electrical, and computer systems with information systems for the design and manufacture of products and processes. The synergy is generated by the right combination of parameters; the final product can be better than just the sum of its parts. Mechatronic products exhibit performance characteristics that were previously difficult to achieve without the synergistic combination. The key elements of the mechatronics approach are presented in Figure 1-1.

Even though the literature often adopts this concise representation, a clearer but more complex representation is shown in Figure 1-2.

Mechatronics is the result of applying information systems to physical systems. The physical system (the rightmost dotted block of Figure 1-2) consists of mechanical, electrical, and computer systems as well as actuators, sensors, and real-time interfacing. In some of the literature, this block is called an *electromechanical* system.

FIGURE 1-1 MECHATRONICS CONSTITUENTS**FIGURE 1-2 MECHATRONICS KEY ELEMENTS**

A mechatronic system is not an electromechanical system but is more than a control system.

Mechatronics is really nothing but good design practice. The basic idea is to apply new controls to extract new levels of performance from a mechanical device. Sensors and actuators are used to transduce energy from high power (usually the mechanical side) to low power (the electrical and computer side). The block labeled “*Mechanical systems*” frequently consists of more than just mechanical components and may include fluid, pneumatic, thermal, acoustic, chemical, and other disciplines as well. New developments in sensing technologies have emerged in response to the ever-increasing demand for solutions of specific monitoring applications. Microsensors are developed to sense the presence of physical, chemical, or biological quantities (such as temperature, pressure, sound, nuclear radiations, and chemical compositions). They are implemented in solid-state form so that several sensors can be integrated and their functions combined.

Control is a general term and can occur in living beings as well as machines. The term “*Automatic control*” describes the situation in which a machine is controlled by another machine. Irrespective of the application (such as industrial control, manufacturing, testing, or military), new developments in sensing technology are constantly emerging.

1.2 Integrated Design Issues in Mechatronics

The inherent concurrency or simultaneous engineering of the mechatronics approach relies heavily on the use of system modeling and simulation throughout the design and prototyping stages. Because the model will be used and altered by engineers from multiple disciplines, it is especially important that it be programmed in a visually intuitive environment. Such environments include block diagrams, flow charts, state transition diagrams, and bond graphs. In contrast to the more conventional programming languages such as Fortran, Visual Basic, C++, and Pascal, the visual modeling environment requires little training due to its inherent intuitiveness. Today, the most widely used visual programming environment is the *block diagram*. This environment is extremely versatile, low in cost, and often includes a *code generator* option, which translates the block diagram into a C (or similar) high-level language suitable for target system implementation. Block diagram-based modeling and simulation packages are offered by many vendors, including MATRIXxTM, Easy5TM, SimulinkTM, Agilent VEETM, DASYLabTM, VisSimTM, and LabVIEWTM.

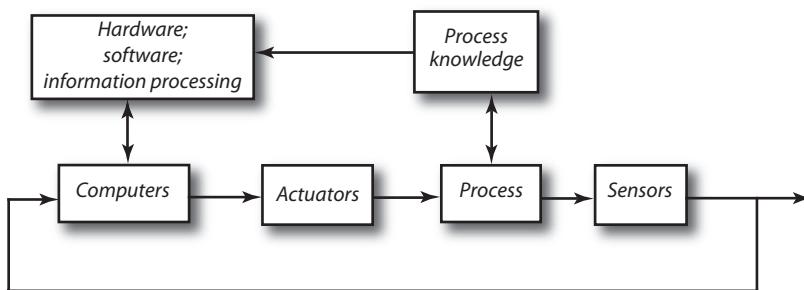
Mechatronics is a design philosophy: an integrating approach to engineering design. The primary factor in mechatronics is the involvement of these areas throughout the design process. Through a mechanism of simulating interdisciplinary ideas and techniques, mechatronics provides ideal conditions to raise the synergy, thereby providing a catalytic effect for the new solutions to technically complex situations. An important characteristic of mechatronic devices and systems is their built-in intelligence that results through a combination of precision in mechanical and electrical engineering, and real-time programming integrated into the design process. Mechatronics makes the combination of actuators, sensors, control systems, and computers in the design process possible.

Starting with basic design and progressing through the manufacturing phase, mechatronic design optimizes the parameters at each phase to produce a quality product in a short-cycle time. Mechatronics uses the control systems to provide a coherent framework of component interactions for system analysis. The integration within a mechatronic system is performed through the combination of hardware (components) and software (information processing).

- Hardware integration results from designing the mechatronic system as an overall system and bringing together the sensors, actuators, and microcomputers into the mechanical system.
- Software integration is primarily based on advanced control functions.

Figure 1-3 illustrates how the hardware and software integration takes place. It also shows how an additional contribution of the process knowledge and information processing is involved besides the feedback process.

FIGURE 1-3 GENERAL SCHEME OF HARDWARE AND SOFTWARE INTEGRATION



The first step in the focused development of mechatronic systems is to analyze the customer needs and the technical environment in which the system is integrated. Complex systems designed to solve problems tend to be a combination of mechanical, electric, fluid power, and thermodynamic parts, with hardware in the digital and analog form, coordinated by complex software. Mechatronic systems gather data from their technical environment using sensors. The next step is to use elaborate modeling and description methods to cover all subtasks of this system in an integrated manner. This includes an effective description of the necessary interfaces between subsystems at an early stage. The data is processed and interpreted, thus leading to actions carried out by actuators. The advantages of mechatronic systems are shorter developmental cycles, lower costs, and higher quality.

Mechatronic design supports the concepts of concurrent engineering.

In the designing of a mechatronic product, it is necessary that the knowledge and necessary information be coordinated amongst different expert groups. Concurrent engineering is a design approach in which the design and manufacture of a product are merged in a special way. It is the idea that people can do a better job if they cooperate to achieve a common goal. It has been influenced partly by the recognition that many of the high costs in manufacturing are decided at the product design stage itself. The characteristics of concurrent engineering are

- Better definition of the product without late changes.
- Design for manufacturing and assembly undertaken in the early design stage.
- Process on how the product development is well defined.
- Better cost estimates.
- Decrease in the barriers between design and manufacturing.

However, the lack of a common interface language has made the information exchange in concurrent engineering difficult. Successful implementation of concurrent engineering is possible by coordinating an adequate exchange of information and dealing with organizational barriers to cross-functional cooperation.

Using concurrent engineering principles as a guide, the designed product is likely to meet the basic requirements:

- High quality
- Robustness
- Low cost
- Time to market
- Customer satisfaction

The benefits that accrue due to the integration of concurrent engineering management strategy are greater productivity, higher quality, and reliability due to the introduction of an intelligent, self-correcting sensory and feedback system. The integration of sensors and control systems in a complex system reduces capital expenses, maintains a high degree of flexibility, and results in higher machine utilization.

1.3 The Mechatronics Design Process

The traditional electromechanical-system design approach attempted to inject more reliability and performance into the *mechanical* part of the system during the development stage. The control part of the system was then designed and added to provide additional performance or reliability and also to correct undetected errors in the design. Because the design steps occur sequentially, the traditional approach is a *sequential engineering* approach. A Standish Group survey of software dependent projects found.

- 31.1% cancellation rate for software development projects.
- 222% time overrun for completed projects.
- 16.2% of all software projects were completed on time and within budget.
- Maintenance costs exceeded 200% of initial development costs for delivered software.

The Boston-based technology think tank, Aberdeen Group, provided key information on the importance of incorporating the right design process for a mechatronic system design. Aberdeen researchers used five key product development performance criteria to distinguish “*best-in-class*” companies, as related to mechatronic design. The key criteria were revenue, product cost, product launch dates, quality, and development costs. Best-in-class companies proved to be twice as likely as “*laggards*” (worst-in-class companies) to achieve revenue targets, twice as likely to hit product cost targets, three times as likely to hit product launch dates, twice as likely to attain quality objectives, and twice as likely to control their development costs. Aberdeen’s research also revealed that best-in-class companies were.

- 2.8 times more likely than *laggards* to carefully communicate design changes across disciplines.
- 3.2 times more likely than *laggards* to allocate design requirements to specific systems, subsystems, and components.
- 7.2 times more likely than *laggards* to digitally validate system behavior with the simulation of integrated mechanical, electrical, and software components.

A major factor in this sequential approach is the inherently complex nature of designing a multidisciplinary system. Essentially, mechatronics is an improvement upon existing lengthy and expensive design processes. Engineers of various disciplines work on a project simultaneously and cooperatively. This eliminates problems caused by design incompatibilities and reduces design time because of fewer returns. Design time is also reduced through extensive use of powerful computer simulations, reducing dependency upon prototypes. This contrasts the more traditional design process of keeping engineering disciplines separate, having limited ability to adapt to mid-design changes, and being dependent upon multiple physical prototypes.

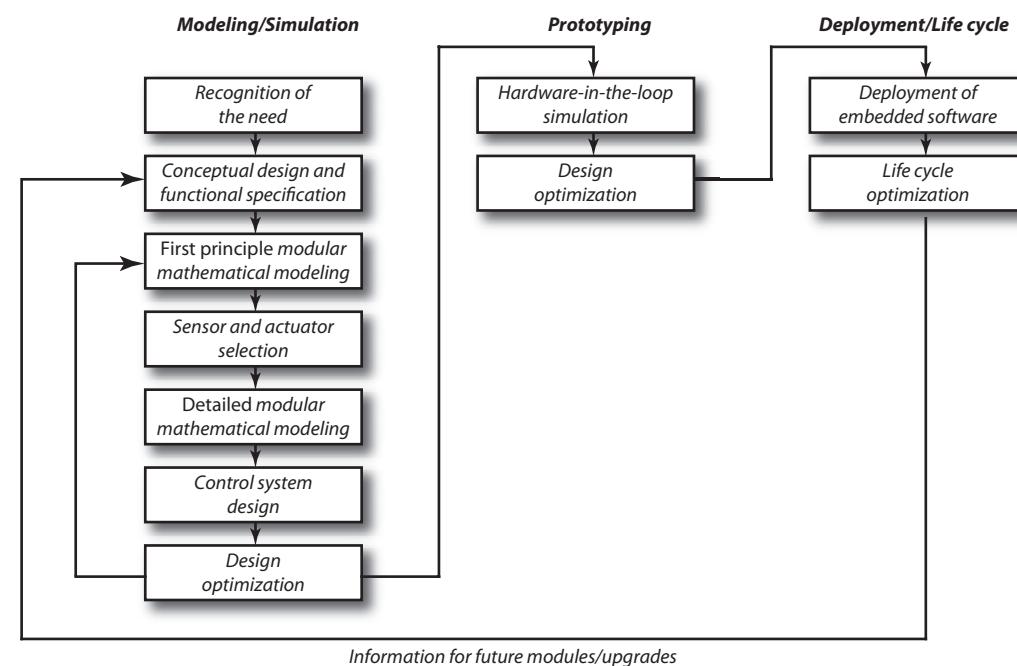
The mechatronic design methodology is not only concerned with producing high-quality products but with maintaining them as well—an area referred to as *life cycle design*. Several important life cycle factors are indicated.

- *Delivery:* Time, cost, and medium.
- *Reliability:* Failure rate, materials, and tolerances.
- *Maintainability:* Modular design.

- *Serviceability*: On board diagnostics, prognostics, and modular design.
 - *Upgradeability*: Future compatibility with current designs.
 - *Disposability*: Recycling and disposal of hazardous materials.

We will not dwell on life cycle factors except to point out that the *conventional design for life cycle* approach *begins* with a product *after* it has been designed and manufactured. In the mechatronic design approach, life cycle factors are *included* during the product design stages, resulting in products which are designed from *conception* to *retirement*. The mechatronic design process is presented in Figure 1-4.

FIGURE 1-4 MECHATRONIC DESIGN PROCESS



The mechatronic design process consists of three phases: modeling and simulation, prototyping, and deployment. All modeling, whether based on first principles (basic equations) or the more detailed physics, should be modular in structure. A first principle model is a simple model which captures some of the fundamental behavior of a subsystem. A detailed model is an extension of the first principle model providing more function and accuracy than the first level model. Connecting the modules (or blocks) together may create complex models. Each block represents a subsystem, which corresponds to some physically or functionally realizable operations, and can be *encapsulated* into a block with input/output limited to input signals, parameters, and output signals. Of course, this limitation may not always be possible or desirable; however, its use will produce modular subsystem blocks which easily can be maintained, exercised independently, substituted for one another (first principle blocks substituted for detailed blocks and vice versa), and reused in other applications.

Because of their modularity, mechatronic systems are well suited for applications that require reconfiguration. Such products can be reconfigured either during the design stage by substituting various subsystem modules or during the life span of the product. Since many of the steps in the mechatronic design process rely on computer-based tasks (such as information fusion, management, and design testing), an efficient computer-aided prototyping environment is essential.

Important Features

- *Modeling:* Block diagram or visual interface for creating intuitively understandable behavioral models of physical or abstract phenomenon. The ability to encapsulate complexity and maintain several levels of subsystem complexity is useful.
- *Simulation:* Numerical methods for solving models containing differential, discrete, hybrid, partial, and implicit nonlinear (as well as linear) equations. Must have a *lock* for real-time operation and be capable of executing faster than real time.
- *Project Management:* Database for maintaining project information and subsystem models for eventual reuse.
- *Design:* Numerical methods for constrained optimization of performance functions based on model parameters and signals. Monte Carlo type of computation is also desirable.
- *Analysis:* Numerical methods for frequency-domain, time-domain, and complex-domain design.
- *Real-Time Interface:* A plug-in card is used to replace part of the model with actual hardware by interfacing to it with actuators and sensors. This is called *hardware in the loop* simulation or *rapid prototyping* and must be executed in real time.
- *Code Generator:* Produces efficient high-level source code from the block diagram or visual modeling interface. The control code will be compiled and used on the embedded processor. The language is usually C.
- *Embedded Processor Interface:* The embedded processor resides in the final product. This feature provides communication between the process and the computer-aided prototyping environment. This is called a *full system prototype*.

Because no single model can ever flawlessly reproduce reality, there always will be error between the behavior of a product model and the actual product. These errors, referred to as *unmodeled errors*, are the reason that so many *model-based designs* fail when deployed to the product. The mechatronic design approach also uses a model-based approach, relying heavily on modeling and simulation. However, *unmodeled errors are accounted for* in the *prototyping* step. Their effects are absorbed into the design, which significantly raises the probability of successful product deployment.

Hardware-in-the-Loop Simulation In the prototyping step, many of the non-computer subsystems of the model are replaced with actual hardware. Sensors and actuators provide the interface signals necessary to connect the hardware subsystems back to the model. The resulting *model* is part mathematical and part real. Because the real part of the model inherently evolves in real time and the mathematical part evolves in simulated time, it is essential that the two parts be synchronized. This process of fusing and synchronizing model, sensor, and actuator information is called *real-time interfacing* or *hardware-in-the-loop simulation*, and is an essential ingredient in the modeling and simulation environment.

TABLE 1-1 DIFFERENT CONFIGURATIONS FOR HARDWARE-IN-THE-LOOP SIMULATION

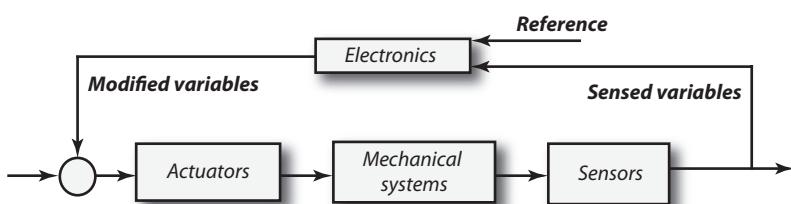
Real Hardware Components	Mathematically Modeled Components	Description
• Sensors • Actuators • Process	• Control algorithm	Modify control system design subject to unmodelled sensor, actuator, and machinery errors.
• Sensors • Actuators • Control (including the embedded computer)	• Process	Evaluate validity of process model.
• Protocol (for distributed applications)	• Control algorithm • Sensors • Actuators • Process	Evaluate the effects of data transmission on design.
• Signal processing hardware	• Control algorithm • Sensors • Actuators • Process	Evaluate the effects of actual signal processing hardware.

So far, we have only discussed one configuration for hardware-in-the-loop simulation. This and other possibilities are summarized in Table 1-1. Table 1-1 assumes the following six functions.

- *Control*: The control algorithm(s) in executable software form.
- *Computer*: The embedded computer(s) used in the product.
- *Sensors*
- *Actuators*
- *Process*: Product hardware excluding sensors, actuators, and the embedded computer.
- *Protocol* (optional): For bus-based distributed control applications.

The comprehensive development of mechatronic systems starts with modeling and simulation, model building for static and dynamic models, transformation into simulation models, programming-and computer-based control, and final implementation. In this atmosphere, hardware-in-the-loop simulation plays a major part. Using visual simulation tools in a real-time environment, major portions of the mechatronic product could be simulated along with the hardware-in-the-loop simulation.

The hardware-in-the-loop model (Figure 1-5) shows the different components of a mechatronic system. It is possible to simulate the electronics where the actuators, mechanics and sensors are the

FIGURE 1-5 HARDWARE-IN-THE-LOOP MODEL

real hardware. On the other hand, if appropriate models of the mechanical systems, actuators, and sensors are available, the electronics could be the only hardware. There are different ways in which hardware-in-the-loop could be simulated, such as electronics simulation, simulation of actuators and sensors, or simulation of mechanical systems alone.

1.4 Mechatronics Key Elements

1.4.1 Information Systems

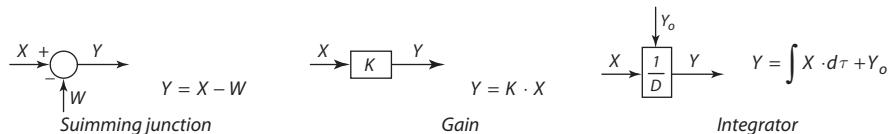
Information systems include all aspects of information transmission—from signal processing to control systems to analysis techniques. An information system is a combination of four disciplines: communication systems, signal processing, control systems, and numerical methods. In mechatronics applications, we are most concerned with modeling, simulation, automatic control, and numerical methods for optimization.

Modeling and Simulation *Modeling* is the process of representing the behavior of a real system by a collection of mathematical equations and logic. The term *real system* is synonymous with *physical system*—that is, a system whose behavior is based on matter and energy. Models can be broadly categorized as either *static* or *dynamic*. In a static model, there is no energy transfer. Systems, which are static produce no motion, heat transfer, fluid flow, traveling waves, or any other changes. On the other hand, a dynamic model has energy transfer which results in power flow. Power, or rate of change of energy, causes motion, heat transfer, and other phenomena that change in time. Phenomena are observed as *signals*, and since time is often the independent variable, most signals are indexed with respect to time.

Models are cause-and-effect structures—they accept external information and process it with their logic and equations to produce one or more outputs. Exogenous, or externally produced, information supplied to the model either can be fixed in value or changing. An external fixed-value unit of information is called a *parameter*, while an external changing unit of information is called an *input signal*. Traditionally, all model output information is assumed to be changing and is therefore referred to as *output signals*.

Because models are collections of mathematical and logic expressions, they can be represented in text-based programming languages. Unfortunately, once in the programming language, one must be familiar with the specific language in order to understand the model. Because most practicing engineers are not familiar with most programming languages, text-based modeling proved to be a poor candidate for mechatronics. The ideal candidate would be picture or *visual* based instead of text-based and intuitive.

All block diagram languages consist of two fundamental objects: signal wires and blocks. A signal wire transmits a signal or a value from its point of origination (usually a block) to its point of termination (usually another block). An arrowhead on the signal wire defines the direction in which the signal flows. Once the flow direction has been defined for a given signal wire, signals may only flow in the forward direction—not backwards. A block is a processing element which operates on input signals and parameters (or constants) to produce output signals. Because block functions can be nonlinear as well as linear, the collection of special function blocks is practically unlimited and almost never the same between vendors. However, there is a three-block basis that all block diagram languages possess: summing junction, gain, and integrator blocks. These blocks and their associated functions are presented in Figure 1-6.

FIGURE 1-6 BASIC BLOCKS

Simulation is the process of *solving* the model and is performed on a computer. Although simulations can be performed on analog computers, it is far more common to perform them on digital computers. The process of simulation can be divided into three sections: initialization, iteration, and termination. If the starting point is a block diagram-based model description, then in the initialization section, the equations for each of the blocks must be sorted according to the pattern in which the blocks have been connected.

The iteration section solves any differential equations present in the model using numerical integration and/or differentiation. An ordinary differential equation is (in general) a nonlinear equation which contains one or more derivative terms as a function of a single independent variable. For most simulations, this independent variable is time. The order of an ordinary differential equation equals the highest derivative term present. Most methods employed for the numerical solution of ordinary differential equations are based on the use of *approximating polynomials*, which fit a truncated Taylor series expansion of the ordinary differential equation. Three steps are required:

- Step 1.** Write a Taylor series expansion of the functional form of the ordinary differential equation solution about its initial condition(s). Since the independent variable considered is time, all derivative terms in the series will be taken with respect to time.
- Step 2.** Truncate the Taylor series at one of the derivative terms, and the resulting truncated series becomes the approximating polynomial.
- Step 3.** Compute all constant terms and each derivative term based on the initial condition values to complete the approximating polynomial.

The display section of a simulation is used to present and post the output process. Output may be saved to a file, displayed as a digital reading, or graphically displayed as a chart, strip chart, meter readout, or even as an animation.

Optimization Optimization solves the problem of distributing limited resources throughout a system so that prespecified aspects of its behavior are satisfied. In mechatronics, optimization is primarily used to establish the optimal system configuration. However, it may be applied to other issues as well, such as

- Identification of optimal trajectories
- Control system design
- Identification of model parameters

In engineering applications, certain conventions in terminology are used. Resources are referred to as *design variables*, aspects of system behavior as *objectives*, and system governing relationships (equations and logic) as *constraints*.

To illustrate the formulation of an optimization problem, consider the following example. A system consists of a piece of box-shaped luggage, where the volume characteristics are to be maximized by appropriate selection of the height, width, and depth resources. The problem is formulated as

Design variables: L (length), W (width), H (height)

Objective: Maximize V (volume) = $\mathbf{V}(L, W, H)$

Constraints: System relationship: $V = LHW$

The objective is written in functional form to show its dependence on the design variables. This problem is easily solved mentally, since the resources are unlimited; the volume becomes infinite. More challenging and realistic situations occur when limits are placed on the resources. Consider placing a limit on the total distance resource (width plus height plus depth) of 80 cm. The problem formulation is presented as

Objective: Maximize V (volume) = $\mathbf{V}(L, W, H)$

Constraints: System relationship: $V = LHW$

Resources: $L + W + H \leq 80$

From basic geometry, we remember that cubic shapes have maximum volume; therefore, the total distance resource must be distributed equally among the height, width, and depth. Next, consider the addition of constraints on each of the three design variables. We will restrict the box length to be less than 40 cm, the width to be less than 30 cm, and the height to be less than 20 cm. The problem formulation becomes

Objective: Maximize V (volume) = $\mathbf{V}(L, W, H)$

Constraints: System relationship: $V = LHW$

Resources: $L + W + H \leq 80$

Side: $0 \leq L \leq 40$

$0 \leq W \leq 30$

$0 \leq H \leq 20$

The system relationship and resource constraints are often called just *constraints*. These are sometimes further divided into equality and inequality constraints. The system constraints are usually equality constraints and the resource constraints may be a combination of both. Constraints on the design variables themselves are called *side constraints*. Furthermore, the objective is called an *objective function*, and it is common in engineering applications to always minimize the objective function. This is because it is often associated with an error signal, which should ideally become zero. Maximizing an objective function is achieved by minimizing the negative of the objective function.

The objective function is the function that is minimized by the search algorithm of the optimization procedure by appropriate choice of the design variables. There is no prescribed general form that an objective function must obey, but the performance of the search algorithm (especially gradient-based algorithms) will be strongly tied to the characteristics of the objective function. These characteristics include: (1) the overall “smoothness” of the function, (2) the

magnitude similarity of the values of the objective function gradient, and (3) the overall numerical “slope” of the objective function.

The basic optimization procedure is the same for any application and requires the following formulation to be started.

1. Design variables: $P = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}$ and their initial *guessed* value $P_o = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}_o$
2. Objective function: $J = J(P)$
 $F(P) = 0$ (system constraints)
3. Constraints:
 $H(P) \leq 0$ (resource constraints)
 $P_{\text{low}} \leq P \leq P_{\text{high}}$ (side constraints)

The optimization process then iterates the equation; $P_{k+1} = P_k + \tau \cdot S_k$, where k is the iteration number, S_k is the search direction in P space, and τ is the stepsize moved in the search direction. The process terminates when no further improvement is made in P . At this point, $P^* = P$ (the asterisk superscript means *optimal*), and the objective function has been extremized (usually minimized) and becomes $J^* = J(P^*)$.

Due to inevitable nonlinearities, most objective functions will have many *local* minimum values, and the one found, $J^* = J(P^*)$, may not be the desired overall minimum (global minimum). One way of finding the global minimum is to make many optimization runs—each using a different initial parameter vector. Assuming enough runs were made, the global minimum becomes the minimum run collection. It is also possible to create an objective function that has no minimum, in which case the optimization process may produce nonsensical results. Care should be exercised when constructing an objective function to insure it has at least one minimum.

1.4.2 Mechanical Systems

Mechanical systems are concerned with the behavior of matter under the action of forces. Such systems are categorized as rigid, deformable, or fluid in nature. A rigid-body system assumes all bodies and connections in the system to be perfectly rigid. In actual systems, this is not true, and some deformation always results as various loads are applied. Normally, the deformations are small and do not appreciably affect the motion of the rigid-body system; however, when one is concerned with material failures, the deformable-body system becomes important. Failure analysis and mechanics of materials are major fields based on deformable-body systems. The field of fluid mechanics consists of compressible and incompressible fluids.

Newtonian mechanics provides the basis for most mechanical systems and consists of three independent and absolute concepts: space, time, and mass. A fourth concept, force, is also present but is not independent of the other three. One of the fundamental principles of Newtonian mechanics is that the force acting on a body is related to the mass of the body and the velocity variation over time. For systems involving the motion of particles with very high velocities, one must resort to relativistic, instead of Newtonian, mechanics (theory of relativity). In such systems, the three concepts are no longer independent (the mass of the particle is a function of its velocity).

Most mechatronic applications involve rigid-body systems, and the study of such systems relies on the following six fundamental laws.

- *Newton's First Law:* If the resultant force acting on a particle is zero, then the particle will remain at rest if it is originally at rest or will move with constant speed in a straight line if it is originally in motion.
- *Newton's Second Law:* If the force acting on a particle is not zero, then the particle will have an acceleration proportional to the magnitude of the force, $F = m \cdot a$.
- *Newton's Third Law:* The forces of action and reaction between bodies in contact have the same magnitude, line of action, and opposite sense.
- *Newton's Law of Gravitation:* Two particles of mass M and m are attracted with equal and opposite forces F and $-F$ according to the formula $F = G \cdot \frac{M \cdot m}{r^2}$, where r is the distance between the two particles and G is the *constant of gravitation*.
- *Parallelogram Law for the Addition of Forces:* Two forces acting on a particle may be replaced by a single *resultant* force obtained by drawing the diagonal of the parallelogram with sides equal to each of the two forces.
- *Principle of Transmissibility:* The point of application of an external force acting on a body (structure) may be transmitted anywhere along the force's line of action without affecting the other external forces (reactions and loads) acting on that body. This means that there is no net change in the static effect upon any body if the body is in equilibrium.

There are three different systems of units commonly found in engineering applications: the meter-kilogram-second (mks) or System International (SI) system, the centimeter-gram-second (cgs) or Gaussian system, and the foot-pound-second (fps) or British engineering system. In the SI and Gaussian systems, the kilogram and gram are mass units. In the British system, the pound is a force unit. In this book we will use the SI system throughout.

1.4.3 Electrical Systems

Electrical systems are concerned with the behavior of three fundamental quantities: charge, current, and voltage (or potential). When a current exists, electrical energy usually is being transmitted from one point to another. Electrical systems consist of two categories: power systems and communication systems. Communication systems are designed to transmit information as low-energy electrical signals between points. Functions such as information storage, processing, and transmission are common parts of a communication system. Electrical systems are an integral part of a mechatronics application. The following electrical components are frequently found in such applications.

- Motors and generators
- Sensors and actuators (transducers)
- Solid state devices including computers
- Circuits (signal conditioning and impedance matching, including amplifiers)
- Contact devices (relays, circuit breakers, switches, slip rings, mercury contacts, and fuses)

Electrical applications in mechatronic systems require an understanding of direct current (DC) and alternating current (AC) circuit analysis, including impedance, power, and electromagnetic as well as semiconductor devices (such as diodes and transistors). Some of the fundamental topics in these areas are introduced in the following sections.

DC and AC Circuit Analysis An electric circuit is a closed network of paths through which current flows. Any path of a circuit consists of circuit elements connected by electrical conductors called wires. Wires are assumed to be ideal or perfect conductors, which implies two conditions.

1. The potential at any point on the wire is the same.
2. Wires store no charge, so the current entering the wire equals the current leaving it.

An *open circuit* exists between two points in a circuit that are not connected by a branch, and a *short circuit* exists if the connection is a wire.

A *node* is a point at which two or more circuit elements are connected, and a path between two nodes is called a *branch*.

Circuit analysis is the process of calculating all voltages and currents in a circuit given the circuit diagram and a description of each element. The process is based on two fundamental laws named after Gustav Robert Kirchhoff (1824–1887). These laws, the current and the voltage law, are summarized here.

Kirchhoff's current law: The sum of all currents entering a node is zero.

Kirchhoff's voltage law: The sum of all voltage drops around a closed loop is zero.

In principle, any circuit can be analyzed by straightforward analytical application of these two laws. However, for large circuits, the algebra becomes tedious, and one often resorts to computer methods for solution.

A common method for describing the behavior of an electrical system element is through its *impedance*, Z , or V - I characteristic. For our purposes, the impedance of an element is the ratio of the voltage drop *across* the element divided by the current drawn *through* the element. The impedance of a resistor is just its resistance, $Z_R = R$. For a capacitor of capacitance C , it becomes

$$Z_C = \frac{1}{C \cdot D}$$

where D is the operator introduced in Figure 1-6.

For an inductor of inductance L , it is

$$Z_L = L \cdot D$$

As will be discovered in Chapter 2, the notion of impedance is an important concept which readily can be extended to other system disciplines (i.e., mechanical, fluid, and thermal).

Various techniques based on Kirchhoff's laws have been established, and combinations of these techniques are often employed to analyze a circuit. Techniques can be categorized depending on the circuit's dependency on time. For time-independent circuits (DC circuits), the following techniques are frequently used.

- Parallel and series branch reductions
- Node and loop analysis
- Voltage and current divider reductions
- Equivalent circuits (Thevenin and Norton equivalents)

Additional techniques for time-dependent circuits, which include periodic (AC) as well as non-periodic or transient, are

- Phasors
- Natural and forced response

Power Energy, which is the capacity to do work, may exist in various forms including *potential, kinetic, electrical, heat, chemical, nuclear, and radiant*. Radiant energy exists only in the absence of matter. The remaining energy forms both exist and can be converted amongst them only in the presence of matter. Power is the rate of energy transfer, and in the SI unit system, the unit of energy is the joule and the unit of power is the watt (1 watt = 1 joule per second).

In electrical systems, power is the product of current and voltage. As current flows through an electrical circuit, so does power, but unlike current, which must *remain* within the circuit, power can be converted to other forms, such as heat, which can *leave* or *enter* the circuit. One often needs to compute the amount of power entering or leaving some part of a circuit to determine how much *useful* power is being delivered. A good example of this process is the diesel-electric locomotive used in railroad applications. The diesel engine is used to power a generator, which in turn powers an electric motor used to move the train. The diesel engine is not *directly* used for motion because of its narrow torque band. By converting its power to electrical (through the generator) and then back to mechanical (through the electric motor), the torque-speed curve can be favorably reshaped to produce a broader torque spectrum more suited to this application. The power conversion does not come without loss, it is primarily through heat. During level and upgrade operation, the locomotive consumes power with a slight loss due to heat. During downgrade operation, the locomotive produces power, which can be either discarded or reused for braking—commonly called regenerative braking. The diesel-electric locomotive discards the power by passing the regenerated current through large resistors located under cooling fans along the top of the locomotive. These fans are used to assist the heat transfer process from the resistors to the atmosphere, keeping the resistors cool (and functional).

Fundamentally, electrical power is categorized as being either instantaneous or time averaged, as defined here.

$$\text{Instantaneous: } P(t) = v(t) \cdot i(t)$$

$$\text{Time averaged: } P_{AV} = \frac{1}{T} \int_0^T v(t) \cdot i(t) \cdot dt$$

1.4.4 Sensors and Actuators

Sensors are required to monitor the performance of machines and processes. Using a collection of sensors, one can monitor one or more variables in a process. Sensing systems also can be used to evaluate operations, machine health, inspect the work in progress, and identify part and tools. The monitoring devices are generally located near the manufacturing process measuring the surface quality, temperature, vibrations, and flow rate of cutting fluid. Sensors are needed to provide real-time information that can assist controllers in identifying potential bottlenecks, breakdowns, and other problems with individual machines and within a total manufacturing environment.

Accuracy and repeatability are critical capabilities; without which sensors cannot provide the reliability needed to perform in advanced manufacturing environments. When used with intelligent processing equipment, sensors must be able to discern weak signals while remaining insensitive to other interfering impulses. Sensors must be able to ascertain conditions instantaneously and accurately, as well as able to provide usable data to system controllers.

Some of the more common measurement variables in mechatronic systems are temperature, speed, position, force, torque, and acceleration. When measuring these variables, several characteristics become important: the dynamics of the sensor, stability, resolution, precision, robustness, size, and signal processing. The need for less expensive and more precise sensors, as well as the need for the integration of the sensor and the signal processing on a common carrier or on one chip, has become important.

Progress in semiconductor manufacturing technology has made it possible to integrate various sensory functions. Intelligent sensors are available that not only sense information but process it as

well. These sensors facilitate operations normally performed by the control algorithm, which include automatic noise filtering, linearization sensitivity, and self-calibration. Microsensors could be used to measure the flow, pressure, or concentration of various chemical species in environmental and mechanical applications.

The resonant microbeams already are being used to sense linear and rotational acceleration. The sensor is mounted on a data glove to detect the characteristic accelerations of human gestures. Many microsensors, including biosensors and chemical sensors can be mass produced. The ability to combine these mechanical structures and electronic circuitry on the same piece of silicon is also important.

Actuators are another important component of a mechatronic system. Actuation involves a physical action on the process, such as the ejection of a work piece from a conveyor system initiated by a sensor. Actuators are usually electrical, mechanical, fluid power or pneumatic based. They transform electrical inputs into mechanical outputs such as force, angle, and position. Actuators can be classified into three general groups.

1. Electromagnetic actuators, (e.g., AC and DC electrical motors, stepper motors, electromagnets)
2. Fluid power actuators, (e.g., hydraulics, pneumatics)
3. Unconventional actuators (e.g., piezoelectric, magnetostrictive, memory metal)

There are also special actuators for high-precision applications which require fast responses. They are often applied to controls which compensate for friction, nonlinearities, and limiting parameters.

Nanofabrication or micromachining refers to the creation of smaller structures—down to the control and arrangement of individual atoms. Such techniques are still being developed but offer fascinating potential. Microfabrication and nanofabrication involve the fabrication and manipulation of materials and objects at microscopic (microfabrication) and atomic (nanofabrication) levels often on a scale of less than one micron. Microfabrication processes include lithography, etching, deposition, epitaxial growth, diffusion, implantation, testing, inspection, and packaging. Nanofabrication includes some of these but also involves atomic-scale tailoring and patterning of materials to utilize their natural properties to achieve desired results.

1.4.5 Real-Time Interfacing

Simulation of a mathematical model is unrelated to real time—the time read from a wall clock. We often would like the model to *run* (or simulate) faster, but there is no harm if it does not. Consider a model which consists of several subsystems categorized as control algorithms, sensors, actuators, and the process (mechanical, thermal, fluid, etc.). The process of simulation requires that all cause-and-effect equations in the model be ordered (or *sorted*) with inputs on the left and outputs on the right prior to simulation. During simulation, the sorted equations are solved, time is advanced, the equations are again solved, and the process continues. One passage through the equations is called a *loop*.

The *real-time interface* process really falls into the electrical and information system categories but is treated independently as was computer system hardware because of its specialized functions. In mechatronics, the main purpose of the real-time interface system is to provide data acquisition and control functions for the computer. The purpose of the acquisition function is to reconstruct a sensor waveform as a digital sequence and make it available to the computer software for processing. The control function produces an analog approximation as a series of small *steps*. The inherent step discontinuities produce new undesirable frequencies not present in the original signal and are often attenuated using an analog *smoothing* filter. Thus, for mechatronic applications, real-time interfacing includes analog to digital (A/D) and digital to analog (D/A) conversion, analog signal conditioning circuits, and sampling theory.

1.5 Applications in Mechatronics

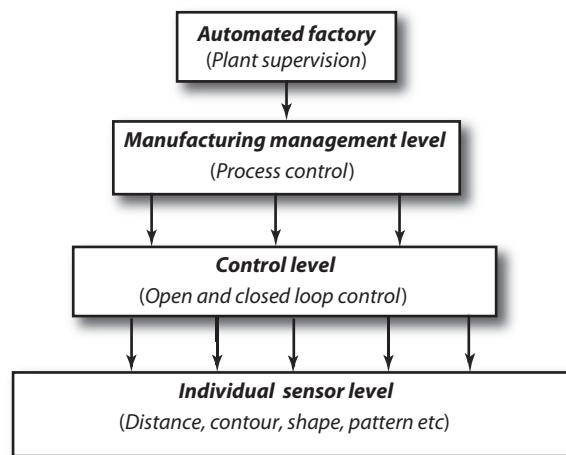
1.5.1 Condition Monitoring

The success of manufacturing process automation hinges primarily on the effectiveness of process monitoring and control systems. An automated factory is required to have sensors at different levels in the production system. Sensors help the production processes by compensating for unexpected disturbances, any tolerance changes in the work pieces, or other changes due to product/process problems. Intelligent manufacturing systems use automated diagnostic systems that handle machinery maintenance and process control operations.

Condition monitoring is defined as the determination of the machine status or the condition of a device and its change with time in order to decide its condition at any given time. The condition of the machines can be determined by physical parameters (like tool wear, machine vibration, noise, temperature, oil contamination, and debris). A change in these parameters provides an indication of the changing machine condition.

If the machine conditions are properly analyzed, they can become a valuable tool in establishing a maintenance schedule and in the prevention of machinery failures and breakdowns. The diagnostic parameters can be measured and monitored continuously at predetermined intervals. In some cases, measurement of secondary parameters such as pressure drop, flow, and power can lead to information on primary parameters such as vibration, noise, and corrosion. The data coming from different levels of the factory provide support for automated manufacturing. Sensors integrated with adaptive processes control capability at the plant level, manufacturing management level, control level, or sensory level and handle the requirements as shown in Figure 1-7.

FIGURE 1-7 SENSOR DISTRIBUTION AT DIFFERENT LEVELS OF PRODUCTION



At the sensory level, frequently required tasks in production processes are distance measurement, contour tracking, pattern recognition, identification of process parameters, and machine diagnostics. The selection of the sensing principle and parameters monitored are shown in Table 1-2.

In the case of manufacturing machinery, sensors can monitor machining operations, conditions of cutting tools, availability of raw material, and work in progress. Sensors can assist in

TABLE 1-2 EXAMPLES OF SENSING PARAMETERS IN AUTOMATED MANUFACTURING

Measurement	Sensing Parameter	Principle
Distance measurement	<ul style="list-style-type: none"> Edge detection; Monitoring the distance between tool and work piece as in laser cutting machines; Collision avoidance in robotics 	<ul style="list-style-type: none"> Potentiometric, inductive, capacitive principle Non-contact sensors, such as optical or ultrasonic sensors Laser interferometer Laser digitizer
Contour measurement	<ul style="list-style-type: none"> Detection of edges and surfaces Robot guided tools in welding operation 	<ul style="list-style-type: none"> Inductive, capacitive Non-contact sensors, such as optical, fiber optic, or ultrasonic sensors
Pattern recognition	<ul style="list-style-type: none"> Shape information Object classification 	<ul style="list-style-type: none"> Optical Tactile Ultrasonic
Machine diagnostics	<ul style="list-style-type: none"> Cutting tool condition Tool wear, breakage Machine vibration Power consumption 	<ul style="list-style-type: none"> Force, torque Current, frequency Amplitude, acceleration Surface roughness, roundness

the recognition of parts, tools, and pallets. They also can be used on the production floor during pre-process situations or at the time when the manufacturing process is in progress.

Figure 1-8 shows the basic elements of condition monitoring for machine tools during a production process. The monitoring system can provide data on the torque produced during machining operation and other data for tool management. The condition monitoring systems can be of two types.

1. Monitoring systems that display the machine conditions to enable the operator to make decisions.
2. Automated monitoring of conditions with adaptive control features.

As shown in Figure 1-9 on the next page, machine condition evaluation is applied for checking the status of cutting tools, work piece assembly, detection of collision, and monitoring of cutting tool

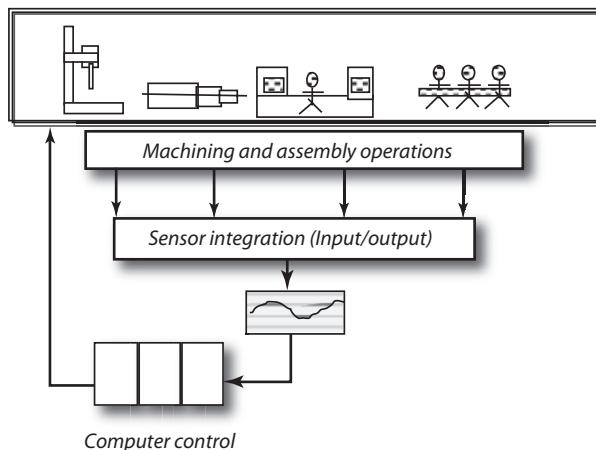
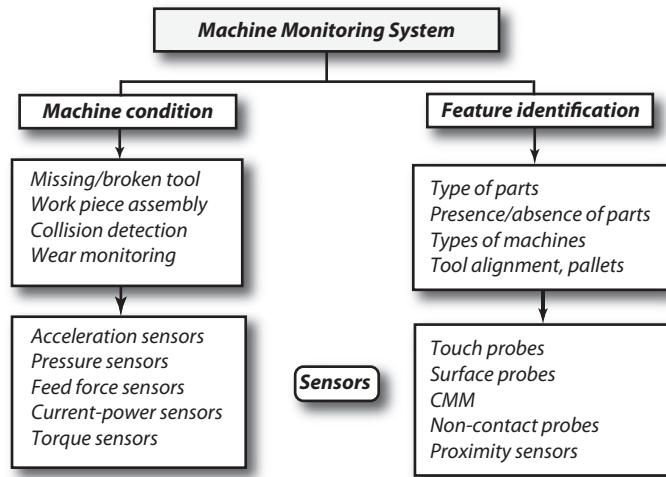
FIGURE 1-8 CONDITION MONITORING SYSTEM FOR TYPICAL PRODUCTION SYSTEMS

FIGURE 1-9 MONITORING SYSTEMS IN MACHINE TOOLS

wear, whereas the feature identification methodology is applied to detect the type of parts, shape of the work piece, alignment of cutting tools, types, and nature of pallets.

Monitoring of Vibration, Temperature, and Wear Vibration, or noise signature, of a machine is very much related to the health of a machine. Precise measurement of vibration levels on bearing housings and measurement of relative translation between shaft and bearings can provide useful information regarding faults such as unbalance, misalignment, lack of lubrication, and wear in machines. In turbo-machinery, resonance and vibration analysis is an established method of diagnosing deteriorating conditions. The frequency spectrum of vibration in a ball bearing can provide a comparison between a defective and a good ball bearing. The level of vibrations and presence of additional peaks are an indication of defects. Figures 1-10 and 1-11 show typical mechatronic systems.

Temperature is also a useful indicator of the condition of a machine. During continuous production, machine faults could cause a deviation in the temperature. Thermocouples, RTD's, optical pyrometers, and fiber-optic gauges are sensors for temperature measurement. Thermography is a technique where a thermal image of a component is obtained. In this process, an infrared camera is used to monitor the temperature patterns in turbines, bearings, piping, furnace linings, and pressure vessels. A thermal image is obtained on a screen that indicates any abnormal condition (like damaged insulation or localized temperature build-up in a bearing).

One factor which influences the cost of the manufacturing process is its tool wear. The increasing dullness of the cutting-tool edge during the cutting process increases the cutting force. In addition, wear in machine tools can provide information of the machine's existing condition. Monitoring the wear and using adaptive optimization methods can improve the manufacturing process. In automotive applications, broken piston rings or wear of the sliding members in contact with the cylinder can be detected. Direct measurement of wear in machine tools is done by incorporating an electrical sensor on the tool tip and observing the change in resistivity. Acoustic probes, imaging devices using position-sensing devices, and fiber-optic wear probes are used for off-line measurement.

FIGURE 1-10 SHADOW CYBERGLOVE

Photo courtesy of Jeremy Sutton-Hibbert/Alamy.

FIGURE 1-11 NEXAN ROBOT

From Mechanical Engineering Magazine, June 2008, Brian Mac Cleery and Nipun Mathur, "Right the First Time." Photo by Nexans.

1.5.2 Monitoring On-Line

The importance of lean production systems has created an opportunity for intelligent autonomous inspection, manufacturing, and decision-making systems that perform tasks without human intervention. Currently, quality is ensured in the product engineering cycle at two distinct levels.

- *At the product design stage:* To ensure that quality is designed into the product. Using the robust design method.
- *At the final inspection stage:* Using statistical process control methods.

Another level of quality assurance, *on-line quality monitoring*, complements robust design and statistical control methods. Continuous quality inspection of critical items in aerospace industry and silicon devices in microelectronic fabrication are done by on-line systems. 100% inspection ensures a quality standard for all products with no sampling error. By linking the process data and quality data, automatic fault correction is achieved. Quality monitoring provides the industrial plants with an ability to take quick corrective actions at the problem source.

Condition monitoring and fault diagnosis in modern manufacturing is also of great practical significance. These improve quality and productivity, and prevent damage to machinery. In a classical implementation of condition monitoring, sensors are deployed to monitor the condition of a system to detect abnormality. For example, the characteristics of frequency spectra originating from vibration in machine bearings can be used as an indicator of progressive bearing wear. Together with expert knowledge about the system, the observation of certain spectral components can be used to detect the onset of specific failure mechanisms. On-line monitoring devices have been available for many years, but they are still not widespread in industry. The main problem so far is the limited functionality and reliability of the devices, in particular when they face rapidly changing production conditions.

Significant progress in optimizing the manufacturing process has been achieved in recent years. Several relevant approaches include stereo matching, 3-D reconstruction, and use of neural networks. The Europe-based program on Intelligent Devices for the On-Line and Real-Time Monitoring, Diagnosis, and Control of Machining Processes (IDMAR) has made effort to connect scientists, machine tool builders, experts in signal processing, developers of monitoring devices and sensors, as well as end-users from the metal-cutting industry. This network helps the sector of European industry to achieve or retain global competitiveness by cutting costs, increasing product and process quality, and providing flexibility at the same time.

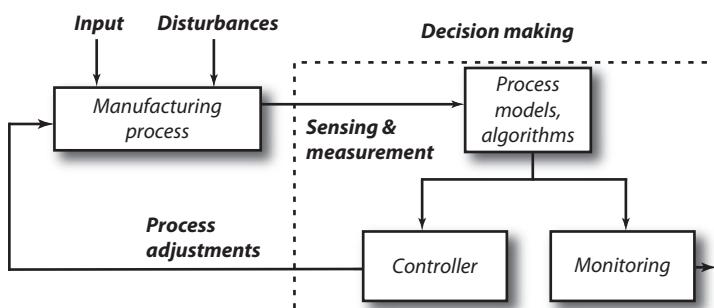
Evidence-Based Diagnostics In fields like healthcare, Internet-based systems are available to help doctors identify possible causes for patient symptoms. One such statistical diagnostic assistant, called “Isabel,” was developed by a father who sought to change the diagnostic system that affected the way his daughter (Isabel) was treated. This system is basically an intuitive system that takes advantage of all previous diagnoses and provides the statistically most likely disease (fault) and treatment (repair).

The application of a condition-based maintenance information system also is available in army and military applications. The system has the ability to integrate information from on-board sensors and diagnostic equipment to develop fleet-wide logistic and situational awareness, implementing a condition-based maintenance service that will enhance the operation and effectiveness of tactical and combat vehicles.

1.5.3 Model-Based Manufacturing

Model-based monitoring systems generally use a set of modeling equations and an estimation algorithm (such as a state observer) to estimate the signal important to the machine performance. In model-based monitoring, the purpose of the model is to represent the behavior of the structure—also sensed externally and recorded. Local sensors provide an output signal related to the measurement. The difference between the model output and the actual process output signals provides a concise mechanism for incorporating diagnostics, which is an attractive alternative to empirical *rule-based decision* systems. Figure 1-12 presents a generic diagram of an intelligent model-based manufacturing system.

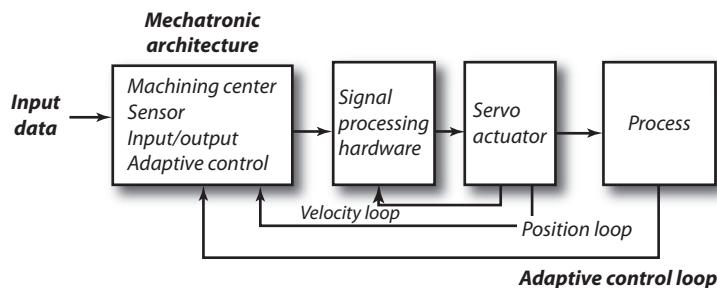
FIGURE 1-12 MODEL-BASED MONITORING SYSTEM



The diagram in Figure 1-12 also shows how the controller applies commands to the process such that various sensed values (related to the machine and/or the process performance) are maintained (or regulated) at desired values. Remote sensors may sense some of the diagnostic signals in difficult-to-access locations. In some cases, estimation algorithms are used based on the system structure and the signal of interest. Modeling procedures (some based on the previous knowledge) are used to produce simple, accurate models to improve estimation accuracy.

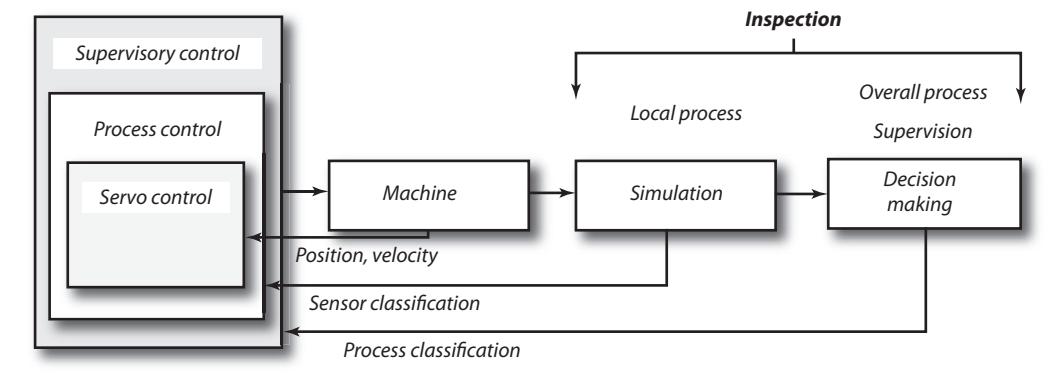
Mechatronics Systems with Open Architecture Process and machine-tool condition monitoring are the keys to an increasing degree of automation and, consequently, to an increasing productivity in manufacturing. One prerequisite for this functionality is the open interface in the NC-kernel. Today, controls with open NC-kernel interfaces are available on the market; however, these interfaces are vendor-specific solutions that do not allow the reuse of monitoring software in different controls. The development of modular, open architecture machine controllers, as shown on the next page in Figure 1-13, have provided improvements to the existing systems to overcome these limitations with vendor-neutral open real-time interfacing for the integration of monitoring functionality into the controller. This trend is also responsible for accelerating the use of intelligent sensors in manufacturing.

Sensor equipped intelligent control systems can be used to evaluate, to control the manufacturing process, and to provide a link to basic design. The multivariate environment of a manufacturing process generally does not produce a good analytical model of the process. However, additional information generally gets generated as a result of the introduction of manufacturing automation in a typical plant floor, and that data becomes available for modeling. Carefully collecting the data and using the knowledge base in a visual simulation environment makes it possible to integrate design,

FIGURE 1-13 MECHATRONIC SYSTEM WITH OPEN ARCHITECTURE PLATFORM

From Furness, 1996.

control, and inspection, as well as planning activities. Figure 1-14 shows a framework for integrating heterogeneous systems, which involves the position and velocity control of a machine tool, local inspection of a process, global inspection of the overall process, and finally, classification.

FIGURE 1-14 FRAMEWORK FOR INTEGRATING HETEROGENEOUS SYSTEMS

1.5.4 Supervisory Control Structure

In addition to influencing the way the products are designed, the developments in mechatronics have created opportunities in autonomous inspection and intelligent manufacturing. Figure 1-14 illustrates a hierarchical control structure where the controller elects position and velocity at the machine level, force and wear at the process level, and quality control issues (like dimension and roughness) at the product level.

This hierarchical control structure consists of servo, process, and supervisory controls.

- The lowest level is servo control, where the motion of the cutting tool relative to the work-piece (such as its position and velocity) is controlled. This involves cycle times of approximately 1 millisecond.
- At the process control level, process variables (such as cutting forces and tool wear) are controlled with typical cycle times of around 10 milliseconds. Control level strategies are

aimed at compensating for factors not explicitly considered in the design of the servo and process level controllers.

- The highest level is the supervisory level, which directly measures product-related variables (such as part dimension and surface roughness). The supervisory level also performs functions such as chatter detection and tool monitoring. The supervisory level operates at cycle times of approximately 1 second. Finally, all of this information can be used to achieve on-line optimization of the machining process at the shop floor and plant control level.

The trend in mechatronics is to optimize the overall manufacturing processes from product design to inspection by integrating all of the information into a common database. For example, knowledge of the parts geometry, as contained in the CAD system, can be used to determine the reference values of process variables. Information from various process-related sensors can be integrated to improve the reliability and quality of sensor information. This shared information (such as the data of the geometry of a part and the materials used from CAD/CAM database) can be used in selecting the optimum machining processes, tool selections, and finishing operations. Finally, all of this information can be used to achieve on-line optimization of the machining process.

Combined with automated monitoring of tool wear and quality inspection, the system helps to ensure efficient manufacturing processes and higher quality products. This will ultimately reduce total production cost and yield a better profit margin.

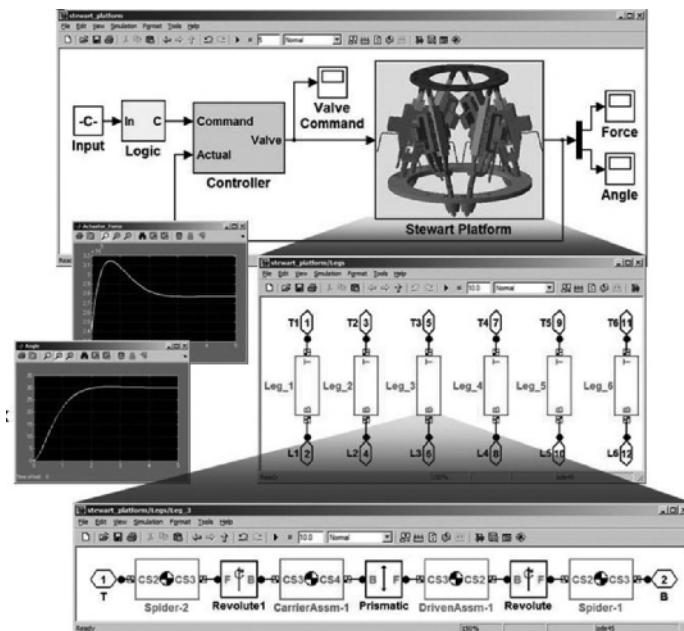
1.5.5 Open Architecture Matters with Mechatronic Models: Speed and Complexity

Mechatronics plays a role irrespective of the possibility of single or multiple microcontrollers handling machine tools or an automobile assembly line of multiple robots. Simulating such complex systems allows designers to develop a system without finalizing the hardware. The simulation procedure can be used as a “what if” scenario when the hardware doesn’t exist. There are two critical issues to consider: speed and complexity. Larger systems involve more detailed simulation and specific system requirements. Trade offs between simulation speed and the level of accuracy is necessary depending on the system resources available. The simulation becomes faster with faster processors, and the use of multicore systems help simulation (MacCleery and Mathur).

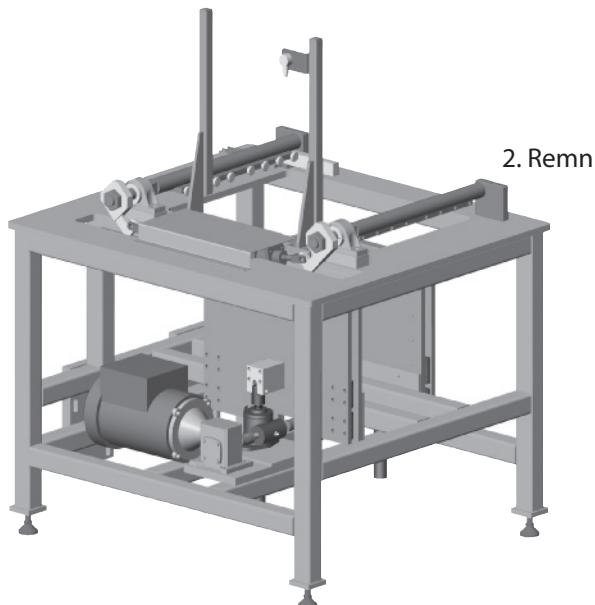
On the next page, Figure 1-15 shows an example of a platform which is used in production lines and in many other industrial applications. In this case, there are effectively two models: the simulated physical model and the application model. The physical model accounts for the physics-based simulated environment. The application model interacts with this environment to simulate the real-world application. Simulink and MATLAB are used as model-based development tools; so the application is a model.

The basic design represented in the physical world by computer-aided design and manufacturing tools (such as CATIA, Autodesk®, and SolidWorks) have advanced simulation tools, although they are oriented toward physical construction rather than process control integration (Figure 1-16 on the next page).

The simulation platform can examine stress under dynamic loading conditions. It also addresses nonlinear analysis (like deflection and impact) with flexible materials (such as foam, rubber, and plastic). In many cases, simulation and analysis of physical entities is useful in a design that doesn’t include a computer-based controller. The contribution by National Instruments facilitates a major integration which facilitates the design engineers to bring in mechanical elements (such as gears, cams, and actuators) while the programmers concentrate on the feedback and control algorithms that will handle the motors and actuators in the system. Linking various objects

FIGURE 1-15 SIMULINK MODEL OF A PLATFORM

From William Wong, Electronic Design, October 23, 2008 © Electronic Design, a Penton Media Publication.

FIGURE 1-16 ASSEMBLY LINE DESIGN USING CAD MODELS

From William Wong, Electronic Design, October 23, 2008 © Electronic Design, a Penton Media Publication.

together enables the models to interact. The provision of rendering permits visualization of the models in action. When creating large models, the modeling environment can demand significant amounts of computing power. The creation of large models can be a challenge to computing. At this stage, open architecture hosts can make a significant difference.

Several CAD and model-based design systems employ interface software that takes advantage of multiple cores. Exploiting a large number of cores and clustered systems has been a challenge in advanced software architectures. The major challenge is communication between cores. The basic requirement of mechatronic simulation is the time-synchronization between various objects in a distributed environment. Simulation in a multiple-core environment is again a challenge when the shared memory cannot handle the synchronization. Typically, there is an amount of limitation in the physical space. A robotic line-assembly simulation can perform well within its region, but it will have limited capability if it has to interact with other cells. Graphical model-based programming can assist in linking multiple cells.

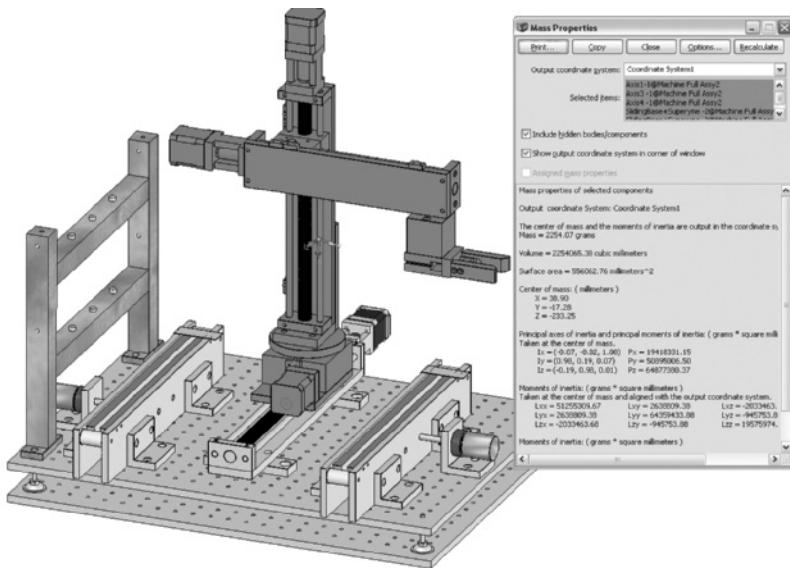
1.5.6 Interactive Modeling

The key aspect of the graphical environments is that the visual representation of system partitioning and interaction lends itself to mechatronic applications. They also reduce system complexity from a developer's standpoint, allowing concentration on the application details. For example, a simulation tool (such as Simscape™) is used as a declarative language that defines implicit relationships between components versus the explicit programming specifications for languages like C and C++ as well as graphical dataflow languages such as LabVIEW™. Simscape targets co-simulation where programming and CAD intersect. This multidomain tool ties together the electronics, mechanical drive elements, and mechanics and hydraulics tools. For example, the Stewart platform simulation discussed earlier can incorporate electrical, hydraulic, mechanical, and signal flow support in addition to software control of the system.

By reducing the amount of expertise required for developing mechatronic applications, developers can spend time and effort on other areas where they do have expertise. Likewise, having a model environment permits a better exchange of ideas and products. The difference these days is that the detail within the models being exchanged as objects within a *mechatronic* application have become more advanced. What used to be just dimensions is now something that can be used within a simulation complete with programmable feedback (and even application interfaces when a model includes application code). Also available is a design verifier, where assertion blocks are able to be included within a model so the system can determine whether an object's use within a system is correct (Figure 1-17 on the next page).

Interactive modeling is crucial to the design process, and it can occur in a mixed environment where real and virtual objects are combined. A real robotic arm may be coupled with a virtual assembly line, for example, if the current task is to determine if the hand on the robotic arm can reorient an object. The robotic arm might be involved in the laser welding of end plates. The key is getting the virtual objects and their control counterparts to interact with the real objects with code that's running on remote devices. The electromechanical control systems once designed for the factory floor have become ubiquitous. For example, a designer may answer a problem concerning vibration by adding a stiffener. In an integrated mechatronic process, however, that small mechanical change may increase the mass of the part; it also may affect how fast the control system ramps up motor speed and how long the part holds in place before returning.

Many top mechatronic performers also use software that routes, tracks, and shares work. Most common are workflow tools, which automatically route work packages, and notify the right people of deadlines and/or changes. Many companies make use of product data management tools to manage multidisciplinary bills of materials.

FIGURE 1-17 SIMULATION VERIFICATION OF A TORQUE LOAD

Images are provided Courtesy of National Instruments and SolidWorks Corporation.

1.5.7 Right First Time—Virtual Machine Prototyping

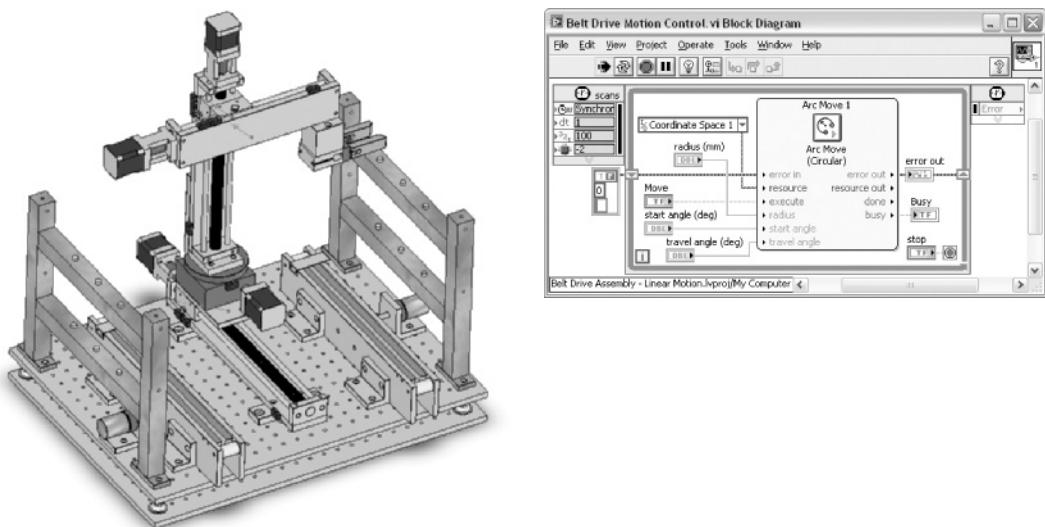
The hardware-in-the-loop facilitates the replacement of conventional mechanical motion-control devices with digital devices. Mechanical systems are increasingly controlled by sophisticated electric motor drives that get their digital intelligence from software running on an embedded processor. Getting electromechanical designs right requires multidisciplinary teamwork and superb communication among team members. A decision like choosing the characteristics of a lead screw actuator has a ripple effect throughout the design and can impact the performance of other systems. To help facilitate a more integrated design process, we need to add motion-simulation capabilities to CAD environments to create a more unified mechatronics workflow.

Integrating motion simulation with CAD simplifies design because the simulation uses information that already exists in the CAD model, such as assembly mates, couplings, and material mass properties. Adding a high-level function block language for programming the motion profiles provides easier access to control those assemblies.

This concept is known as *virtual machine prototyping*. It brings together motion-control software and simulation tools to create a virtual model of an electromechanical machine in operation. Virtual prototyping helps designers reduce risk by locating system-level problems, finding interdependencies, and evaluating performance trade offs (Figure 1-18).

1.5.8 Evaluating Trade Off

Simulations enable everyone to work on development before the first prototype is completed. Engineers can use force and torque data from simulations for stress and strain analysis to validate whether mechanical components are stiff enough to handle the load during operation. They also can

FIGURE 1-18 EVALUATING TRADE OFFS IN A CAD ENVIRONMENT

Images are provided Courtesy of National Instruments and SolidWorks Corporation.

validate the entire operating cycle for the machine by driving the simulation with control-system logic and timing. They can calculate a realistic estimate for cycle time performance (which is typically the top performance indicator for a machine design) and compare force and torque data with the realistic limitations of transmission components and motors. This information can help identify flaws and drive design iterations from within the CAD environment. Simulations also simplify evaluating engineering trade offs between different conceptual designs.

For example, would a SCARA robot be preferable to the four-axis Cartesian Gantry robot system? Simulations are faster and can be run again whenever you make design changes. Consider an analysis of the torque load for the bottom lead screw actuator. Using simulation software, you can find the mass of all the components mounted on the lead screw, determine the resulting center of mass by creating a reference coordinate system located at the center of the lead screw table, and calculate the mass properties with respect to that coordinate system. With this information, you can calculate the static torque on the lead screw due to gravity caused by the overhanging load. If you violate the limits specified by the manufacturer, the mechanical transmission parts may not last for their rated life cycles.

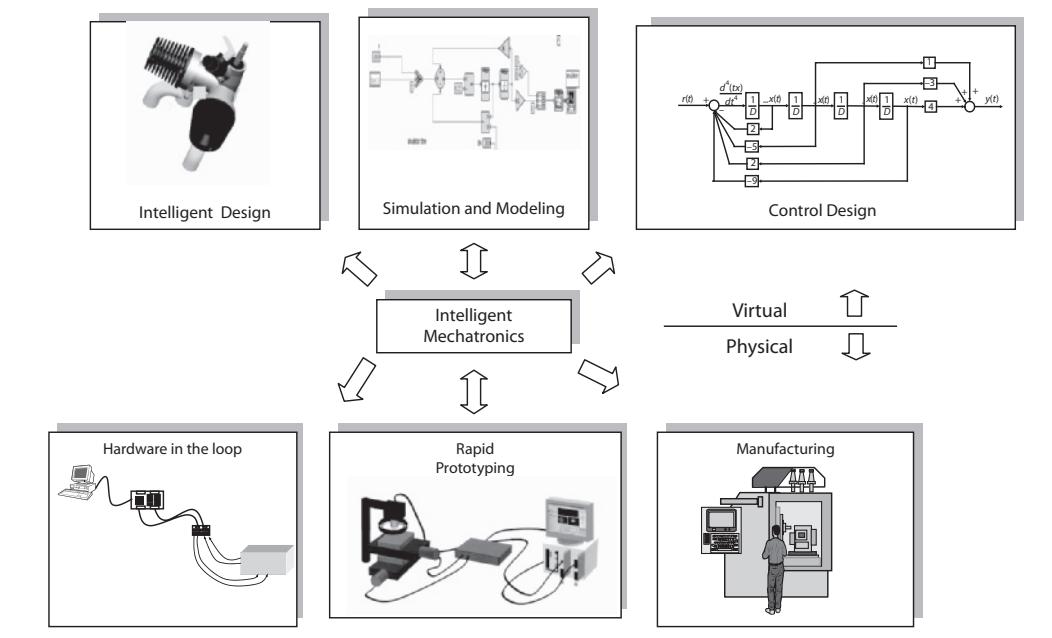
Evaluating the *dynamic torque* induced by motion is important because it tends to be much larger than the static torque load. Realistic motion profiles will help us to simulate inverse vehicle dynamics. This can provide more accurate torque and velocity requirements based on the motion profiles and the mass, friction, and gear ratio properties of the transmission.

At times, the designer may consider compliance issues when he designs the assemblies, but incorrect assumptions about operational forces and torques may lead to problems. In mechatronics systems, compliance issues take two main forms: rotational compliance and linear compliance. *Rotational compliance* is affected by the flexibility of mechanical transmission components, such as the connecting rods and couplings. Each rotating part acts like a spring with a particular stiffness,

and the entire drive train acts like a series of springs connected in series. *Linear compliance* problems are caused by the flexibility of mechanical assemblies, such as the gripper arm in a pick-and-place machine. The length of the moment arm, the weight of the payload, and the speed of the motion profiles all play a role.

Another phenomenon is *backlash*, which is caused by the clearance between mating components (gear teeth) and appears during a change of direction. Compliance and backlash issues can make the proportional-integral-derivative feedback devices difficult or impossible to tune, causing the system to literally hum during operation. If the system is de-tuned by reducing the PID gains to try to avoid the problem, the cycle time performance is affected.

FIGURE 1-19 PHOTOGRAPH OF INTELLIGENT MANUFACTURING SYSTEM



1.5.9 Embedded Sensors and Actuators

The advances in MEMS and wireless, information, and other enabling technologies are leading to new sensor system functionality and allows access to more accurate sensing. Smart sensor-on-a-chip concepts include on-board calibration and temperature compensation, self-test capabilities, embedded software for data analysis, and a wireless communication interface to provide a useful output signal when it is appropriate to act on the sensed data. A smart sensor system has the capability to measure data for the presence of a biological or chemical agent and to process the data to evaluate that agent's concentration.

In conjunction with other on-chip software, a control algorithm consists of a model that is updated with sensed data from multiple sensors on that same chip. In addition, through wireless data infusion from neighboring smart sensors, the smart sensor can generate an appropriate output to trigger a variety of actions. Several networked smart sensors share information. Should one of

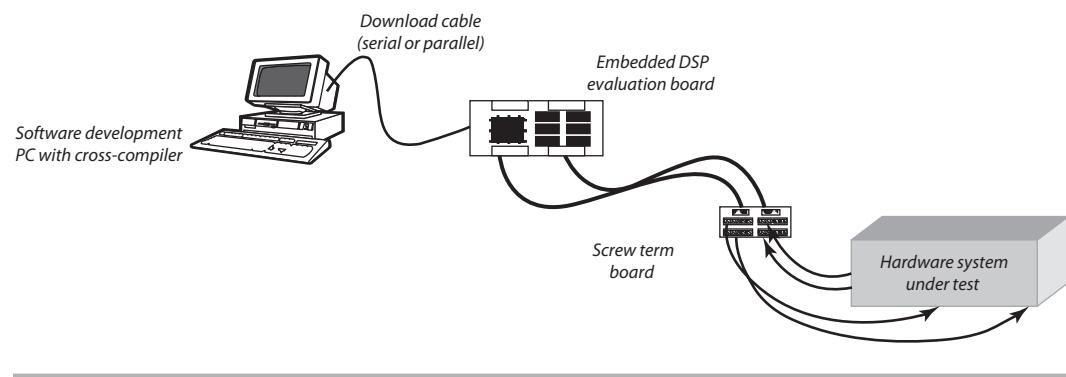
the network components fail, they have built-in mechanisms for system reconfiguration. Examples of this technology are piezoelectric films embedded in the work holder for precise motion control or eddy current probes mounted in the tool holder to monitor cutting-tool wear. A MEMS linear motor can be used to control thermal deformation errors with the level of precision of nanometers. With the increasing bandwidth of digital electronics and the greatly increasing application of Internet communication, the joint location of manufacturing shop floors and process monitoring/control systems is no longer a must.

1.5.10 Rapid Prototyping of a Mechatronic Product

Rapid prototyping and hardware-in-the-loop simulation are integral parts of today's product development process. Hardware-in-the-loop simulation testing provides the designer with reassurance that any assumptions made on the plant model were correct. PC-based integration of systems benefits from various software packages that often use graphical programming to create virtual instrumentation. Hardware-in-the-loop simulation is also a cost-effective method to perform system tests in a virtual environment. It demonstrates a level of interaction with the modeling of a system that is not possible when code is directly ported to the final target platform.

Mathematical models replace most of the components of the system environment when the components to be tested are inserted into the closed loop. If any assumptions were incorrect, the designer does have the opportunity to continue the optimization of the design before committing to the real-target hardware platform. There are two methods currently used to accomplish hardware-in-the-loop simulation testing. One method utilizes the virtual-instrumentation-based user interface coupled with standard data acquisition and control interface. The actual plant environment is used in place of the plant simulation model, and actual sensors and actuators are connected between the plant and the interface. Figure 1-20 shows a typical configuration for this type of hardware-in-the-loop simulation.

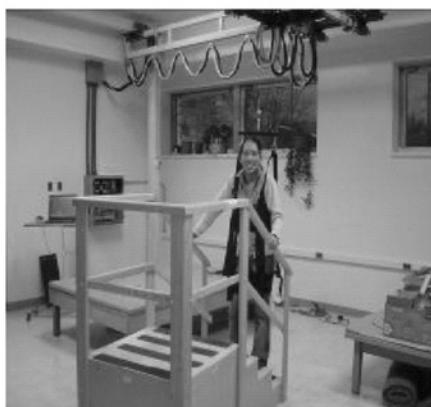
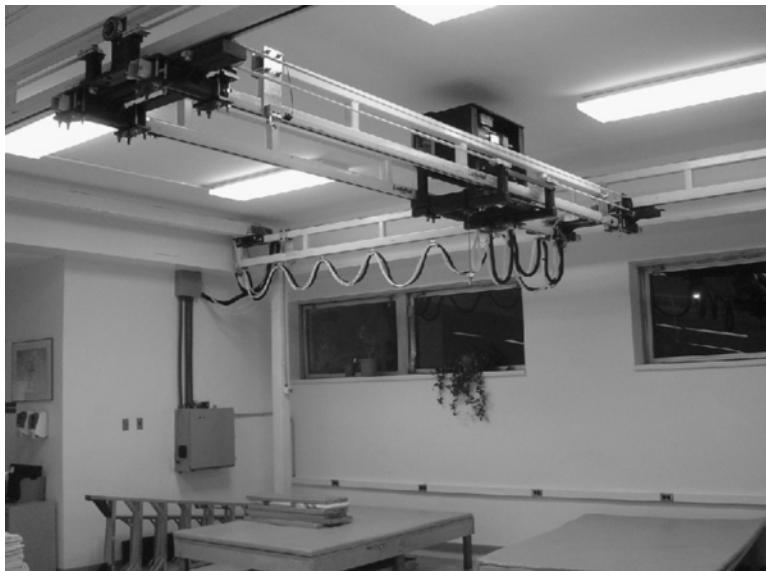
FIGURE 1-20 TYPICAL PC-BASED HARDWARE-IN-THE-LOOP SIMULATION



Another method for accomplishing hardware-in-the-loop testing involves cross-compiling the control algorithm to target an embedded real-time processor platform. The embedded processor platform is a digital signal processor with I/O that is customized for embedded system products. The cross-compiled code is then downloaded to the embedded processor, sensors are connected to the inputs of the embedded processor board, and actuators are connected to the outputs of the embedded processor board.

Mechatronically Designed Ambulatory Rehabilitation Walker The rehabilitation walker device (as shown in Figure 1-21) is an apparatus developed with the intent of aiding in the rehabilitation of hospital patients learning to walk again. This apparatus and control system are of industrial quality and would be reproducible in its entirety using off the shelf parts. The idea behind the rehabilitation walker is that it will relieve a certain percentage of body weight by carrying the patient in a harness which is attached to a hoist. The hoist is actively controlled using feedback from strain-gauge sensors. As the patient walks around within the confines of the

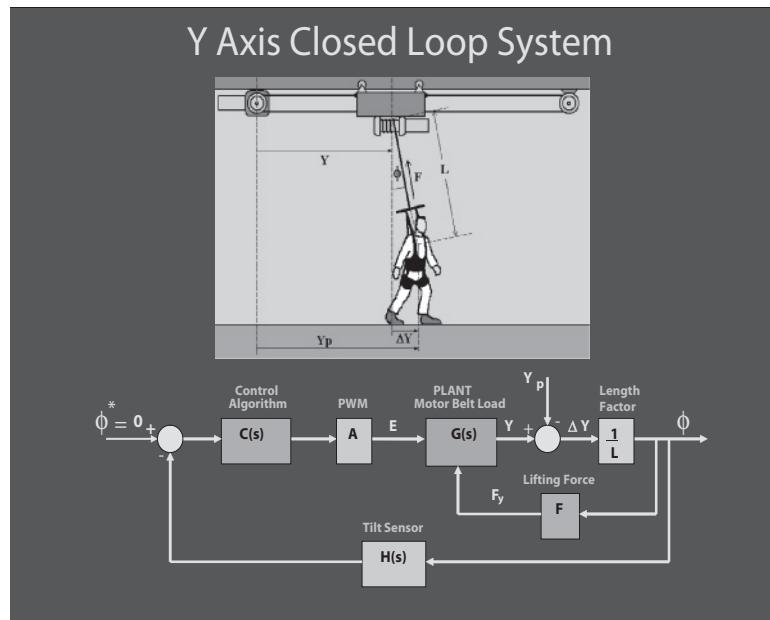
FIGURE 1-21 MECHATRONIC APPLICATION FOR REHABILITATION EQUIPMENT (COPYRIGHT US PATENT 7,462,138B2, SHETTY, FAST AND CAMPANA)



Shetty and Fast.

room-sized gantry, the hoist will follow the patient around. The overhead gantry is motorized in the x and y directions (Figure 1-22). The closed-loop motor control reacts to feedback from multi-axis tilt sensors on the hoist line. If ever the patient were to fall, the hoist system would react and remove the full load of the patient's weight. The base of the control system consists of a National Instruments Compact Reconfigurable Input Output Programmable Automation Controller (CRIo). The CRIo system is based on a Field Programmable Gate Arrays (FPGA) backplane and a real-time controller.

FIGURE 1-22 EXAMPLE OF MONITORING OF Y-AXIS IN THE REHABILITATION DEVICE

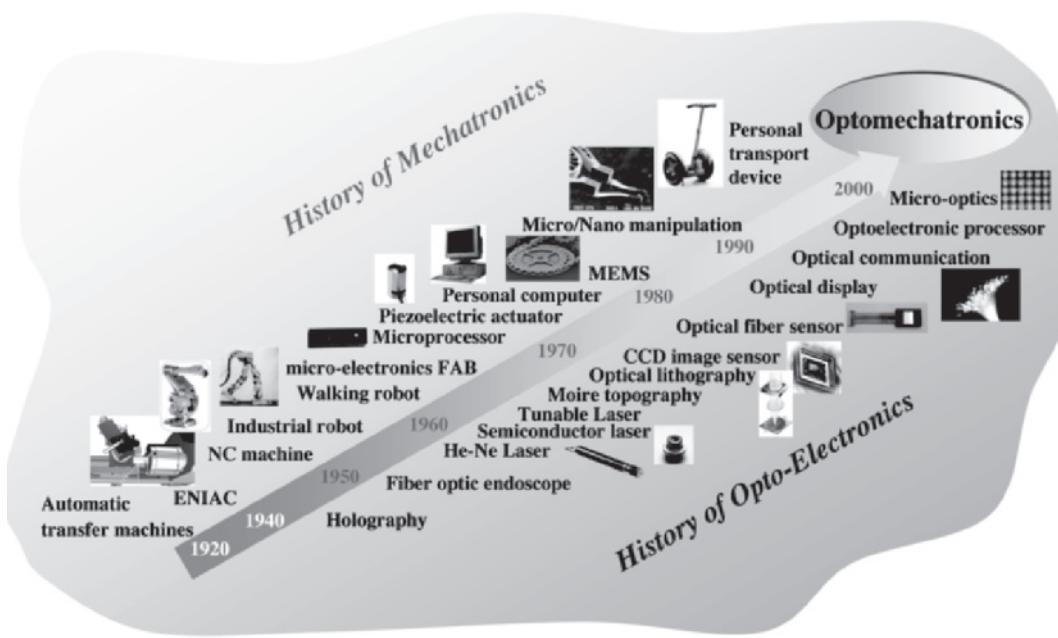


Shetty and Bravo, University of Hartford.

The backplane accepts modules which perform various I/O functions. The modules are chosen to interact with the rehabilitation walker sensors as well as handle the motor-drive output signals. The motors are driven by industrial amplifiers, while position is tracked via quadrature encoder feedback.

1.5.11 Optomechatronics

In recent years optical technology has been increasingly incorporated into mechatronic systems, resulting in a greater number of smart products. Optically integrated technology provides enhanced characteristics. On the next page Figure 1-23 shows the development of mechatronic technology in the upper line above the arrow and that of optical engineering in the lower line. With an array of choices available to measure critical dimensions, non-contact techniques from vision to high-tech lasers are increasingly offered to inspection as well as material processing. Three-dimensional, five-axis laser processing has become attractive due to advances in control systems and programming.

FIGURE 1-23 HISTORY OF OPTOMECHATRONICS

IEEE Transactions on Industrial Electronics, 52.4, © 2005, IEEE.

1.5.12 E-Manufacturing

Web-enabled monitoring is the fastest way to bring your real-time data onto the Web to provide real-time data from a factory floor line directly onto the Web. A Web-enabled platform is an integrated, visual environment that supports real-time Information systems and allows flexible monitoring and analyzing. Remote monitoring device interface and system technologies need to be developed based on a generic equipment model. The major purpose is to minimize the need for a struggle with distributed application development and deployment issues, and to allow industry engineers to focus on application functionality instead. The platform contains all the information related to monitoring

- The number of machines, devices and installation.
- The data server.
- The application server.
- The web server.
- Web-browsers.

All of the data collected from the devices and machines will be stored in databases, which can be integrated with different systems.

E-manufacturing is a methodology system that enables the manufacturing operations to successfully integrate with the functional objectives of an enterprise through the use of the Internet, with tether-free (i.e., wireless, web, etc.) and predictive technologies. E-manufacturing includes the ability to monitor the plant floor assets, predict the variation and performance loss to dynamically reschedule production and maintenance operations, and to synchronize related and consequent actions to achieve a complete integration between manufacturing systems and upper-level enterprise applications. Rockwell Automation Annual Report outlines a statement of competencies that are required of world class companies. These are design, operate, maintain and synchronize. E-manufacturing should include intelligent maintenance and performance assessment systems to provide reliability, dependability, and minimum downtime, allowing equipment to run smoothly at their highest performance.

FIGURE 1-24 (A) OPTICALLY IGNITED MECHATRONIC WEAPON SYSTEM (B) WELDING SYSTEM WITH MONITORING AND CONTROL

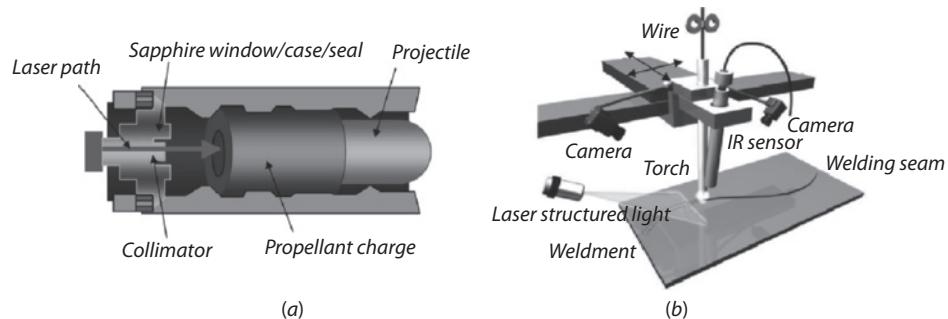
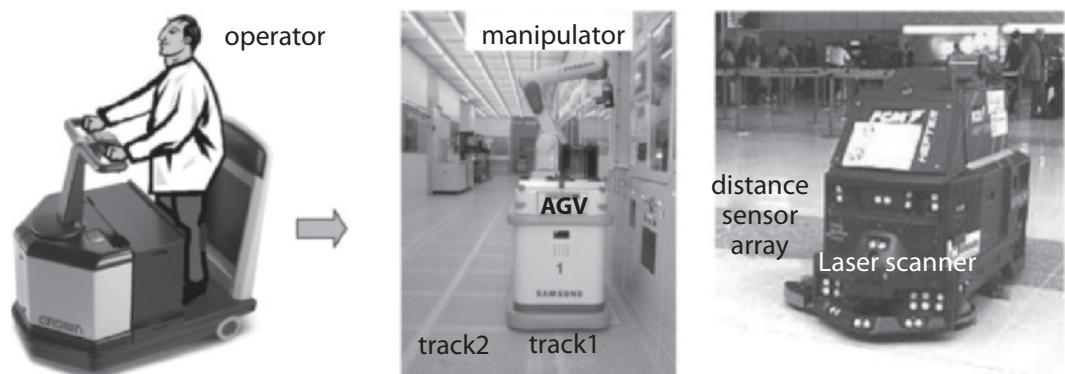


FIGURE 1-25 (A) HUMAN GUIDED VEHICLE (B) AUTOMATICALLY GUIDED ROBOT (C) MOBILE ROBOT FOR CLEANING

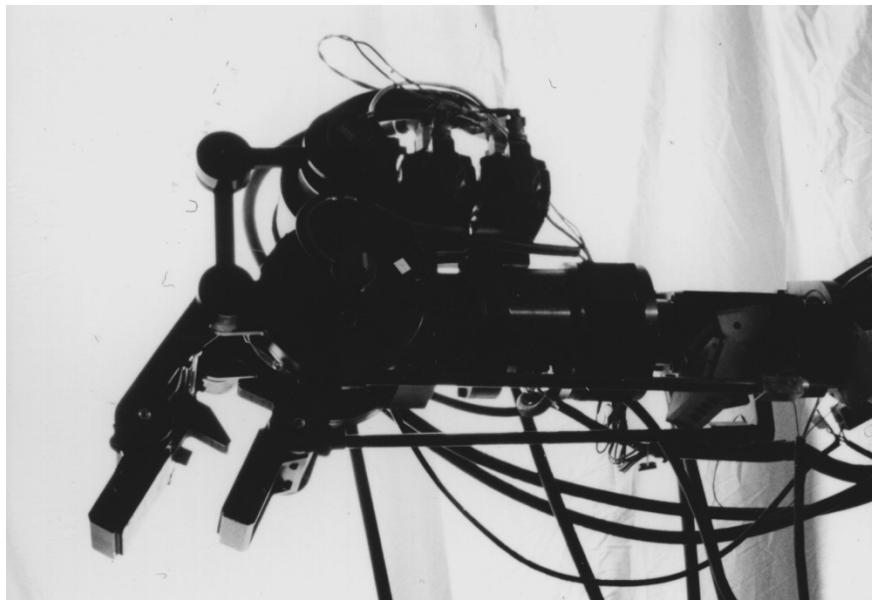


IEEE Transactions on Industrial Electronics, 52.4, © 2005, IEEE.

1.5.13 Mechatronic Systems in Use

Examples of mechatronic systems for industrial use are found in many areas. Mechatronic monitoring systems have been applied to products such as aircraft, machine tools, and automobiles. These systems are designed to measure plant parameters (such as compliance and inertia), plant states (such as current and velocity), and production states (such as force and wear). Figure 1-26 illustrates a recent application of mechatronics in a six degree-of-freedom hydraulic extender used for loading and unloading aircraft.

FIGURE 1-26 EXPERIMENTAL SIX-DEGREES-OF-FREEDOM HYDRAULIC EXTENDER FOR LOADING AND UNLOADING AIRCRAFT



Courtesy Professor Kazerooni, University of California, Berkeley.

Noteworthy Mechatronic Applications

Automotive Industry:

- Vehicle diagnostics and health monitoring. Various sensors are used to detect the environment or road conditions; Sensors to monitor engine coolant, temperature and quality; Engine oil pressure, level, and quality; tire pressure; brake pressure.
- Pressure, temperature sensing in various engine and power train locations Manifold control with pressure sensors; exhaust gas analysis and control; Crankshaft positioning; Fuel pump pressure and fuel injection control; Transmission force and pressure control.
- Airbag safety deployment system. Micro-accelerometers and inertia sensors mounted on the chassis of the car measures car deceleration in x or y directions can assist in airbag deployment.

- Antilock brake system, cruise control. Position sensors to facilitate antilock braking system; Displacement and position sensors in suspension systems.
- Seat control for comfort and convenience. Displacement sensors and micro actuators for seat control; Sensors for air quality, temperature and humidity, Sensors for defogging of windshields.

Health Care Industry:

- Medical diagnostic systems, non-invasive probes such as ultrasonic probe. Disposable blood pressure transducer; Intrauterine pressure monitor during child delivery.
- Pressure sensors in several diagnostic probes. Systems to control the intravenous fluids and drug flow; Catheter tip pressure sensor.
- Endoscopic and orthopedic surgery. Angioplasty pressure sensor; Respirators; Lung capacity meters.
- Other products such as Kidney dialysis equipment; MRI equipment.

Aerospace Industry:

- Landing gear systems; Cockpit instrumentation; Pressure sensors for oil, fuel, transmission; Air speed monitor; Altitude determination and control systems.
- Fuel efficiency and safety systems; Propulsion control with pressure sensors; Chemical leak detectors; Thermal monitoring and control systems.
- Inertial guidance systems; Accelerometers; Fiber-optic gyroscopes for guidance and monitoring.
- Communication and radar systems; High bandwidth, low-resistance radio frequency switches; Optical instrumentation using laser communications.

Consumer Industry:

- Consumer products such as auto focus camera, video, and CD players; Consumer electronic products; User-friendly washing machines with water level controls, dish washers, and other home appliances.
- Video game entertainment systems; Virtual instrumentation in home entertainment.
- Home support systems; Garage door opener; Sensors with heating, ventilation, and air-conditioning system; Home security systems.

Industrial Systems and Products:

- Monitoring and control of the manufacturing process; CNC machine tools; Advanced high speed machining and quality monitoring; Intelligent machining and on-line quality check; Digital torque wrenches, variable speed drilling and other hand tools.
- Rapid prototyping; Manufacturing cost saving by rapid creation of models done by CAD/CAM integration and rapid prototyping equipment.
- Autonomous production cells with image-based object recognition; Flexible manufacturing and other factory automation systems.

- Specialized manufacturing process such as the use of welding robots; Procedure for automatically programming and controlling a robot from CAD data; Robotics in nuclear inspection and space applications.
- Automatic guided vehicles, space application; Use of automated navigation system for NASA projects; Use of automated systems in under water monitoring and control.

Other Applications:

- Telecommunications.
- Biorobotics, which utilize the biofunctions for applications in environmental control.
- Magnetically levitated vehicles.
- Scanners and copying machines and other office products.

Numerical computation, simulation, computer-aided design, and experimental validation are important technologies which must be considered when evaluating the feasibility of complex mechatronic systems. Other technologies include artificial intelligence, expert systems, fuzzy logic, neural networks, and nano-technology. The usefulness of these technologies is expected to be at the higher levels of the control hierarchy in machining processes.

EXAMPLE 1.1 Step-by-Step Mechatronic Design

A simple mechatronic system consisting of a permanent magnet (PM), DC gear motor, and a Hall effect sensor is used for demonstrating how the contents of various chapters in this book are used in the design of a mechatronic system. The intent is to understand the approach that can be followed while designing a mechatronic system. However, it is also important to know that design approach will differ based on the problem. Figure 1-27 shows the components of a mechatronic system in general for the position control of the PM-DC gear motor. Table 1-3 gives an insight of how these components are covered in this book to fulfill the task of designing the simple mechatronic system.

FIGURE 1-27 COMPONENTS OF MECHATRONIC SYSTEM OF A DC MOTOR POSITION

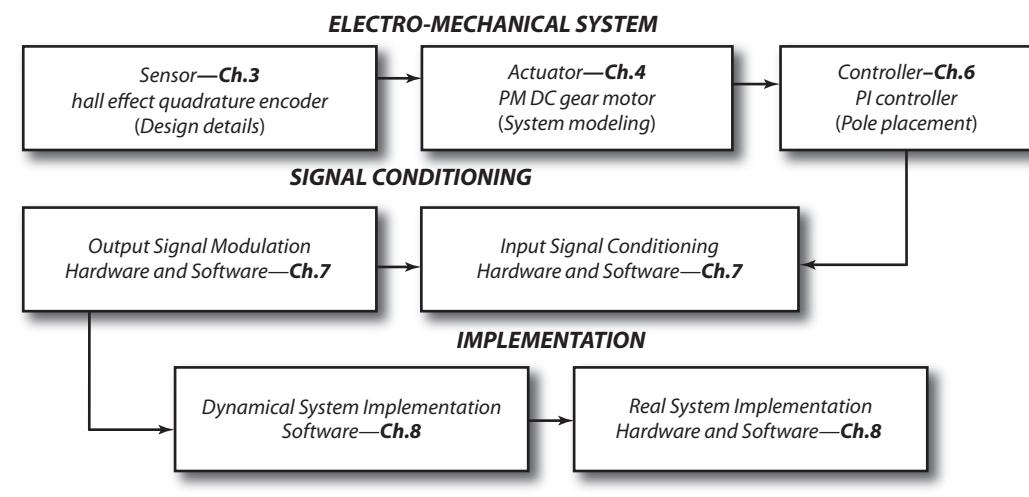


TABLE 1-3 CHAPTER-BREAKDOWN OF COMPONENT DISTRIBUTION FOR THE DC MOTOR EXAMPLE

Mechatronic System Components for DC Motor Example	Chapter
Theory and design details of Hall effect sensor	3
Mathematical modeling of the PM DC motor and the system as a whole	4
Design of a PI controller for accurate positioning of the motor shaft based on the required performance characteristics	6
Hall effect sensor application	7
Modulation of the PI output data both at hardware and software level	7
Implementation of the dynamical system and real system	8

1.6 Summary

Successful mechatronics design can lead to products that are extremely attractive to the consumer in terms of quality and cost effectiveness. Conversely, products designed in the more traditional sequential manner do not possess optimum design capabilities and lack consumer appeal. A major factor in the development of an intelligent and flexible mechatronic system is the concurrent use of automated diagnostic systems using sensors to handle machinery-maintenance and process-control operations. Sensor-fused intelligent control systems can be used to evaluate and control the manufacturing process, and to provide a link to basic design. Increasing demands on the productivity of machine tools and their growing technological complexity call for improved methods in future product development processes. Mechatronics is also influenced by intelligent devices for the on-line and real-time monitoring, which includes diagnosis and control of processes.

REFERENCES

- Aberdeen Group., “System design: New product development for mechatronics.” Boston, MA, January 2008 and *NASA Tech Briefs*, May 2009. (www.aberdeen.com)
- Ali, A., Chen, Z., and Lee, J., “Web-enabled platform for distributed and dynamic decision making systems.” *International Journal of Advanced Manufacturing Technology*, August 2007.
- Brian Mac Cleery and Nipun Mathur. “Right the first time” *Mechanical Engineering*, June 2008.
- Bedini, R., Tani, Giovanni, et. al., “From traditional to virtual design of machine tools, a long way to go- Problem identification and validation.” *Presented at the International Mechanical Engineers Conference (IMECE)*, November 2006.
- Pavel, R., Cummings, M., and Deshpande, A., “Smart Machining Platform Initiative.” *Manufacturing Engineering*, 2008.
- Cho Hyungsuck. “Optomechatronics—Fusion of Optical and Mechatronic Engineering”. *Taylor and Francis & CRC Press*, 2006.
- Fan, H. and Wu, S., “Case Studies on Modeling Manufacturing Processes Using Artificial Neural Networks,” *Neural Networks in Manufacturing and Robotics*, ASME, PED-Vol., 57, 1992.
- Furness, R., “Supervisory Control of the Drilling Process,” *Ph.D. Dissertation, Department of Mechanical Engineering and Applied Mechanics*, University of Michigan, Ann Arbor, MI, 1992.
- Gopel, W., Hesse, J., and Zemel, J.N. “*Sensors, A Comprehensive Survey*, (Vol.1) VCH Publishers Inc, 1989.
- Jay Lee. “E-manufacturing—fundamental, tools, and transformation.” *Robotics and Computer Integrated Manufacturing*, 2003.

- Landers, R.G. and Ulsoy, A.G., “A supervisory machining control example.” *Recent Advances in Mechatronics*, ICRAM 1995, Turkey, 1995.
- Nise, Norman S., *Control Systems Engineering*. Benjamin/Cummings Publishing Co., Redwood City, California, 1992.
- Ryoji Ohba., *Intelligent Sensor Technology*. John Wiley & Sons, 1992.
- Philpott, M.L., Mitchell, S.E., Tobolski, J.F., and Green, P.A., “In-process surface form and roughness measurement of machined sculptured surfaces,” *Manufacturing Science and Engineering*, Vol. 1, ASME, PED-Vol. 68-1, 1994.
- Rockwell Automation e-Manufacturing Industry Road Map*. <http://www.rockwellautomation.com>
- Stein, J. L. and Huh, Kunsoo, “A design procedure for model based monitoring systems: cutting force estimation as a case study.” *Control of Manufacturing Processes*, ASME, DSC, Vol 28/PED-Vol 52, 1991.
- Stein, J. L. and Tseng, Y. T., “Strategies for automating the modeling process.” *ASME Symposium For Automated Modeling*, ASME, New York, 1991.
- Shetty, D. and Neault, H., “Method and Apparatus for Surface Roughness Measurement Using Laser Diffraction Pattern.” United States Patent, Patent Number: 5,189,490, 1993.
- NI LabVIEW-SolidWorks Mechatronics Toolkit, <http://www.ni.com/mechatronics>.
- Shetty, D., *Design For Product Success* Society of Manufacturing Engineers, Dearborn, Michigan, 2002.
- Sze, S.M., *Semiconductor Sensors*. John Wiley & Sons, Inc., 1994.
- Tarbox, G.H. and Gerhardt, L., “Evaluation of a hierarchical architecture for an automated inspection system.” *Proceedings of Manufacturing International*, ASME, Vol. V, pp. 121–126, 1990.
- Ulsoy, A.G. and Koren, Y., “Control of Machining Processes,” *Journal of Dynamic Systems, Measurement, and Control*. Vol. 115, pp. 301–308, 1993.
- Van de Vegte, John., *Feedback Control Systems*, Second Edition, Prentice Hall, Englewood Cliffs, New Jersey, 1990.
- William Wong. “Muticore matters with mechatronic models,” *Electronic Design*, October 23, 2008.

PROBLEMS

- 1.1. What is mechatronics? How is it different from the traditional approach of designing? State the advantage of using the mechatronic design methodology?
- 1.2. What is the function of a sensor and a actuator in a mechatronic system? List different types of actuators with at least two examples of each type.
- 1.3. Understand the purpose of the following mechatronic system and recommend appropriate sensor and actuator to carry out the specified task.

a. Temperature Control System

Purpose: To maintain the temperature of a confined space at the specified temperature.

(Hint: Decide how to sense the temperature. Decide how to increase or decrease temperature.)

b. Anti-Lock Braking System

Purpose: To prevent wheels from locking up by automatically modulating the brake pressure during an emergency stop.

(Hint: Decide how to sense that the wheels are locked, i.e., the wheels are not rolling. Decide how to apply or release brakes.)

CHAPTER 2

MODELING AND SIMULATION OF PHYSICAL SYSTEMS

- 
- 2.1 Operator Notation and Transfer Functions
 - 2.2 Block Diagrams, Manipulations, and Simulation
 - 2.2.1 Block Diagrams—Introduction
 - 2.2.2 Block Diagrams—Manipulations
 - 2.2.3 Simulation
 - 2.3 Block Diagram Modeling—Direct Method
 - 2.3.1 Transfer Function (or ODE) Conversion to Block Diagram Model
 - 2.3.2 Conversion of Mechanical Illustration to Block Diagram Models
 - 2.4 Block Diagram Modeling—Analogy Approach
 - 2.4.1 Potential and Flow Variables, PV and FV
 - 2.4.2 Impedance Diagrams
 - 2.4.3 Modified Analogy Approach
 - 2.5 Electrical Systems
 - 2.6 Mechanical Translational Systems
 - 2.7 Mechanical Rotational Systems
 - 2.8 Electrical-Mechanical Coupling
 - 2.8.1 Lorentz's Law—Electrical to Mechanical Coupling
 - 2.8.2 Faraday's Law—Mechanical to Electrical Coupling
 - 2.8.3 Electrical-Mechanical Coupling Linear Relationships
 - 2.9 Fluid Systems
 - 2.10 Summary
 - References
 - Problems
 - Appendix to Chapter 2

Component modeling, which is the derivation of mathematical equations suitable for computer simulation, plays a critical role during the design stages of a mechatronic system. For all but the simplest systems, the performance aspects of components (such as sensors, actuators, and mechanical geometry) and their effect on system performance can only be evaluated by simulation.

Any modeling task requires the formulation of mathematical models suitable for computer simulation or solution—the terms are analogous. This chapter presents one method, the *analogy approach*, which can be used for such modeling tasks. It was developed by electrical engineers to model mechanical, thermal, and fluid systems for simulation on analog computers. Because the analog computer was used for the simulation environment, it was fitting that models were constructed using standard electrical elements, such as resistors, capacitors, and inductors.

Analog computer simulation environments have two attractive features: precise integration and real-time operation, but they are limited in their ability to represent and solve complex nonlinear equations. For example, a nonlinear table function cannot be incorporated using the standard electrical elements, instead the function must be approximated by a truncated power series and represented as a polynomial. Being a sequence of multiplications and additions, the polynomial then can be represented using the standard electrical elements. If one table entry is modified, the approximating polynomial must be completely regenerated—a time consuming process. Today the digital computer is used extensively for simulation. Instead of using standard electrical elements and circuits, digital computer models are constructed using block elements and represented as *block diagrams*. Block diagrams are much more powerful, flexible, and intuitive than circuit models.

In this chapter, we will present two approaches for developing block diagram models from system illustrations: (1) the direct method and (2) the analogy method with slight modifications.

2.1 Operator Notation and Transfer Functions

For ease in writing linear lumped-parameter differential equations, the D operator is introduced. Any linear lumped differential equation can be converted to *operator form* by simply substituting the operation using differentiation or integration with the appropriate operator. Table 2-1 summarizes the operators for differentiation and integration and presents several examples.

TABLE 2-1 D OPERATOR FOR DIFFERENTIATION AND INTEGRATION

Type	Operation	Operator	Operator Form Examples
Continuous	Differentiation	$D \equiv \frac{d(\cdot)}{dt}$	$\dot{x}(t) - 3\ddot{x}(t) + x(t) = \dot{r}(t) - 1$ $\Rightarrow D^2x(t) - 3Dx(t) + x(t) = Dr(t) - 1$
Continuous	Integration	$\frac{1}{D} \equiv \int_{t_0}^t (\cdot) d\tau$	$\dot{x}(t) + x(t) - \int x(\tau) d\tau + r(t) = 0$ $\Rightarrow Dx(t) + x(t) - \frac{1}{D}x(t) + r(t) = 0$

Oftentimes, we wish to do more than just write a differential equation in a concise form. We want to solve it and analyze its behavior. The Laplace transform is used to represent a continuous time domain system, $f(t)$, using a continuous sum of complex exponential functions of the form e^{st} where s is a complex variable defined as $s \equiv \sigma + j\omega$. The complex domain (or s plane as it's often called) is just a plane with a rectangular x - y coordinate system where σ is the real part and ω is the imaginary part. Applying the Laplace transform to a time-domain differentiation operation results in a frequency-domain multiplication operation where s is the operator. The Laplace s operator is identical to the D operator previously introduced, except when a differential equation is written in s -operator or Laplace format, it is no longer in the time domain but rather in the frequency (complex variable) domain.

The cause–effect relationship for many systems can be approximated by a linear ordinary differential equation. For example, consider the following second-order dynamic system with one input, $r(t)$, and one output, $y(t)$.

$$\ddot{y}(t) - 2\dot{y}(t) + 7y(t) = \dot{r}(t) - 6r(t)$$

This type of system is called a single input–single output or *SISO* system. The transfer function is another way of writing a SISO system. The transfer function is *the ratio of the output variable over the input variable represented as the ratio of two polynomials in the D or s operator*.

Any linear ordinary differential equation can be converted to transfer function form using the following three step procedure. To illustrate the procedure, we'll convert the second-order differential equation to its transfer function form.

Step 1. Rewrite the equation using operator notation

$$D^2y(t) - 2Dy(t) + 7y(t) = Dr(t) - 6r(t)$$

Step 2. Collect and factor all output terms on the left side and input terms on the right side:

$$y(t) \cdot (D^2 - 2D + 7) = r(t) \cdot (D - 6)$$

Step 3. Obtain the transfer function by solving for the ratio of the output over the input signal:

$$\frac{y(t)}{r(t)} = \frac{(D - 6)}{(D^2 - 2D + 7)} = \text{Transfer function}$$

The transfer function consists of two polynomials in the D or s operator, a numerator polynomial, and a denominator polynomial. A *monic* polynomial has its highest D or s -power coefficient set to 1. To minimize the number of coefficients in a transfer function, the numerator and denominator polynomials are usually written in monic form with any gain factored out. For example, the following transfer function is converted to monic form by factoring 16 out of the numerator and 5 out of the denominator.

$$\frac{16D - 4}{5D^2 + 3D + 1} \xrightarrow{\text{Monic form}} \frac{16}{5} \cdot \left(\frac{D - \frac{4}{16}}{D^2 + \frac{3}{5}D + \frac{1}{5}} \right)$$

2.2 Block Diagrams, Manipulations, and Simulation

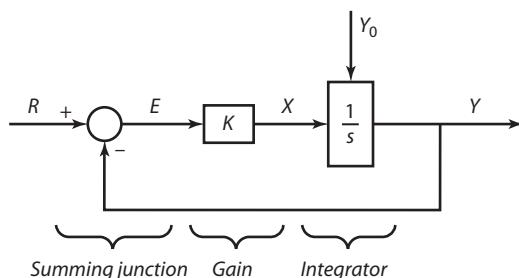
Simulation is the process of *solving* a block diagram model on a computer. Generally, simulation is the process of solving *any* model, but since block diagram models are so widely used, we will use block diagrams for all modeling tasks in this text. Block diagrams are usually part of a larger *visual programming environment*. Other parts of the environment may include numerical algorithms for integration, real-time interfacing, code generation, and hardware interfacing for high-speed applications. Visual programming environments are offered by many vendors and, depending on the supplier, will support different environment features.

2.2.1 Block Diagrams—Introduction

Block diagram models consist of two fundamental objects: *signal wires* and *blocks*. The function of a signal wire is to transmit a signal or value from its origination point (usually a block) to its termination point (usually another block). The *flow direction* of the signal is defined by an arrowhead on the signal wire. Once the flow direction has been defined for a given signal wire, all signals traveling on that wire must flow in the specified direction. A *block* is a processing element which operates on input signals and parameters to produce output signals. Because block functions may be nonlinear as well as linear, the collection of special function blocks is practically unlimited and almost never the same between vendors of block diagram languages. There is, however, a fundamental set of three basic blocks that all block diagram languages possess. These blocks are the summing junction, the gain, and the integrator. An example system using these three blocks is presented in Figure 2-1.

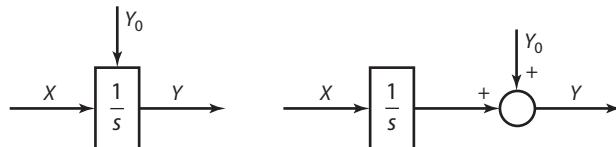
The vertical signal, Y_0 , entering the integrator from the top represents the initial condition on the integrator. When this signal is omitted, the initial condition is assumed to be 0.

FIGURE 2-1 THREE BLOCK SYSTEM EXAMPLE



The initial condition also could be represented as a summing junction downstream of the integrator, as shown in Figure 2-2.

FIGURE 2-2 METHODS FOR REPRESENTATION OF INTEGRATOR INITIAL CONDITION IN A BLOCK DIAGRAM



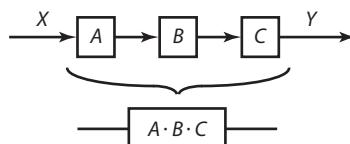
The block diagram will be used extensively in this text to represent system models. Once a system is represented in block diagram form, it can be analyzed or simulated. Analysis of block diagram systems involves reductions, usually to obtain the transfer characteristic between signals. These manipulations are discussed in the next section.

2.2.2 Block Diagrams—Manipulations

Block diagrams are rarely constructed in a standard form, and it is often necessary to reduce them to more efficient or understandable forms. The ability to simplify a block diagram is often a critical step in understanding its function and behavior. This section presents several basic rules which may be used to reduce a block diagram.

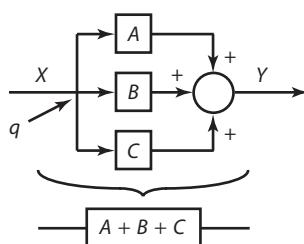
Series Block Reduction (Figure 2-3)

FIGURE 2-3 SERIES MANIPULATION—SERIES BLOCKS MULTIPLY



Parallel Block Reduction (Figure 2-4)

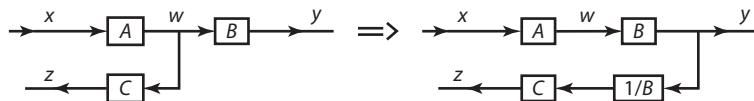
FIGURE 2-4 PARALLEL MANIPULATION—PARALLEL BLOCKS ADD



Moving Pick-Off Points Pick-off points are wire origination points located on a wire as opposed to a block output. When a signal is picked-off of a wire, both the signals and the picked-off signal are identical. It is often necessary to move pick-off points either downstream or upstream in order to create a parallel block configuration which can then be reduced using the parallel block reduction rule.

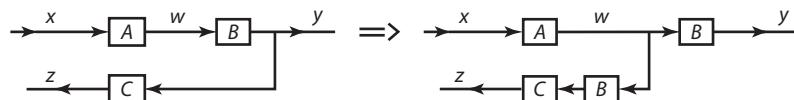
Downstream When a pick-off point is shifted downstream over a block, the inverse of the block appears in the feedback path. Figure 2-5 illustrates this reduction.

FIGURE 2-5 PICK-OFF POINT SHIFTED DOWNSTREAM



Upstream When a pick-off point is shifted upstream over a block, the block appears in the feedback path. Figure 2-6 illustrates this reduction.

FIGURE 2-6 PICK-OFF POINT SHIFTED UPSTREAM



Moving Blocks Through Summing Junctions Moving blocks through summing junctions is based on the distributive property of the summation operation, $y = k(A + B) = kA + kB$. Care must be taken to preserve the correct sign conventions. Two situations are considered: moving a block through a summing junction in the **upstream** direction (Figure 2-7) and moving a block through a summing junction in the **downstream** direction (Figure 2-8).

FIGURE 2-7 MOVING BLOCKS UPSTREAM THROUGH A SUMMING JUNCTION

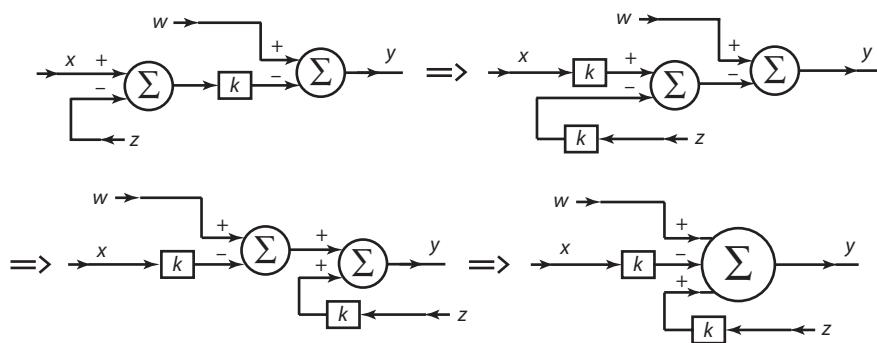
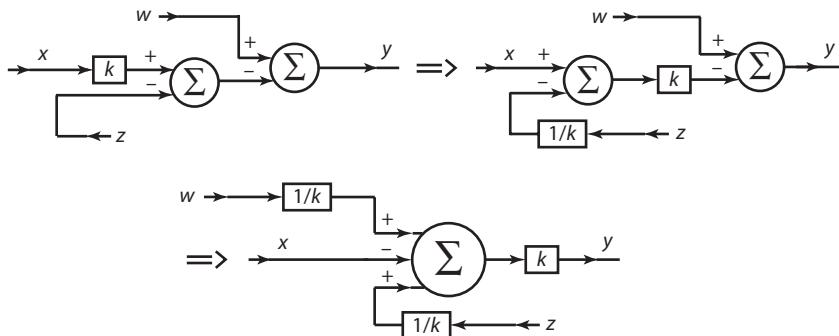
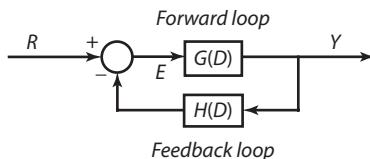


FIGURE 2-8 MOVING BLOCKS DOWNSTREAM THROUGH A SUMMING JUNCTION

Basic Feedback System Form One of the crucial ingredients of automatic control is *feedback*. It provides the mechanism for attenuating the effects of parameter variations and disturbances and enhancing dynamic tracking ability. The *basic feedback system* (BFS) shown in Figure 2-9 is the fundamental block diagram representing a feedback system.

FIGURE 2-9 BASIC FEEDBACK SYSTEM (BFS) BLOCK DIAGRAM

The variable R is the input to the BFS, E is the control or error variable, and Y is the output. The closed-loop transfer function for the BFS is computed by writing two equations in three variables, R , E , and Y ; then combining the equations to eliminate E ; and solving for the ratio of Y/R . These steps are illustrated here.

$$\text{Step 1. } E = R - H(D) \cdot Y$$

$$\text{Step 2. } Y = G(D) \cdot E$$

$$\text{Steps 1.} \rightarrow \text{2. } Y = G(D) \cdot (R - H(D) \cdot Y)$$

$$Y + G(D) \cdot H(D) \cdot Y = G(D) \cdot R$$

$$Y \cdot (1 + G(D) \cdot H(D)) = G(D) \cdot R$$

$$\frac{Y}{R} = \frac{G(D)}{1 + G(D) \cdot H(D)}$$

The function $G(D) \cdot H(D)$ represents the transfer function around the loop of the feedback system and is called the *loop transfer function* (LTf). If a system is in BFS form, its *closed loop transfer function* (CLTF or T) can be written directly as

$$T(D) = \frac{\text{forward loop transfer function}}{1 + \text{loop transfer function}} = \frac{G(D)}{1 + G(D) \cdot H(D)}$$

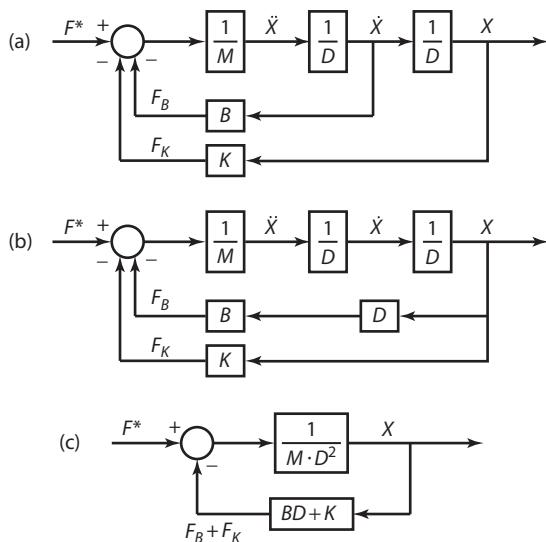
The denominator of T is called the *return difference* and is defined as $1 + \text{loop transfer function}$.

To illustrate the use of the block diagram reduction techniques just discussed, several examples are presented which utilize all of the manipulations presented so far.

EXAMPLE 2.1 Simple Feedback Diagram Reduction

Frequently, block diagram models consist of a series of nested feedback loops—each originating from a different pick-off point but terminating at one summing junction. For example, the mass–spring–damper system model in Figure 2-10 has two feedback loops which represent the reaction forces exerted by the damper and the spring.

FIGURE 2-10 SIMPLIFYING A MASS-SPRING-DAMPER BLOCK DIAGRAM



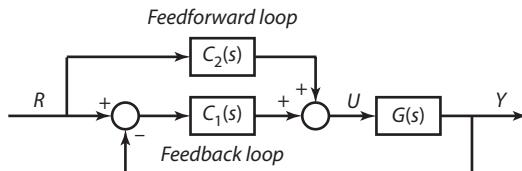
Solution

- Starting block diagram.
- The block diagram can be simplified by moving the \dot{X} pick-off point to X and making the appropriate scaling change, a multiplication by $1/D$, in the F_B path.
- The two feedback loops now originate from the same pick-off point, X , and terminate at the same summing junction so they can be combined as a parallel combination. Similarly the entire forward loop can be reduced as a series combination.

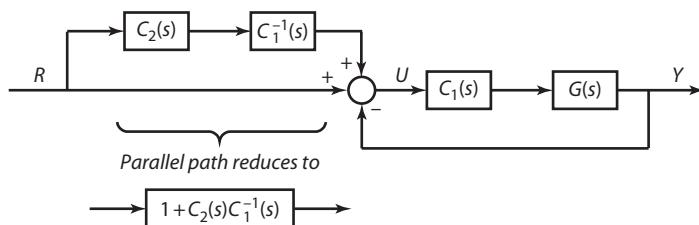
In this case, the price paid for the simplification is the loss of the \ddot{X} and \dot{X} signals. It is normal to expect the loss of some signal points, as a block diagram is simplified.

EXAMPLE 2.2 High-Performance Control

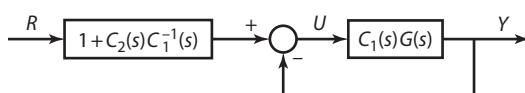
A control structure used in many high-performance systems combines feedforward control for fast response and feedback control for accuracy at lower frequencies. A block diagram of such a control structure being used to control a plant, $G(s)$, is presented in Figure 2-11

FIGURE 2-11 HIGH-PERFORMANCE FEEDFORWARD–FEEDBACK CONTROL SYSTEM**Solution**

The point of this example is to illustrate how the manipulations discussed previously may be applied to simplify the control section of the system block diagram. We begin by sliding the feedback-loop transfer function, $C_1(s)$, to the right side of the second summing junction and making the appropriate modification to the feedforward path using multiplication by $C_1^{-1}(s)$. Figure 2-12 presents the results.

FIGURE 2-12 FIRST STEP IN THE SIMPLIFICATION OF THE BLOCK DIAGRAM

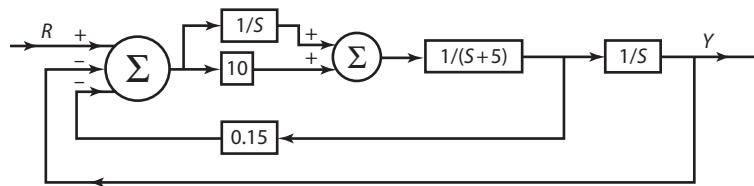
The two summing junctions now may be collapsed into a single *super* summing junction creating two parallel paths between it and the input pick-off point. The final simplified block diagram is shown in Figure 2-13.

FIGURE 2-13 FINAL SIMPLIFICATION OF THE BLOCK DIAGRAM

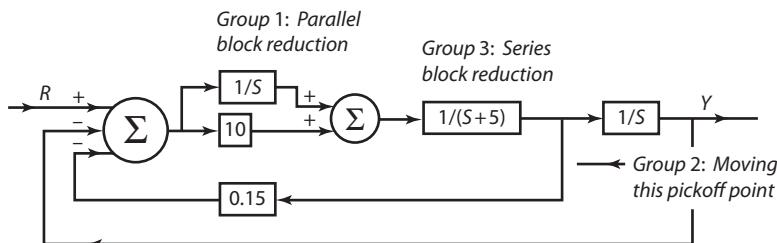
By selecting the feedforward-loop transfer function, such that $C_2(s) \cong G^{-1}(s)$, the effect of R on Y approaches 1, which means that changes in the *setpoint*, R , are *felt* immediately at the output, Y . The feedback-loop transfer function is usually selected for tracking accuracy and is often a proportional (PI or PID) type.

EXAMPLE 2.3 Feedback Plus Parallel Forward-Loop Diagram Reduction

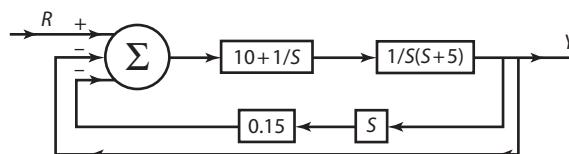
This example demonstrates series, parallel, and pick-off point movement manipulations. The block diagram, Figure 2-14, is to be reduced such that two blocks are present: one in the forward loop and one in the feedback loop. The reduced system will be in BFS form.

FIGURE 2-14 BLOCK DIAGRAM REDUCTION**Solution**

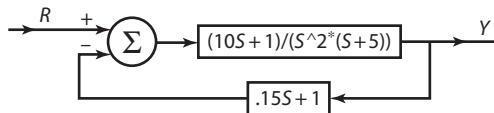
Identify subdiagrams as groups to which you can apply the manipulation rules (Figure 2-15).

FIGURE 2-15 BLOCK SUBDIAGRAMS

Group 1 is a parallel block manipulation, Group 2 is moving a pick-off point downstream, and Group 3 is a series block combination combining the $1/(S + 5)$ and $1/S$ blocks. Notice that the group operations are performed in a certain order. In this case, Group 2 is performed before Group 3 because the intermediate point disappears during the Group 3 series operation. Note also that the reason for moving the pick-off point in the first place was to create two parallel feedback loops. After performing these three group operations, the block diagram becomes that given in Figure 2-16.

FIGURE 2-16

The forward loop and the feedback loops now can be reduced using the series and parallel rule to produce the BFS form (Figure 2-17).

FIGURE 2-17

2.2.3 Simulation

Most visual simulation environments perform three basic functions.

- *Graphical Editing:* Used for the creation, editing, storage, and retrieval of models. Also used to create model inputs, orchestrate the simulation, and to present the model results.
- *Analysis:* Used to obtain transfer functions, compute frequency response, and evaluate sensitivity to disturbances.
- *Simulation:* Numerical solution of the block diagram model.

All models in a visual simulation environment are block-diagram based, so a textual programming is not necessary; however, some environments supplement their block libraries with such a language for greater flexibility. Since block diagrams were introduced in the previous section, we will proceed directly to the simulation process.

Simulation is the process through which the model equations are numerically solved.

The simulation process consists of three steps.

Step 1. Initialization

Step 2. Iteration

Step 3. Termination

In the initialization step, the equations for each block in the system model are *sorted* according to the pattern in which the blocks are connected. For example, a model consisting of three blocks (A, B, and C) connected in series (input to A is exogenous, output of A to input of B, output of B to input of C) would have its equations sorted with the Block A equations first, followed by those in Block B, and then by those in Block C. The exogenous input to A would precede the sorted list, as it is needed to process the A block. In the iteration step, differential equations present in the model are solved using numerical integration and/or differentiation, and the simulation time is advanced. Discrete equations are also solved in the iteration section. Results are presented in the termination step along with any other post-processed calculations. Output may be saved to a file, displayed as a digital reading, or graphically displayed as a chart, strip chart, meter readout, or even as an animation.

All visual modeling environments include the simulation function. Some of the most commonly used environments are MATRIXX/System Build (National Instruments), MATLAB/Simulink (Mathworks), LabVIEW (National Instruments), VisSim (Visual Solutions), and Easy5 (Boeing).

In the remainder of this chapter, we present two approaches for developing block diagram models from system illustrations: the *direct method* and the *modified analogy method*.

2.3 Block Diagram Modeling—Direct Method

The direct method for block diagram modeling is well suited for the modeling of simple, single-discipline models or of multidiscipline models with minimal coupling between disciplines. Normally, the starting point in these applications is either a set of linear ordinary differential equations, a transfer function, or an illustration of the system itself.

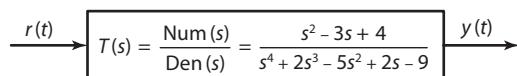
2.3.1 Transfer Function (or ODE) Conversion to Block Diagram Model

The procedure for converting a transfer function (or ODE) to a block diagram model is presented in this section as a six-step process. An *ordinary differential equation* (ODE) is a differential equation with all derivatives taken with respect to time. Time is the independent variable. A complete set of initial conditions must be specified for each (time) derivative term. It is assumed that the transfer function is in proper form, which means that the order of the numerator polynomial is less than or equal to the order of the denominator polynomial.

Given A transfer function is used here with input r , output y , and all required initial conditions. To better illustrate the procedure, we will apply it to the following illustrative transfer function, $T(s)$.

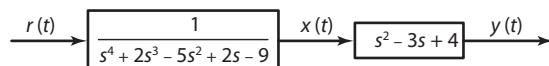
$$T(s) = \frac{Y(s)}{R(s)} = \frac{s^2 - 3s + 4}{s^4 + 2s^3 - 5s^2 + 2s - 9}; \quad y(0) = 1, \dot{y}(0) = -2, \ddot{y}(0) = 6, \dddot{y}(0) = 3$$

This transfer function can be written as the following top-level block diagram to show the numerator and polynomial polynomials.



Solution

Step 1. Create the state variable, $x(t)$, by “sliding” the numerator part of the transfer function into a new block located to the right of the denominator part of the transfer function. Connect the denominator and numerator blocks with an arrow and label the signal, $x(t)$, as the state variable. Include any transfer function gain term with the numerator block. The resulting block diagram is shown here.



Compute the order of the transfer function as the order of its denominator, ny . In this case, $ny = 4$.

Step 2. From step 1, write the *state equation* (SE) as the differential equation relating the input, $r(t)$, to the state, $x(t)$.

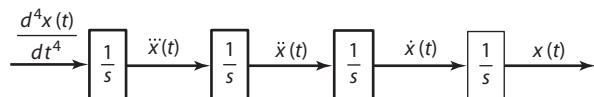
$$\text{SE: } \frac{x(t)}{r(t)} = \frac{1}{s^4 + 2s^3 - 5s^2 + 2s - 9}$$

or

$$\frac{d^4x(t)}{dt^4} + 2\frac{d^3x(t)}{dt^3} - 5\frac{d^2x(t)}{dt^2} + 2\frac{dx(t)}{dt} - 9x(t) = r(t)$$

Step 3. Begin constructing the block diagram by placing ny -integrator blocks in series and connect them from left to right. The input to the leftmost integrator block will be the highest derivative of

the state equation, in this case $\frac{d^4x(t)}{dt^4}$, and the output of the rightmost integrator block will be $x(t)$. Using our example system, there are four integrators written as follows.



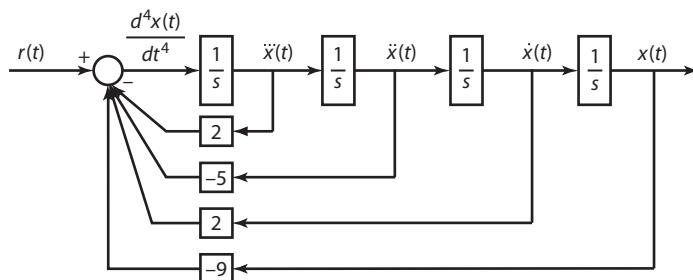
For now we'll ignore the initial conditions, they will be added in the last step of the procedure, step 6.

Step 4. Solve the state equation from step 2 for the highest derivative of the state variable. In this case we'd solve for

$$\frac{d^4x(t)}{dt^4} = -2 \frac{d^3x(t)}{dt^3} + 5 \frac{d^2x(t)}{dt^2} - 2 \frac{dx(t)}{dt} + 9x(t) + r(t)$$

Using a summing junction to represent the equality condition, we implement the previous state equation onto the block diagram (Figure 2-18) started in step 3 using the existing state variable and its derivatives (for the feedback parts) and also add a new external signal, $r(t)$.

FIGURE 2-18 STATE EQUATIONS TO BLOCK DIAGRAM



Notice the diagram that we have chosen to make all feedbacks at the summing junction negative, the other sign information is included in the feedback gains (i.e., -5 and -9).

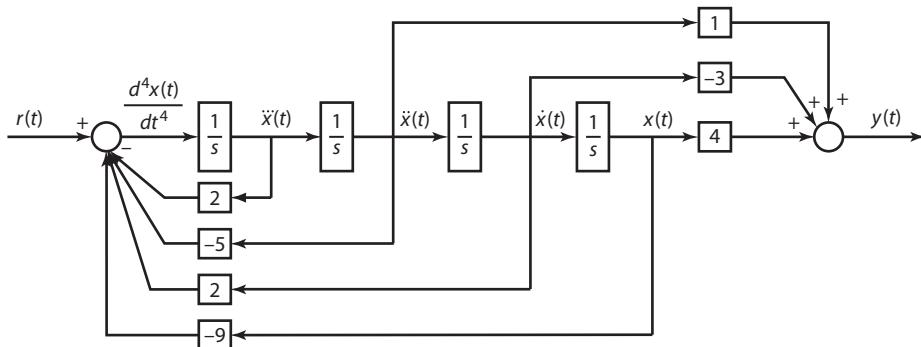
Step 5. From step 1, write the *output equation* (OE) as the differential equation relating the output, $y(t)$, to the state, $x(t)$, and its derivatives.

$$\text{OE: } \frac{y(t)}{x(t)} = s^2 - 3s + 4$$

or

$$\dot{x}(t) - 3\ddot{x}(t) + 4x(t) = y(t)$$

To complete this step, we implement the output equation on the block diagram from step 4 by combining the existing state variable and its derivatives through the appropriate gains and a summing junction to create the output signal, $y(t)$, as in Figure 2-19.

FIGURE 2-19 OUTPUT EQUATIONS IN BLOCK DIAGRAM

Step 6. Add the initial conditions to the block diagram in step 5. In order to do this, we must translate the initial conditions from the output variable, $y(t)$, to the state variable, $x(t)$, its derivatives, and possibly the input. Officially, a *state* is defined as the output of an integrator. In this example, there are four states given as $[x(t), \dot{x}(t), \ddot{x}(t), \dddot{x}(t)]$. Note that $\frac{d^4x(t)}{dt^4}$ is NOT a state; however, it can be written in terms of the states and input using the state equation from step 4 as

$$\frac{d^4x(t)}{dt^4} = -2\frac{d^3x(t)}{dt^3} + 5\frac{d^2x(t)}{dt^2} - 2\frac{dx(t)}{dt} + 9x(t) + r(t)$$

The translation process uses the output equation and its derivatives to perform this translation. We will also use the state equation to eliminate any $\frac{d^4x(t)}{dt^4}$ terms and represent them in terms of the states and possibly the input.

The following initial conditions are,

$$y(0) = 1, \dot{y}(0) = -2, \ddot{y}(0) = 6, \dddot{y}(0) = 3$$

The four output initial conditions are written in terms of the output equation evaluated at $t = 0$. The four equations are presented here.

1. $\ddot{x}(0) - 3\dot{x}(0) + 4x(0) = y(0) = 1$
2. $\dddot{x}(0) - 3\ddot{x}(0) + 4\dot{x}(0) = \dot{y}(0) = -2$
3. $\frac{d^4x(0)}{dt^4} - 3\ddot{x}(0) + 4\dot{x}(0) = \ddot{y}(0) = 6$

Substituting the state equation for $\frac{d^4x(0)}{dt^4}$ in this third equation yields

$$[-2\ddot{x}(0) + 5\dot{x}(0) - 2x(0) + 9x(0) + r(0)] - 3\ddot{x}(0) + 4\dot{x}(0) = \ddot{y}(0) = 6$$

so we have

$$-5\ddot{x}(0) + 9\dot{x}(0) - 2x(0) + 9x(0) + r(0) = \dot{y}(0) = 6$$

$$4. -5 \frac{d^4x(0)}{dt^4} + 9\ddot{x}(0) - 2\dot{x}(0) + 9x(0) + \dot{r}(0) = \ddot{y}(0) = 3$$

Again substituting the state equation for $\frac{d^4x(0)}{dt^4}$ in this fourth equation yields

$$-5[-2\ddot{x}(0) + 5\dot{x}(0) - 2x(0) + 9r(0)] + 9\ddot{x}(0) - 2\dot{x}(0) + 9x(0) + \dot{r}(0) = \ddot{y}(0) = 3$$

so we have

$$19\ddot{x}(0) - 27\dot{x}(0) + 19x(0) + 45r(0) + \dot{r}(0) - 5r(0) = \ddot{y}(0) = 3$$

Normally, the input and its derivatives are set to zero at time 0, and we are left with the task of solving four equations for four unknowns, $[x(0), \dot{x}(0), \ddot{x}(0), \ddot{\ddot{x}}(0)]$. In matrix form, this is written as

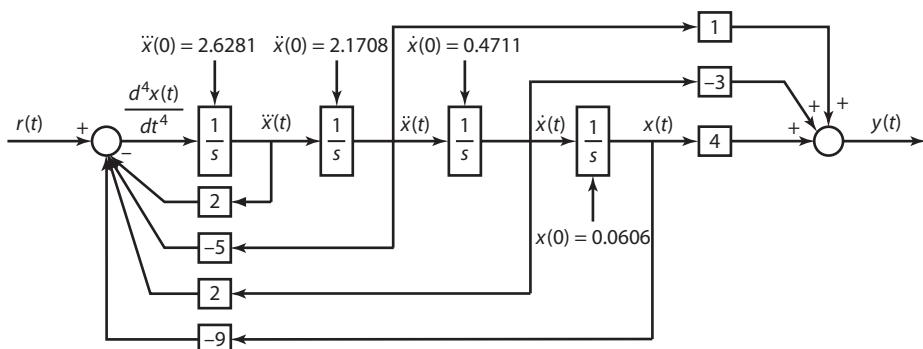
$$\begin{bmatrix} 0 & 1 & -3 & 4 \\ 1 & -3 & 4 & 0 \\ -5 & 9 & -2 & 9 \\ 19 & -27 & 19 & 45 \end{bmatrix} \cdot \begin{bmatrix} \ddot{x}(0) \\ \dot{x}(0) \\ x(0) \\ r(0) \end{bmatrix} = \begin{bmatrix} y(0) = 1 \\ \dot{y}(0) = -2 \\ \ddot{y}(0) = 6 \\ \ddot{\ddot{y}}(0) = 3 \end{bmatrix}$$

Solving for the state and its derivatives yields

$$\begin{bmatrix} \ddot{x}(0) \\ \dot{x}(0) \\ x(0) \\ r(0) \end{bmatrix} = \begin{bmatrix} 2.6281 \\ 2.1708 \\ 0.4711 \\ 0.0606 \end{bmatrix}$$

Step 6 is completed by adding the initial conditions to the block diagram from step 5. The completed block diagram is shown in Figure 2-20.

FIGURE 2-20 BLOCK DIAGRAM WITH INITIAL CONDITIONS



This example is somewhat complicated due to the order of the transfer function; however, the procedure for computing initial conditions will be the same for any transfer function.

EXAMPLE 2.4 Transfer Function to Block Diagram with No Input Dynamics

This example applies the six-step procedure to a transfer function having a denominator polynomial and only a gain term in the numerator. The transfer function and initial conditions are given as

$$T(s) = \frac{Y(s)}{R(s)} = \frac{3}{5s^2 + 8s + 13} = \frac{\text{Num}(s)}{\text{Den}(s)} ; \quad y(0) = 2, \dot{y}(0) = -2$$

Solution

Step 1. In problems like this, it will be simpler if we factor out the leading coefficients of the $\text{Num}(s)$ and $\text{Den}(s)$ to make them both monic polynomials. A *monic polynomial* has its highest s -power coefficient equal to 1. The monic form for $T(s)$ is written as

$$T(s) = \frac{Y(s)}{R(s)} = \frac{3}{5} \frac{1}{s^2 + 8/5s + 13/5} = 0.6 \frac{1}{s^2 + 1.6s + 2.6}$$

Next, we create the state variable, $x(t)$, by “sliding” the numerator part of the transfer function into a new block located to the right of the denominator part of the transfer function. Connect the denominator and numerator blocks with an arrow and label the signal, $x(t)$, the state variable. The resulting block diagram is



The order of the transfer function is $ny = 2$.

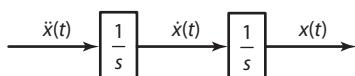
Step 2. From step 1, write the state equation (SE) as the differential equation relating the input, $r(t)$, to the state, $x(t)$, as

$$\text{SE: } \frac{x(t)}{r(t)} = \frac{1}{s^2 + 1.6s + 2.6}$$

or

$$\frac{d^2x(t)}{dt^2} + 1.6 \frac{dx(t)}{dt} + 2.6x(t) = r(t)$$

Step 3. Begin constructing the block diagram by placing ny -integrator blocks in series and connect them from left to right. The input to the leftmost integrator block will be the highest derivative of the state equation, in this case $\frac{d^2x(t)}{dt^2}$, and the output of the rightmost integrator block will be $x(t)$. Using our example system, there are two integrators, and they are written as

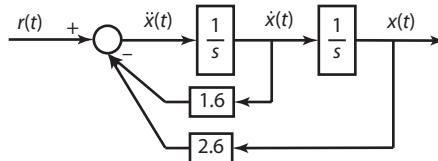


As before, we'll add the initial conditions in the last step of the procedure, step 6.

Step 4. Solve the state equation from step 2 for the highest derivative of the state variable, in this case we'd solve for $\frac{d^2x(t)}{dt^2}$ as,

$$\frac{d^2x(t)}{dt^2} = -1.6 \frac{dx(t)}{dt} - 2.6x(t) + r(t)$$

Using a summing junction to represent the equality condition, we implement the above form of the state equation onto the block diagram started in step 3. The right hand side of the state equation will always be a function of the state variable, its derivatives, and the input. The state variable and its derivatives have already been created as a result of step 3. At this point we will need to add a new external signal for the input, $r(t)$. The resulting updated block diagram is presented in Figure 2-21.

FIGURE 2-21


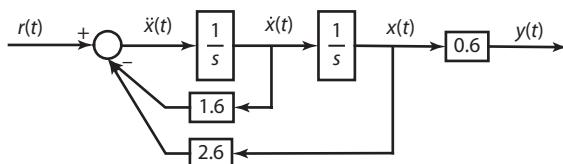
Step 5. From step 1, write the output equation (OE) as the differential equation relating the output, $y(t)$, to the state, $x(t)$, and its derivatives. Since there are no input dynamics, the output equation in this case is particularly simple. It is not a differential equation but rather a static equation.

$$\text{OE: } \frac{y(t)}{x(t)} = 0.6$$

or

$$0.6x(t) = y(t)$$

Step 5 is completed by implementing the output equation onto the block diagram from step 4. Since the output equation is only a gain, the implementation is straightforward and presented in Figure 2-22.

FIGURE 2-22


Step 6. Add the initial conditions to the block diagram in step 5. In order to do this, we must translate the initial conditions from the output variable, $y(t)$, to the state variable, $x(t)$, its derivatives, and possibly the input. In this example, there are two states given as $[x(t), \dot{x}(t)]$. Note that $\ddot{x}(t)$ is NOT a state, however, it can be written in terms of the states and input using the state equation from step 4 as

$$\frac{d^2x(t)}{dt^2} = -1.6 \frac{dx(t)}{dt} - 2.6x(t) + r(t)$$

The translation process uses the output equation and its derivatives to compute the state initial conditions. We will also use the state equation to eliminate any $\frac{d^2x(t)}{dt^2}$ -terms and represent them in terms of the states and input.

The following output initial conditions were given as

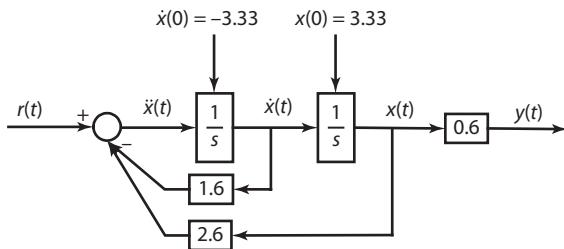
$$y(0) = 2, \dot{y}(0) = -2$$

These two output initial conditions are written in terms of the state initial conditions using the output equation and its derivatives. The equations are presented as

1. $0.6x(0) = y(0) = 2 \rightarrow x(0) = 3.33$
2. $0.6\dot{x}(0) = \dot{y}(0) = -2 \rightarrow \dot{x}(0) = -3.333$

Step 6 is completed by adding the initial conditions to the integrators in the block diagram created in step 5. The completed block diagram is shown in Figure 2-23.

FIGURE 2-23



This example is much less complicated than the previous example due to the absence of numerator dynamics.

EXAMPLE 2.5 | ODE to Block Diagram

This example applies the six-step procedure for converting a transfer function to a block diagram and then to a differential equation. A mass–spring–damper system defined by its free-body equations is to be modeled as a block diagram.

An illustration of the mass–spring–damper system is presented in Figure 2-24 along with its free-body equations. Prior to application of the input signal, $F(t)$, the system is initially at rest with the initial conditions $x(0) = x_0$, $\dot{x}(0) = 0$.

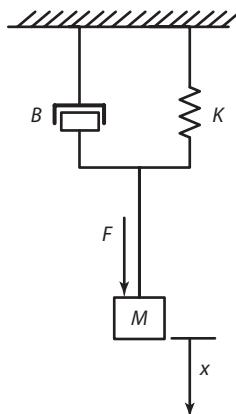
1. Sum of force equation: $\sum F(t) = M\ddot{x}(t)$
2. Restraining force due to spring: $F_k(t) = K(x(t) - x_0)$
3. Restraining force due to damper: $F_B(t) = B\dot{x}(t)$

Solution

Noting that $\sum F(t)$ equals $F(t) - F_k(t) - F_B(t)$, Equation (1) is rewritten, after substitution of $F_k(t)$ and $F_B(t)$ as

$$4. F(t) - B\dot{x}(t) - K(x(t) - x_0) = M\ddot{x}(t)$$

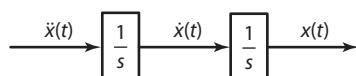
Note: x_0 is the initial displacement of the spring before application of the Force, F .

FIGURE 2-24

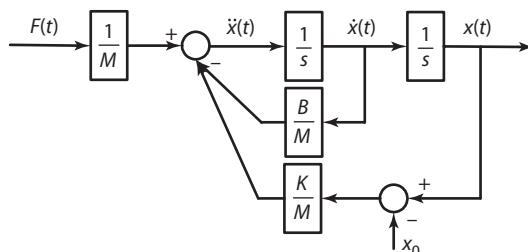
Step 2. For this example, we will take the mass displacement, $x(t)$, as the output, $y(t)$. With some manipulation, the state equation for the mass–spring–damper system is written as

$$\ddot{x}(t) = -\frac{B}{M} \dot{x}(t) - \frac{K}{M} (x(t) - x_0) + \frac{1}{M} F(t)$$

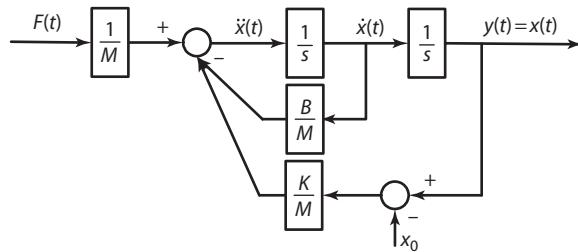
Step 3. This puts us at step 3 in our procedure. Noting that the equation is second order, we will begin construction of the block diagram with two integrators as



Step 4. We have already solved the state equation for the highest derivative of the state variable, so in the remainder of this step, we'll implement it onto the block diagram started in step 3. The resulting updated block diagram is presented in Figure 2-25. Note the input has been scaled by $1/M$ before entering the summing junction and the Δx input to the spring has been represented using a summing junction to remove the initial displacement, x_0 , from $x(t)$.

FIGURE 2-25

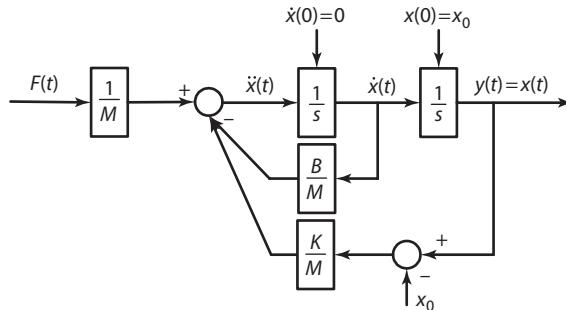
Step 5. The Output Equation (OE) for this example is $y(t) = x(t)$. (See Figure 2-26.)

FIGURE 2-26 PARALLEL SIMPLIFICATION FOR THE SIMPLE IMPEDANCE SYSTEM

Step 6. In this last step, we apply the initial conditions to each of the two states using the output equation, $y(t) = x(t)$. The calculations are presented here.

1. $x(0) = y(0)$, $x(0) = x_0$
2. $\dot{x}(0) = \dot{y}(0)$, $\dot{x}(0) = 0$

Adding the initial condition information (Equations 1 and 2) to the block diagram from step 5 produces the completed block diagram (Figure 2-27) for the mass–spring–damper system.

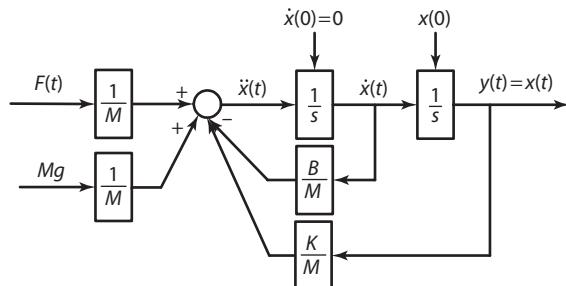
FIGURE 2-27

In some situations, the x_0 value is used to represent the spring displacement value that causes the spring force to equal the force of gravity on the mass. The force due to gravity is represented in Figure 2-28 as an additional force input on the summing junction and the displacement initial condition is shown as $x(0)$. Prior to application of the force input, $F(t)$, the system is motionless, (i.e., $\dot{x}(0) = \dot{x}(0) = 0$). In this state, the equation at the summing junction and displacement initial condition becomes

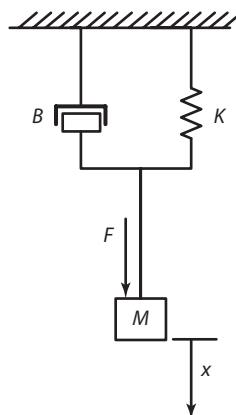
$$\frac{Mg}{M} - \frac{K}{M}x(0) = 0 \rightarrow x(0) = \frac{Mg}{K}$$

2.3.2 Conversion of Mechanical Illustrations to Block Diagram Models

The procedure for converting a system illustration to a block diagram model is primarily applicable to single domain systems such as mechanical translation or mechanical rotation. The method makes use of the basic force relationships for the three basic mechanical components: the mass, spring, and damper.

FIGURE 2-28

Given A system illustration is used with input r , output y , and all required initial conditions. As in the transfer function approach, we'll develop the modeling steps using an illustrative example. In this case, we'll use the mass–spring–damper system introduced in the Example 2.5: ODE to Block Diagram presented previously. The system is presented in Figure 2-29 for reference.

FIGURE 2-29

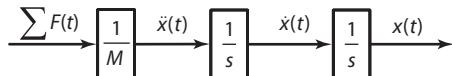
As before, the input is defined as the force, $F(t)$, and the output as the displacement, $x(t)$.

Solution

Step 1. For each mass in the illustration, write the $\sum F(t) = M\ddot{x}(t)$ equation and solve it for the acceleration of the particular mass. In our example system, we would write

$$\ddot{x}(t) = \frac{1}{M} \sum F(t)$$

Next we begin the block diagram by writing the $\ddot{x}(t)$ equation with input $\sum F(t)$ passed through a gain block of $\frac{1}{M}$ to create $\ddot{x}(t)$ followed by two series integrators to create the motion variables $\dot{x}(t)$, $x(t)$ for the mass. For our example system, the following block diagram is written as



Step 2. For each mass in the illustration, write the $\sum F(t)$ equation in terms of its components, the input (external force), spring force, and damping force. From the equation, we see that $F(t)$ moves in the same direction as $x(t)$. Also the force due to the spring, $F_K(t)$, and the force due to the damper, $F_B(t)$, restrain the motion (move in the opposite direction). We can write the following equation for the sum of forces as

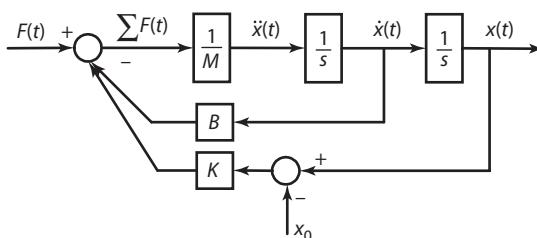
$$\sum F(t) = F(t) - F_K(t) - F_B(t)$$

In this step, we further define the spring and damper forces in terms of the states from each of the masses. In this example, there is only one mass, and the states are $\dot{x}(t)$, $x(t)$. The spring and damper forces are defined as

$$\begin{aligned} F_K(t) &= K(x(t) - x_0) \\ F_B(t) &= B\dot{x}(t) \end{aligned}$$

Step 3. Implement the step 2 equations on the diagram begun in step 1. You will probably find it necessary to redraw the resulting block diagram (Figure 2-30) to obtain the most concise and readable form. The block diagram obtained is slightly different in form from the previous example, which modeled the ODE's directly; however, the functionality is identical. We'll present several examples to illustrate this modeling method to mechanical systems with multiple masses.

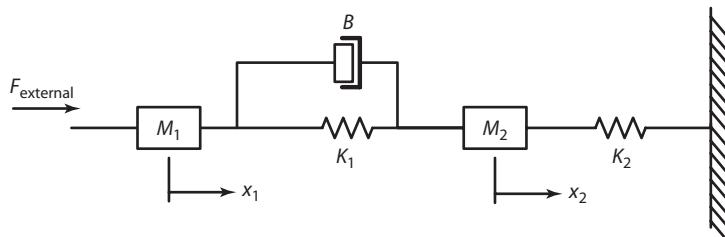
FIGURE 2-30



EXAMPLE 2.6 Two-Mass Mechanical System

This example illustrates how a two-mass mechanical translation system is modeled using the approach described previously. The system is described by Figure 2-31.

As before, the input is defined as the force, $F(t)$, and the output as the displacement, $x(t)$.

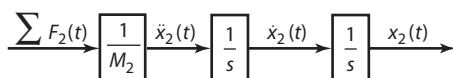
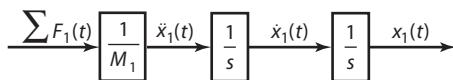
FIGURE 2-31**Solution**

Step 1. For each mass in the illustration, write the $\sum F(t) = M\ddot{x}(t)$ equation and solve it for the acceleration of the particular mass. In this example, we have two masses with the following equations.

$$\text{Mass 1: } \ddot{x}_1(t) = \frac{1}{M_1} \sum F_1(t)$$

$$\text{Mass 2: } \ddot{x}_2(t) = \frac{1}{M_2} \sum F_2(t)$$

We represent these equations by the following block diagram fragments



Step 2. For each of the two masses, write the $\sum F(t)$ equation in terms of its components, the input (external force), spring force, and damping force. From step 1, we can write the following equations.

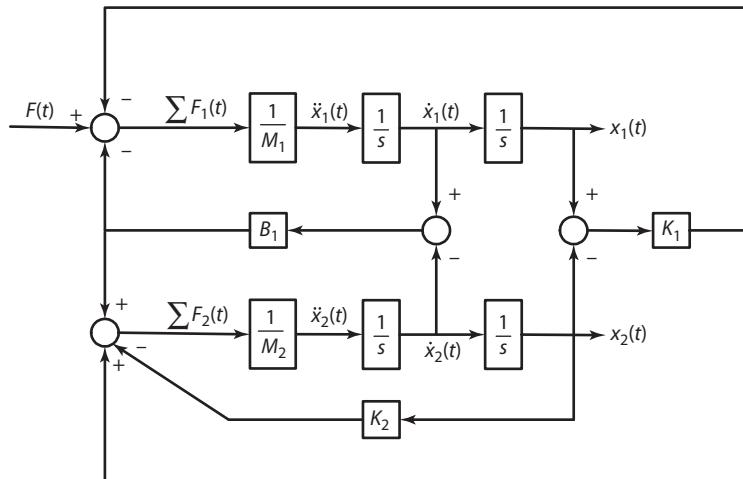
$$\sum F_1(t) = F_1(t) - K_1(x_1(t) - x_2(t)) - B(\dot{x}_1(t) - \dot{x}_2(t))$$

$$\sum F_2(t) = K_1(x_1(t) - x_2(t)) + B(\dot{x}_1(t) - \dot{x}_2(t)) - K_2 x_2(t)$$

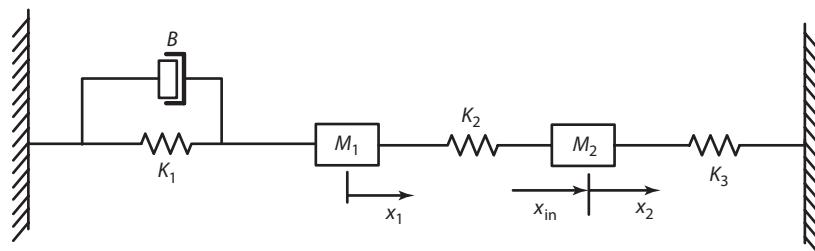
Note the sign convention used. In the first equation, the spring and damping force act to restrain the motion of mass 1 and are therefore negative. Since the masses are connected by the spring damper pair, the effect on mass 2 is equal and opposite, hence the positive sign. Note also that in the second equation, we have defined the ground displacement where spring \$K_2\$ is attached to be zero. In general, this could be any value.

Step 3. Implement the step 2 equations on the diagram that was started in step 1. After some minor manipulation, the final diagram is presented in Figure 2-32.

This example as well as the previous examples have all used force as the input signal. Occasionally, one will encounter models that use displacement or some other motion variable as the input. The next example presents such a system and the direct approach used to obtain the block diagram model.

FIGURE 2-32
EXAMPLE 2.7 Mechanical System with Displacement Input

This example illustrates how a two-mass mechanical translation system is modeled using the approach described in Example 2.6. The system is described by Figure 2-33.

FIGURE 2-33

In this diagram, the input is defined as a displacement, $x_{in}(t)$, instead of a force. We will apply the direct approach to obtain the block diagram for this system.

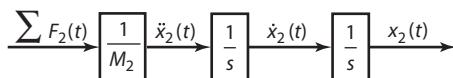
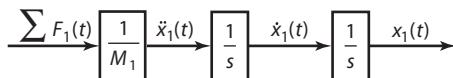
Solution

Step 1. For each mass in the illustration, write the $\sum F(t) = M\ddot{x}(t)$ equation, and solve it for the acceleration of the particular mass. In this example, we have two masses with the following equations.

$$\text{Mass 1: } \ddot{x}_1(t) = \frac{1}{M_1} \sum F_1(t)$$

$$\text{Mass 2: } \ddot{x}_2(t) = \frac{1}{M_2} \sum F_2(t)$$

As in the previous example, we represent these equations by the following block diagram fragments.



Step 2. For each of the two masses, write the $\sum F(t)$ equation in terms of its components, the input (external force), spring force, and damping force. From step 1, we can write the following equations.

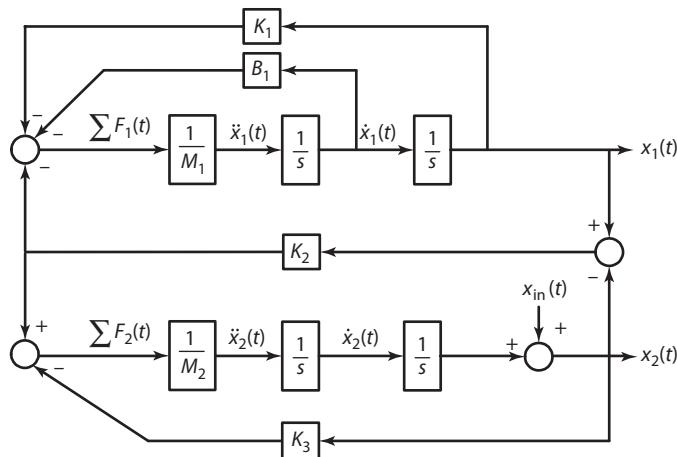
$$\sum F_1(t) = -K_1 x_1(t) - B \dot{x}_1(t) - K_2(x_1(t) - x_2(t)) - x_{in}(t)$$

$$\sum F_2(t) = K_2(x_1(t) - x_2(t)) - x_{in}(t) - K_d(x_2(t) + x_{in}(t))$$

Since the input displacement, $x_{in}(t)$ is aligned in direction with $x_2(t)$ and is added into the second equation with the same sign convention as $x_2(t)$. Also note that the displacements of the two grounds has been defined to be zero.

Step 3. Implement the step 2 equations on the diagram that was started in step 1. The diagram in Figure 2-34 is quite similar to the one in Example 2.6, however, the input force signal is absent.

FIGURE 2-34



2.4 Block Diagram Modeling—Analogy Approach

All disciplines of engineering are based on sets of fundamental laws or relationships. Electrical engineering relies on Ohm's and Kirchoff's laws, mechanical engineering on Newton's law, electromagnetics on Faraday's and Lenz's laws, fluids on continuity and Bernoulli's law, and so on. These laws are used to predict the behavior (both static and dynamic) of systems. Systems may exist completely in one engineering discipline (such as an electric circuit, a gear system, or a water distribution system), or they may be coupled between several disciplines (such as electromechanical, electromagnetic, etc). Although analytic solutions are appropriate for single discipline static equations it is more often the case that computer based solution methods are required, especially when dynamics are present in the equations.

System modeling is the derivation and representation of the equations which describe the behavior of a system. The term representation is used to indicate that the equations have been *prepared* for computer solution as a computer program. System modeling requires knowledge of the fundamental laws in each discipline of engineering to derive equations. Taken separately, application of the various laws is straightforward; however, for coupled systems (such as electromechanical, electrothermal, or fluidmechanical), it is often difficult to combine the equations. This section presents a method based on electrical analogies for deriving the *fundamental* equations of systems (single or coupled) in five disciplines of engineering: electrical, mechanical, electromagnetic, fluid, and thermal.

The *modeling by analogy* method, or *analogy method* as it is often called, became popular during the era of the analog computer. Although the method was originally intended for use on linear or linearized systems, it may be applied to some nonlinear systems as well. The *analogy* method becomes even more powerful when combined with block diagram modeling. By using the *analogy* method to first derive the fundamental relationships in a system, the equations then can be represented in block diagram form, allowing secondary and nonlinear effects to be added. This two-step approach is especially useful when modeling large coupled systems using block diagrams.

2.4.1 Potential and Flow Variables, PV and FV

Systems consist of components such as springs and dampers in mechanical systems, tanks and restrictions in fluid systems, and insulators and thermal capacitances in thermal systems. When in motion, the energy in a system can be *increased* by an energy-producing source outside the system, *redistributed* between components within the system, or *decreased* by energy loss through components out of the system. In this context, a *coupled system* becomes synonymous with energy transfer between systems.

Since the analogy method was developed for use on analog computers, it is fitting that the approach be described from a basic electrical viewpoint. Electrical systems are based on three fundamental components:

- Resistor
- Capacitor
- Inductor

The capacitor and inductor are capable of storing energy. The energy stored in a capacitor is $Cv^2/2$ and the energy stored in an inductor is $Li^2/2$. The resistor cannot store energy but can transfer electrical energy into heat energy.

In an ideal, lossless LC circuit with nonzero initial energy, all energy remains in the circuit and is transferred back and forth in sustained oscillations between the inductor and capacitor. Addition of a resistor establishes an energy leak to the surrounding air through which heat energy is transferred, causing the oscillations to decay in amplitude and eventually disappear. If the resistor were immersed in a fluid such as water, the temperature of the fluid would rise due to the heat energy transferred to it. In the steady state, all electrical energy in the circuit would be converted to heat energy in the fluid. Further addition of a voltage or current source to the circuit would provide an external source of energy into the circuit. If the source had a nonzero mean value, the heat energy transferred to the fluid would be sustained.

Total energy, E , in the LC circuit consists of potential energy, U , and kinetic energy, K . Potential energy is associated with the *potential* to perform work and kinetic energy with the work to change motion or *flow*. Based on this association two energy related are defined as

$$\text{Potential variable} = PV$$

$$\text{Flow variable} = FV$$

For a given system, the choice of the potential and flow variables are not unique. For example, in an LC circuit, the initial energy may exist in either the capacitor as a potential, in the inductor as a current, or in both. If the potential energy is stored entirely in the capacitor, voltage becomes the natural choice for the potential variable and, current becomes the flow variable. On the other hand, if the potential energy is stored entirely in the inductor, then current may be used as the potential variable and voltage as the flow variable.

Since it is natural to picture current as *flowing* and voltage drops as *accumulating* through an electrical circuit, the flow variable in an electrical circuit is current, and the potential variable is voltage.

2.4.2 Impedance Diagrams

In an electrical circuit the impedance of a component is defined as the ratio of the voltage phasor, v , across the component over the current phasor, i , through the component. Since voltage and current are complex numbers, the impedance is also a complex number. A complex number consists of a real part and an imaginary part. The placeholder for the imaginary part is j , and no placeholder is required for the real part.

The impedance of an electrical circuit element is a complex phasor quantity defined as the ratio of the voltage phasor divided by the current phasor. The impedance phasors for the capacitor, inductor, and resistor are summarized in Figure 2-35 and are shown as bold arrows. Positive phase occurs when the phasor is rotated in the *counterclockwise* direction beginning from the positive real axis (which is the zero phase direction). When the phasor is lined up with the positive imaginary axis (vertically upward) 90° of the phase has been accumulated. When the phasor is pointing leftward, 180° of the phase has been accumulated. When the phasor is pointing downward along the negative imaginary axis, 270° or -90° of the phase has been accumulated.

Keeping in mind that impedance is voltage divided by current, a positive imaginary component indicates voltage leading current, and a negative imaginary component indicates voltage lagging current. Because j occurs in the denominator of the capacitor impedance, the capacitor voltage *lags* its current by 90°. Similarly, because j occurs in the numerator of the inductor impedance, the

FIGURE 2-35 IMPEDANCE PHASORS FOR THE CAPACITOR, INDUCTOR, AND RESISTOR

Element	Impedance	Phasor
Capacitor	$Z_C = \frac{1}{j\omega C}$	<p>A phasor diagram in a 2D plane with a horizontal "Real" axis and a vertical "Imaginary, j" axis. A bold arrow points vertically upwards along the positive imaginary axis, representing a 90° phase lead.</p>
Inductor	$Z_L = j\omega L$	<p>A phasor diagram in a 2D plane with a horizontal "Real" axis and a vertical "Imaginary, j" axis. A bold arrow points vertically upwards along the positive imaginary axis, representing a 90° phase lag.</p>
Resistor	$Z_R = R$	<p>A phasor diagram in a 2D plane with a horizontal "Real" axis and a vertical "Imaginary, j" axis. A bold arrow points horizontally to the right along the positive real axis, representing no phase shift.</p>

inductor voltage *leads* its current by 90° . The imaginary component of impedance for a resistor is zero, indicating that the current and voltage are in phase with one another.

Consider the sinusoid $x(t) = \sin \omega t$. If we differentiate $x(t)$ analytically with respect to time, we obtain

$$\dot{x}(t) = \frac{d(\sin(\omega t))}{dt} = \omega \cdot \cos(\omega t)$$

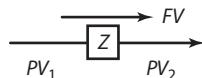
Furthermore, since $\cos \lambda = \sin(\lambda + 90^\circ)$, the right side of $\dot{x}(t)$ may be written as $\omega \cdot \sin(\omega t + 90^\circ)$ or simply $j\omega \cdot \sin(\omega t)$. This means that differentiation of a sinusoid of frequency ω is the same as multiplication of the sinusoid by $j\omega$.

The impedance of a component is often represented as Z_X , where X is the component name or description. In terms of the potential and flow variables, the impedance of a component is defined as the ratio of the potential variable to the flow variable, as given in Equation 2-1.

$$Z_{\text{Component}} \equiv \frac{\Delta PV}{FV} \quad (2-1)$$

For example, consider the circuit element shown in Figure 2-36.

FIGURE 2-36 UNKNOWN CIRCUIT ELEMENT



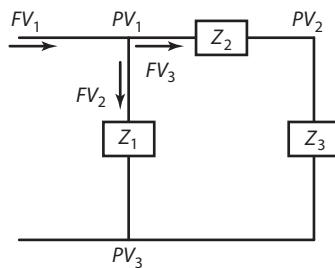
In accordance with Equation 2-1, the impedance of the circuit element becomes

$$Z = \frac{PV_1 - PV_2}{FV} \equiv \frac{\Delta PV}{FV}$$

EXAMPLE 2.8 Impedance Calculations for a Parallel System

This example illustrates how impedances are calculated in a parallel system. The system shown in Figure 2-37 has three impedance's, three flow variables, and three potential variables.

FIGURE 2-37 SIMPLE CIRCUIT FOR IMPEDANCE CALCULATIONS



Solution

Using Equation 2-1, the impedance's are calculated as

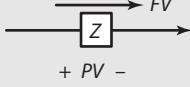
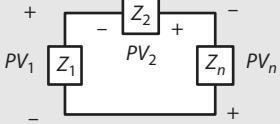
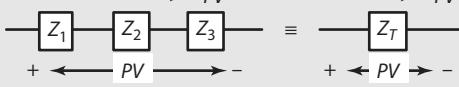
$$\begin{aligned} Z_1 &= \frac{PV_1 - PV_3}{FV_2} \equiv \frac{\Delta PV_{13}}{FV_2} \\ Z_2 &= \frac{PV_1 - PV_2}{FV_3} \equiv \frac{\Delta PV_{12}}{FV_3} \\ Z_3 &= \frac{PV_2 - PV_3}{FV_3} \equiv \frac{\Delta PV_{23}}{FV_3} \end{aligned}$$

PV_3 is a common potential point in the circuit. It is usually set to either zero or a reference value. Setting PV_3 to zero, the impedance equations may be reduced to

$$\begin{aligned} Z_1 &= \frac{PV_1}{FV_2} \\ Z_2 &= \frac{PV_1 - PV_2}{FV_3} \equiv \frac{\Delta PV_{12}}{FV_3} \\ Z_3 &= \frac{PV_2}{FV_3} \end{aligned}$$

In many situations, an impedance diagram can be simplified by applying any of six fundamental impedance relationships. These relationships, which are based on Ohm's and Kirchoff's Laws, are summarized in Table 2-2.

TABLE 2-2 FUNDAMENTAL IMPEDANCE RELATIONSHIPS

Impedance Configuration	Relationship (Name)
	$PV = Z \cdot FV$ (Basic impedance relationship)
	$0 = \sum_{k=1}^n FV_k$ (FV node)
	$0 = \sum_{k=1}^n PV_k$ (PV around a closed loop)
	$Z_T = Z_1 + Z_2 + Z_3$ (Series impedance's)

	$\frac{1}{Z_T} = \frac{1}{Z_1} + \frac{1}{Z_2} + \frac{1}{Z_3}$ <p>(Parallel impedance's)</p>
	$PV_{\text{out}} = \frac{Z_2 + Z_3}{Z_1 + Z_2 + Z_3} \cdot PV$ <p>(Potential divider)</p>
	$FV_1 = \frac{Z_2 Z_3}{Z_1 Z_2 + Z_1 Z_3 + Z_2 Z_3} FV$ $FV_2 = \frac{Z_1 Z_3}{Z_1 Z_2 + Z_1 Z_3 + Z_2 Z_3} FV$ $FV_3 = \frac{Z_1 Z_2}{Z_1 Z_2 + Z_1 Z_3 + Z_2 Z_3} FV$ <p>(Flow divider)</p>

Parallel and series impedance reductions will be used frequently in our manipulations. The following properties will be used repeatedly.

- *Series Impedance's Add:* The total impedance of a series combination is the sum of the individual impedance's.
- *Parallel Impedance's–Inverses Add:* The inverse of the total impedance of a parallel combination is the sum of the inverses of the individual impedance's.

To illustrate how the impedance relationships are applied, several examples are presented.

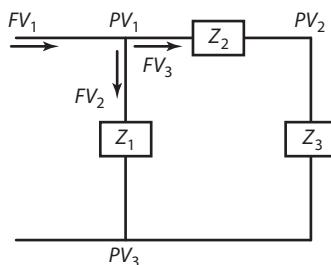
EXAMPLE 2.9 Impedance Diagram Simplification—Simple System

This example illustrates how series and parallel reductions can be applied to the previous example to derive a single representative impedance, Z_{Total} , for the entire system. The system, which is rewritten in Figure 2-38, is reduced in two steps.

Step 1. Combine the Z_2 and Z_3 impedance's into a single series impedance, Z_{23} .

Step 2. Combine the Z_1 and Z_{23} impedance's into a single parallel impedance, Z_{Total} .

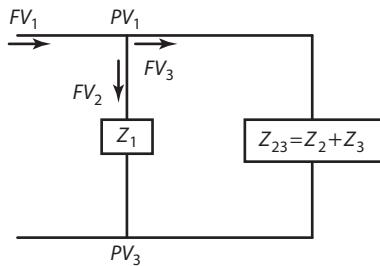
FIGURE 2-38 SIMPLE IMPEDANCE SYSTEM



Solution

Step 1. The Z_2 and Z_3 impedance's are combined into the single series impedance, Z_{23} , according to the series relationship, $Z_{23} = Z_2 + Z_3$. The impedance diagram is presented in Figure 2-39.

FIGURE 2-39 SERIES SIMPLIFICATION FOR THE SIMPLE IMPEDANCE SYSTEM



Inevitably, some signals are lost as a result of impedance diagram simplifications. In this simplification, we have lost the PV_2 signal.

Step 2. The Z_1 and Z_{23} impedance's are combined into the single parallel impedance, Z_{Total} , according to the series relationship, $\frac{1}{Z_{\text{Total}}} = \frac{1}{Z_1} + \frac{1}{Z_{23}}$. It is awkward to leave this calculation in this form, so it is simplified to produce Z_{Total} as

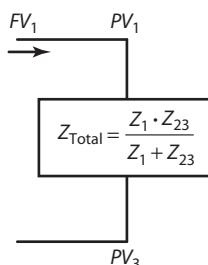
$$Z_{\text{Total}} = \left(\frac{1}{Z_1} + \frac{1}{Z_{23}} \right)^{-1} = \frac{Z_1 \cdot Z_{23}}{Z_1 + Z_{23}}$$

This result is important because it is encountered so frequently. It is summarized as

The combined impedance of parallel branches is equal to the product of the two impedance's divided by the sum of the two impedance's.

You may find it helpful to use this relationship in place of the parallel relationship presented in Table 2-2. The final result of this simplification produces the impedance diagram shown in Figure 2-40.

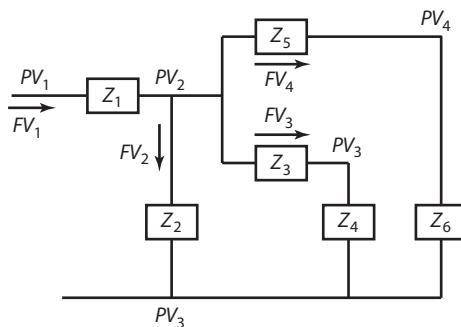
FIGURE 2-40 PARALLEL SIMPLIFICATION FOR THE SIMPLE IMPEDANCE SYSTEM



It is important to note that the flow through Z_{Total} is FV_1 and not FV_2 . Also, in this step of the simplification, we have lost the flow variables, FV_2 and FV_3 .

EXAMPLE 2.10 Impedance Diagram Simplification—Complex System

This example illustrate how series and parallel reductions can be applied to a more complex system. The system, Figure 2-41, is typical of the type encountered in mechanical systems with several masses. The objective is to reduce the diagram to a single equivalent impedance.

FIGURE 2-41 COMPLEX IMPEDANCE SYSTEM


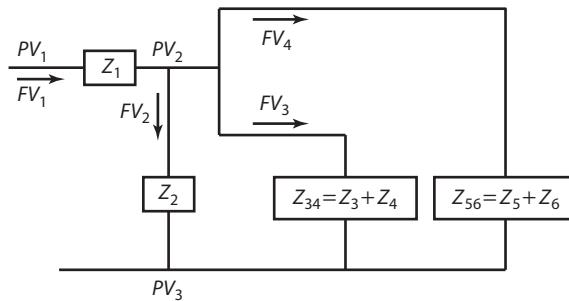
We will solve the problem in the four steps outlined below.

- Step 1.** Combine the Z_5 and Z_6 impedance's into a single series impedance, Z_{56} .
- Step 2.** Combine the Z_3 and Z_4 impedance's into a single series impedance, Z_{34} .
- Step 3.** Combine the Z_2 , Z_{34} , and Z_{56} impedance's into a single parallel impedance, Z_{23456} .
- Step 4.** Combine the Z_1 and Z_{23456} impedance's into a single series impedance, Z_{Total} .

Solution

Steps 1 and 2. The Z_5 and Z_6 impedance's are combined into a single series impedance, Z_{56} according to the series relationship, $Z_{56} = Z_5 + Z_6$. A similar combination is performed on the Z_3 and Z_4 impedance's forming $Z_{34} = Z_3 + Z_4$. The impedance diagram is presented in Figure 2-42.

The two potential variables, PV_3 and PV_4 , are lost in this simplification.

FIGURE 2-42 COMPLEX IMPEDANCE SYSTEM SIMPLIFIED AS PER STEPS 1 AND 2


Step 3. The Z_2 , Z_{34} , and Z_{56} impedance's are reduced to a single parallel impedance, Z_{23456} , by applying the parallel reduction as

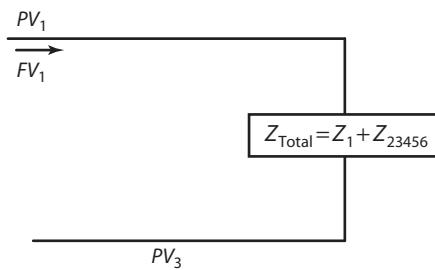
$$\frac{1}{Z_{23456}} = \frac{1}{Z_2} + \frac{1}{Z_{34}} + \frac{1}{Z_{56}}$$

$$Z_{23456} = \frac{Z_2 \cdot Z_{34} \cdot Z_{56}}{Z_{34}Z_{56} + Z_2Z_{56} + Z_2Z_{34}}$$

As a result of this simplification, the three flow variables, FV_2 , FV_3 , and FV_4 , are lost. Also notice that the flow through the entire diagram is now FV_1 .

Step 4. The reduction is completed by combining the series Z_1 and Z_{23456} impedance's into the single final impedance, Z_{Total} . The completed impedance diagram is presented in Figure 2-43.

FIGURE 2-43 FINAL REDUCTION OF COMPLEX IMPEDANCE SYSTEM



Not all electrical circuit components have an impedance, for example, an ideal voltage source does not have a fixed impedance. Although the voltage value is constant, the current is determined by the circuit to which the source is connected, making the impedance a variable. The same is true for an ideal current source.

2.4.3 Modified Analogy Approach

The modified analogy approach is a process which allows you to convert an illustration of a physical system to a block diagram model. The approach is based on the electrical notion of impedance and a four-step conversion process explained in this section.

The difference between the modified analogy approach and the basic analogy approach is the manner in which nonlinearities are handled. The basic analogy approach presented in many texts is restricted to linear applications. If a nonlinearity exists, it must be linearized prior to incorporating it into the model. Linearization provides only an approximation to the behavior of the nonlinearity; the difference between the linearized and actual behavior becomes an undesirable modeling error. The modified approach removes this limitation by allowing the actual nonlinearity to be incorporated into the model. This results in a more accurate model with better predictive capability and less modeling error.

Given a system illustration, analogies are first established for the PV and FV . Once the analogies have been established, the following four-step procedure is applied to obtain the block diagram model.

Step 1. Create and (if possible) simplify the impedance diagram using the manipulations presented in Table 2-2. Simplifications of this nature include minor parallel and series branches which can be easily reduced to single equivalent branches.

Step 2. Circle all nodes (*FV* and *PV*) in the impedance diagram and label all signals entering and leaving these nodes. A *FV* node is a point in the impedance diagram where three or more branches intersect. A *PV* node occurs when two or more impedance elements exist in series. The *PV* node relates the individual *PV* drops of the elements to a single overall *PV* drop.

Step 3. Construction of the block diagram is initiated by representing select nodes (*PV* and *FV*) from the previous step as summing junctions with inputs and outputs labeled according to signals from the impedance diagram. In general, it usually is not necessary to implement all *PV* and *FV* nodes, because often they are dependent upon one another. Select the output of each summing junction such that, when it is applied to the corresponding impedance block, a causal operation (either an integration or multiplication by a gain) results.

For example, an element with impedance $Z = D \cdot L = \frac{PV}{FV}$ where ($D \equiv d(\cdot)/dt$) must have *PV* as input to have integral causality. Similarly, an element with impedance $Z = \frac{1}{D \cdot C} = \frac{PV}{FV}$ must have *FV* as input to have integral causality.

It should be noted that in some situations it will not be possible to create a block diagram with only gain or integral causality. In these situations, we either attempt to differentiate the noncausal elements directly or modify the model to achieve causality.

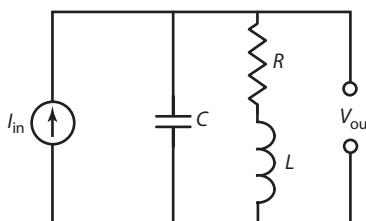
Step 4. The block diagram is completed by placing each component impedance from the impedance diagram onto the block diagram and connecting them with signals from either summing junctions or other impedances. Other intermediate, input, and output signals necessary to complete the block diagram are also added during this step.

This procedure is somewhat complicated and best illustrated through examples. Throughout the remainder of this chapter, we will apply this procedure in each example to illustrate the steps involved in the construction of the block diagram. As you become more familiar with the procedure and gain experience, you may find it easier to go directly from the illustration of the system to the block diagram without drawing the intermediate impedance diagram at all.

EXAMPLE 2.11 Block Diagram Construction—Parallel Resonant Electrical Circuit

The parallel resonant circuit exhibits a controllable resonant peak suitable for notch filtering applications. Notch filters are used to remove unwanted frequencies from a signal leaving the other frequencies unaltered. The parallel resonant circuit diagram with the resistance lumped in the inductor branch is presented in Figure 2-44.

FIGURE 2-44 PARALLEL RESONANT CIRCUIT



The impedance variables are chosen as *FV* = current and *PV* = voltage.

Solution**Step 1.** Create/simplify the impedance diagram.

The impedance's of the circuit elements are summarized as

$$\text{Capacitor: } Z_C = \frac{1}{j\omega C} \quad \text{Inductor: } Z_L = j\omega L \quad \text{Resistor: } Z_R = R$$

Substituting these impedances into the original circuit produces the impedance diagram shown in Figure 2-45. Using D as the time differentiation operator ($D \equiv d(\cdot)/dt$), the impedance diagram is rewritten in operator notation in Figure 2-46.

FIGURE 2-45 PARALLEL RESONANT CIRCUIT IMPEDANCE DIAGRAM

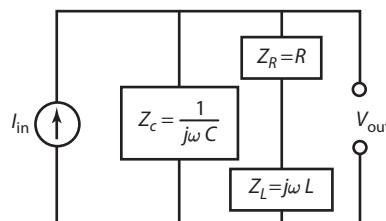
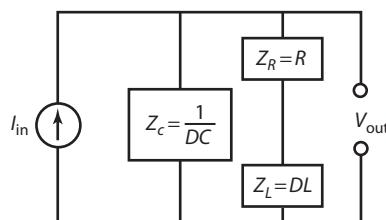


FIGURE 2-46 PARALLEL RESONANT CIRCUIT IMPEDANCE DIAGRAM USING D OPERATOR

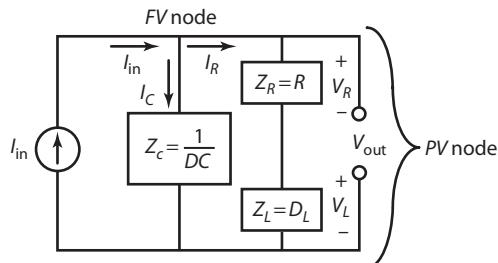
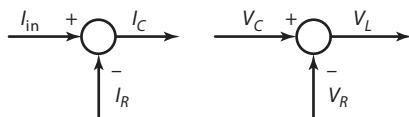
**Step 2.** Identify all independent nodes (FV and PV) in the impedance diagram and label all signals.

A FV node is a point in the impedance diagram where three or more branches intersect. A PV node relates the individual PV drops over a series of impedance's to an overall PV drop. Our diagram has one FV node and one PV node as shown in Figure 2-47.

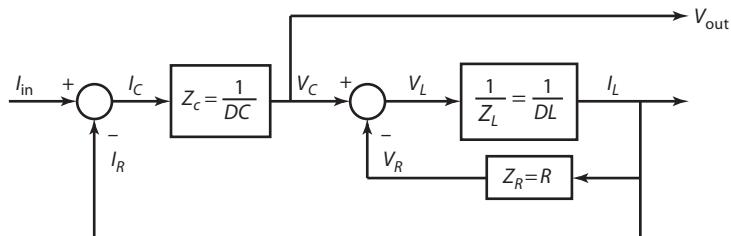
Step 3. Represent select nodes as a summing junction, and select the output of the summing junction such that (when it is connected to its associated impedance blocks) either gain or integral causality results.

The two nodes in our impedance diagram produce the two summing junctions shown in Figure 2-48. We have arbitrarily selected the summing junction output in step 3. If we encounter causality problems in step 4, we may need to modify either or both of these summing junctions.

Step 4. Add the impedance blocks; connect and create all necessary intermediate and output signals to complete the block diagram.

FIGURE 2-47 NODES IN THE PARALLEL RESONANT CIRCUIT IMPEDANCE DIAGRAM**FIGURE 2-48 PARTIAL BLOCK DIAGRAM REPRESENTATION OF THE PARALLEL RESONANT CIRCUIT**

Noting that $I_R = I_L$ and that $V_{\text{out}} = V_C$, the block diagram is constructed by first adding the three impedance blocks. Next, the appropriate signal connections are made using wires. Luckily, we have selected the summing junction outputs which provide integral causality, so no modifications are needed in step 3. The completed block diagram is presented in Figure 2-49.

FIGURE 2-49 COMPLETED BLOCK DIAGRAM REPRESENTATION OF THE PARALLEL RESONANT CIRCUIT

The system equations can be derived by simplifying the block diagram. For example, the transfer function relating the input current to the output voltage is presented in Equation 2-2.

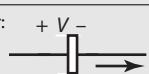
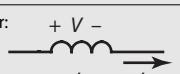
$$V_{\text{out}} = \frac{DL + R}{D^2LC + DRC + 1} I_{\text{in}} \text{ or } \ddot{V}_{\text{out}}LC + \dot{V}_{\text{out}}RL + V_{\text{out}} = I_{\text{in}}L + I_{\text{in}}R \quad (2-2)$$

2.5 Electrical Systems

Electrical circuits rely on two variables, voltage and current, to transport energy. Since current flows through an electrical circuit, it is natural to associate current with the flow variable and voltage with the potential variable. Using this convention, the impedances of six basic ideal circuit components

are discussed: the resistor, capacitor, inductor, voltage source, current source, and transformer. The impedances of these components will provide the fundamental analogies for components in other disciplines. Of the six basic electrical components, only the resistor, capacitor, and inductor have impedance's which are not functions of the circuit to which they are attached. The resistor, capacitor, and inductor impedance characteristics are summarized in Table 2-3.

TABLE 2-3 RESISTOR, CAPACITOR, AND INDUCTOR IMPEDANCES

Analogy		Component		
$PV =$ Voltage, v	$FV =$ Current, i	<p>Resistor:</p>  $\Rightarrow Z_R = R$	<p>Capacitor:</p>  $\Rightarrow Z_C = \frac{1}{CD}$	<p>Inductor:</p>  $\Rightarrow Z_L = LD$

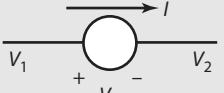
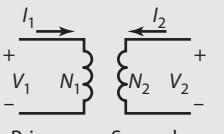
The remaining three components have impedances which are functions of the circuit to which they are attached. The ideal voltage source is used to create a specified potential at any point in a circuit. The potential exists between the two terminals of the voltage source. The current which passes through the voltage source is determined by the circuit to which the source is connected. Due to the current being an unknown, it is not possible to write the impedance relationship for the voltage source without knowledge of the rest of the circuit. Sometimes the voltage value for the source will be a function of another variable of the circuit (such as a current or voltage). In this situation, the voltage source is called *dependent*, since its value is dependent on another signal in the circuit.

The ideal current source is used to create a specified current at any point in a circuit. The voltage which exists between the two terminals of the current source is determined by the circuit to which the source is connected. Due to the voltage being an unknown, it is not possible to write the impedance relationship for the current source without knowledge of the rest of the circuit. Similar to the voltage source, sometimes the value for the current source will be a function of another variable of the circuit (such as a current or voltage). In this situation, the current source is called *dependent*, since its value is dependent on another signal in the circuit.

A transformer is a magnetically coupled electrical device consisting of two coils wound along each side of a closed conducting core. One winding is called the *primary* (winding 1) and the other winding called the *secondary* (winding 2). The number of windings in the primary and secondary coils are N_1 and N_2 , respectively. The impedance characteristics of the ideal transformer are dependent on the circuit to which it is connected. The impedance characteristics of the voltage source, current source, and transformer are presented in Table 2-4.

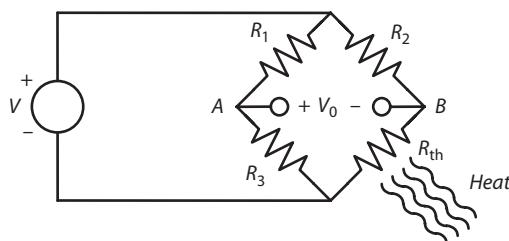
To illustrate how the analogy approach is applied to electrical circuits to create block diagrams, two examples are presented: a bridge circuit and a transformer circuit. Bridges can be constructed entirely of resistors or capacitors depending on the quantity being measured. The transformer is an important electric circuit component, because (as will be seen later) it is analogous to gear trains in mechanical rotational systems and lever arms in mechanical translation systems. Transformers have many applications, including impedance matching, voltage step up, and voltage step down. Electric power-transmission systems rely heavily on step-up and step-down transformers to efficiently send electricity over large distances.

TABLE 2-4 VOLTAGE SOURCE, CURRENT SOURCE, AND TRANSFORMER IMPEDANCES

Component	Impedance Relationship
 Voltage source	Defining Equations: $V = V_1 - V_2$ $I = f_1(\text{Attached circuit})$ Impedance: $Z_{VS} = f_2(\text{Attached circuit})$
 Current Source	Defining Equation: $I = \text{Specified current}$ $V_1, V_2 = f_1(\text{Attached circuit})$ Impedance: $Z_{CS} = f_2(\text{Attached circuit})$
 Transformer	Defining Equations: $\frac{V_2}{V_1} = \frac{N_2}{N_1}$ and $\frac{I_1}{I_2} = \frac{N_2}{N_1}$ Impedance: $Z_T = f(\text{Attached circuit})$

EXAMPLE 2.12 Bridge Circuit System

A thermistor is a semiconductor device whose resistance changes with temperature. Temperature readings in terms of voltage can be obtained by installing the thermistor as one of the resistances in a bridge circuit. A typical configuration is shown in Figure 2-50.

FIGURE 2-50 BRIDGE CIRCUIT FOR TEMPERATURE MEASUREMENT

When a constant voltage is applied to the circuit, V , heat source variations cause the thermistor resistance to change, thus creating a potential difference between points A and B, which is proportional to temperature.

The objective of this example is to apply the analogy method to develop a block diagram model of the bridge circuit.

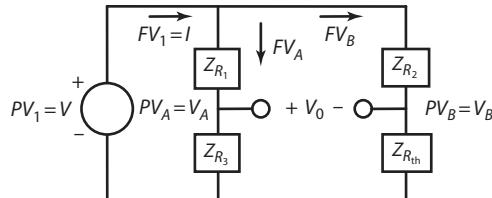
Solution

Step 1. Create/simplify the impedance diagram.

The first step of the procedure is the construction of the impedance diagram. This is relatively straightforward. All flow paths, potentials, and branches remain intact; the only difference is

the replacement of each component with its associated impedance. The impedance diagram is presented in Figure 2-51.

FIGURE 2-51 IMPEDANCE DIAGRAM FOR TEMPERATURE MEASUREMENT CIRCUIT



Step 2. Identify all independent nodes (*FV* and *PV*) in the impedance diagram and label all signals.

The impedance diagram has one *FV* node and two *PV* nodes. The node equations are given as

$$FV \text{ node equation: } FV_1 = FV_A + FV_B$$

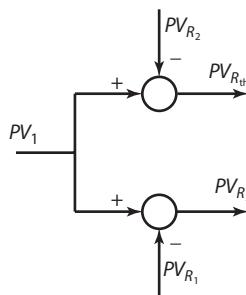
$$PV \text{ node 1 equation: } PV_1 = PV_{R_1} + PV_{R_3} \text{ (note that } PV_A = PV_{R_3})$$

$$PV \text{ node 2 equation: } PV_1 = PV_{R_2} + PV_{R_{th}} \text{ (note that } PV_B = PV_{R_{th}})$$

Step 3. Represent select nodes as a summing junction, and select the output of the summing junction such that (when it is connected to its associated impedance blocks) either gain or integral causality results.

The initial block diagram is constructed with two summing junctions to model the two *PV* nodes, Figure 2-52.

FIGURE 2-52 SUMMING JUNCTIONS FOR TEMPERATURE MEASUREMENT CIRCUIT BLOCK DIAGRAM

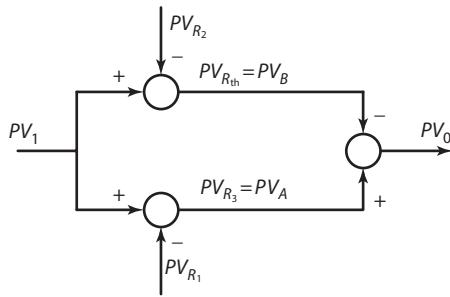


Next, the diagram is slightly modified to include the definitions $PV_A \equiv PV_{R_3}$, $PV_B \equiv PV_{R_{th}}$, and $V_0 = PV_0 \equiv PV_A - PV_B$. These additions are presented in Figure 2-53.

Step 4. Add the impedance blocks; connect and create all necessary intermediate and output signals to complete the block diagram.

The *FV* node equation was not directly implemented using a summing junction; however, since Z_{R_1} and Z_{R_3} both have the same flow, FV_A , and since Z_{R_2} and $Z_{R_{th}}$ have FV_B flowing through them, the following two constraint relationships are written.

FIGURE 2-53 SUMMING JUNCTIONS FOR TEMPERATURE MEASUREMENT CIRCUIT BLOCK DIAGRAM WITH SLIGHT MODIFICATION



$$FV_A = \frac{PV_{R_1}}{Z_{R_1}} = \frac{PV_A}{Z_{R_3}} \Rightarrow PV_{R_1} = \frac{Z_{R_1}}{Z_{R_3}} \cdot PV_A$$

$$FV_B = \frac{PV_{R_2}}{Z_{R_2}} = \frac{PV_B}{Z_{R_{th}}} \Rightarrow PV_{R_2} = \frac{Z_{R_2}}{Z_{R_{th}}} \cdot PV_B$$

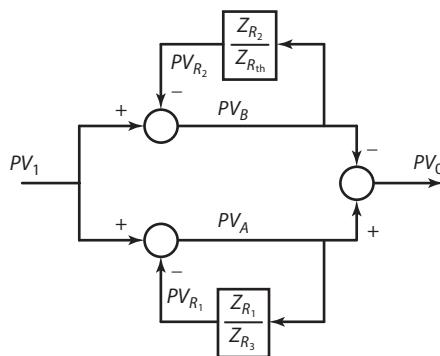
The final block diagram, Figure 2-54, is constructed by adding these two relationships to the block diagram to define the PV_{R_1} and PV_{R_2} signals. From the revised block diagram, the system equations may be derived after substituting the appropriate resistance values and noting that $V = PV_1$, $V_A = PV_A$, and $V_B = PV_B$, we have

$$V_A = \frac{R_3}{R_1 + R_3} V$$

$$V_B = \frac{R_{th}}{R_2 + R_{th}} V$$

$$\text{Potential difference A - B: } V_{AB} = V_0 = \left(\frac{R_3}{R_1 + R_3} - \frac{R_{th}}{R_2 + R_{th}} \right) V$$

FIGURE 2-54 BLOCK DIAGRAM FOR TEMPERATURE MEASUREMENT CIRCUIT

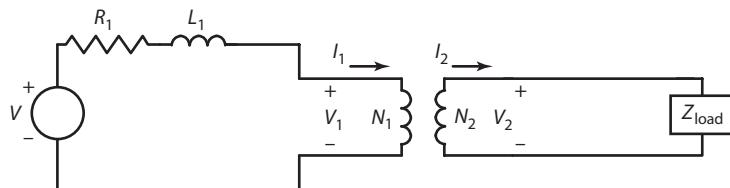


As written, the system equation represents the output voltage as a function of the thermistor resistance and the input voltage; $V_0 = V_0(R_{\text{th}}, V)$. With a constant input voltage, the output voltage becomes only a function of the thermistor resistance.

EXAMPLE 2.13 Transformer System

The basic transformer circuit with input, V , and output, i_2 , is shown in Figure 2-55.

FIGURE 2-55 BASIC TRANSFORMER CIRCUIT



Voltage, V_1 , is applied to the transformer primary side coil which consists of a series resistance, R_1 and L_1 . The secondary side coil of the transformer consists of a load impedance, Z_{load} .

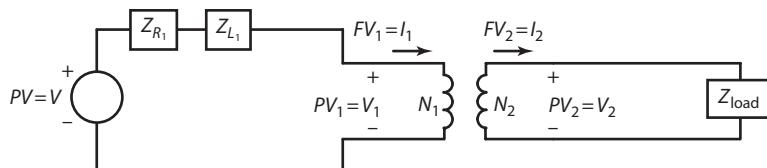
Again, the objective of this example is to develop the block diagram model for the transformer circuit.

Solution

Step 1. Create/simplify the impedance diagram.

The impedance diagram for the transformer is created by replacing each element of the circuit with its associated impedance. The impedance diagram is presented in Figure 2-56.

FIGURE 2-56 BASIC TRANSFORMER IMPEDANCE DIAGRAM



Step 2. Identify all independent nodes (FV and PV) in the impedance diagram and label all signals.

The impedance diagram has two PV nodes which represent the potential drops around the primary winding and secondary winding loops. These equations are summarized here.

$$\text{Primary winding loop equation: } PV - PV_{R_1} - PV_{L_1} - PV_1$$

$$\text{Secondary winding loop equation: } PV_2 - Z_{\text{load}} \cdot FV_2$$

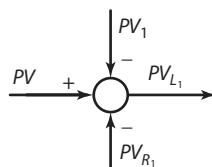
In addition to the loop equations, the auxiliary transformer equations which relate PV and FV across the transformer ratio are

$$PV_1 = \frac{N_1}{N_2} PV_2 \text{ and } FV_1 = \frac{N_2}{N_1} FV_2$$

Step 3. Represent select nodes as a summing junction and select the output of the summing junction such that (when it is connected to its associated impedance blocks) either gain or integral causality results.

The block diagram construction is initiated with the primary winding loop PV equation and presented in Figure 2-57.

FIGURE 2-57 BLOCK DIAGRAM BASED ON PRIMARY WINDING LOOP EQUATION



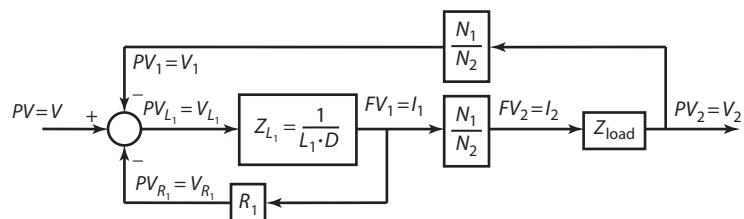
The signal PV_1 is computed from the transformer equation as $PV_1 = \frac{N_1}{N_2} PV_2$. The primary flow is computed using the causal impedance relationship for Z_{L_1} , which produces FV_1 . Using the transformer equation, this primary flow is converted to a secondary flow as

$$FV_2 = \frac{N_1}{N_2} FV_1$$

Step 4. Add the impedance blocks; connect and create all necessary intermediate and output signals to complete the block diagram.

The block diagram is completed by incorporating these definitions and presented in Figure 2-58.

FIGURE 2-58 BASIC TRANSFORMER BLOCK DIAGRAM



We have assumed that the load impedance has current causality in the formulation. If this were not the case, for example, if it had voltage causality, the diagram would need to be modified.

Depending on the desired output, many system relationships can be computed from the block diagram. For example, Equation (2-3) relates input voltage to secondary current and is computed as

$$\frac{I_2}{V} = \frac{(N_1/N_2)Z_{\text{load}}}{L_1D + R_1 + (N_1/N_2)^2Z_{\text{load}}} \quad (2-3)$$

2.6 Mechanical Translational Systems

Mechanical systems can be either translational or rotational. Although the fundamental relationships for both types are derived from Newton's law, they are different enough to warrant separate consideration.

Mechanical translation system analysis is based on Newton's law, which states:

The vector sum of all forces applied to a body equals the product of the vector acceleration of the body times its mass.

The equation for Newton's law is presented in Equation 2-4.

$$F = Ma \quad (2-4)$$

where the units in the British system are

F = total force, newtons, N

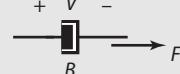
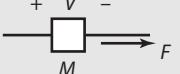
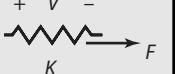
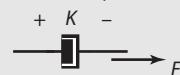
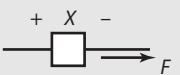
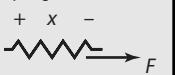
M = mass, kg

a = total acceleration, $\frac{\text{m}}{\text{s}^2}$

Two elements typically encountered in mechanical systems are the linear damper and the linear spring. The linear damper produces a force proportional to the applied velocity, and the linear spring produces a force proportional to the applied displacement.

Depending on the system, either velocity or displacement may be used as the PV . Regardless of the choice of PV , force is used for the FV . Table 2-5 summarizes the impedance's of the three mechanical translation system components for both analogies.

TABLE 2-5 MECHANICAL SYSTEM IMPEDANCE ANALOGIES

Analogy		Component		
$PV =$ Velocity, v	$FV =$ Force, F	Viscous damper:  $\Rightarrow Z_B = \frac{1}{B}$	Mass:  $\Rightarrow Z_M = \frac{1}{MD}$	Spring:  $\Rightarrow Z_K = \frac{D}{K}$
$PV =$ Displacement, x	$FV =$ Force, F	Viscous damper:  $\Rightarrow Z_B = \frac{1}{DB}$	Mass:  $\Rightarrow Z_M = \frac{1}{MD^2}$	Spring:  $\Rightarrow Z_K = \frac{1}{K}$

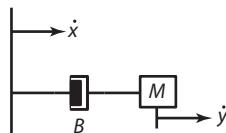
In the remainder of this section, several examples are presented illustrating how the analogy approach is applied to mechanical translational systems to develop a block diagram model.

EXAMPLE 2.14 Mass–Damper System

The basic mass–damper system is modeled in this example. Selection of logical *PV* and *FV* variables will create a causality problem which is also discussed.

An illustration of the mass–damper system is shown in Figure 2-59. Since the input, \dot{x} , and output, \dot{y} , of the system are both velocities and no springs are involved, velocity is the logical choice for the potential variable. The flow variable is force.

FIGURE 2-59 MASS–DAMPER SYSTEM ILLUSTRATION



Solution

Step 1. Create/simplify the impedance diagram.

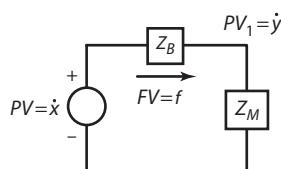
The impedance diagram for the mass–damper system is created by replacing each element of the circuit with its associated impedance. The impedances are defined as

$$Z_B = \frac{1}{B}$$

$$Z_M = \frac{1}{MD}$$

The impedance diagram is presented in Figure 2-60.

FIGURE 2-60 MASS–DAMPER SYSTEM IMPEDANCE DIAGRAM



Step 2. Identify all independent nodes (*FV* and *PV*) in the impedance diagram and label all signals.

The impedance diagram consists of one *PV* node represented by the following equation.

$$PV - PV_{ZB} - PV_{ZM} = 0$$

The three auxiliary equations are also required.

$$PV_{ZB} = Z_B \cdot FV$$

$$PV_{ZM} = Z_M \cdot FV$$

$$PV_{ZM} = PV_1$$

Step 3. Represent select nodes as a summing junction and select the output of the summing junction such that (when it is connected to its associated impedance blocks) either gain or integral causality results.

Integral causality for the Z_M element requires that FV be its input. Our strategy is to model the PV node equation such that PV_{ZB} is the output. The damper, which has no causality problems because the potential variable is velocity, is used to create the FV required as input to the Z_M block.

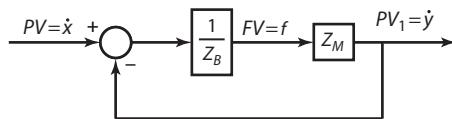
Step 4. Add the impedance blocks; connect and create all necessary intermediate and output signals to complete the block diagram.

The resulting block diagram is presented in Figure 2-61.

The output velocity, \dot{y} , is computed by reducing the block diagram and substituting for the two impedances as

$$\dot{y} = \frac{Z_M}{Z_M + Z_B} \dot{x} = \frac{B}{MD + B} \dot{x}$$

FIGURE 2-61 MASS-DAMPER BLOCK DIAGRAM



The force flowing through the system, FV , may also be computed from the block diagram as

$$f = \frac{\dot{x} - \dot{y}}{Z_B} = (\dot{x} - \dot{y})B$$

One also could solve this problem using displacement instead of velocity as the potential variable. The input and output variables become x and y . Since displacement is the integral of velocity and integration is represented in operator notation as $\frac{1}{D}$, the impedances in the displacement–voltage analogy system are equivalent to the impedances of the velocity–voltage system multiplied by $\frac{1}{D}$. These impedances become

$$Z_B = \frac{1}{BD} \text{ and } Z_M = \frac{1}{MD^2}$$

Because the system is linear, the transfer function relating \dot{x} to \dot{y} is

$$\dot{y} = \frac{B}{MD + B} \dot{x}$$

We can compute the transfer function from x to y by integrating both sides. This is analogous to division by the D operator. The resulting transfer function becomes

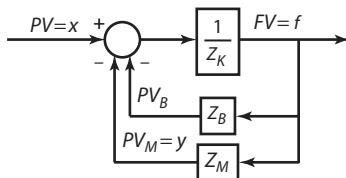
$$y = \frac{B}{MD + B} x$$

This is no surprise, however. Suppose we were confronted with the task of modeling the system with displacement used as the potential variable. Causality now becomes an issue. For integral causality, both elements Z_B and Z_M must have an FV input. Investigation of the PV node equation for this system reveals that this is not possible; however, all is not lost. We recognize that the real problem is that the only causality independent element capable of converting a PV to an FV signal in this situation is the spring, which is not present in our diagram.

We can solve this problem using an approximate system which includes an additional spring with its stiffness set to a very large value. The approximate system will be of integral causality and will approximate the actual response closer and closer as the spring stiffness is increased. Setting this limit, the original transfer function will result.

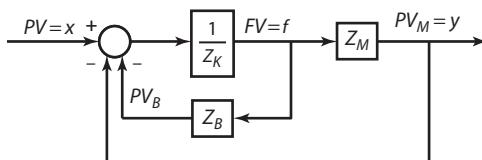
The approximate system block diagram is presented in Figure 2-62. The added spring is placed just to the right of the PV node summing junction to produce the required FV output.

FIGURE 2-62 APPROXIMATE SYSTEM BLOCK DIAGRAM



Since we are interested in computing the system transfer function from x to y , it is beneficial to redraw the block diagram before any reductions are performed, as in Figure 2-63.

FIGURE 2-63 REDRAWN APPROXIMATE SYSTEM BLOCK DIAGRAM



Since displacement is the PV , the impedance's are

$$Z_K = \frac{1}{K}, \quad Z_B = \frac{1}{BD}, \quad \text{and} \quad Z_M = \frac{1}{MD^2}$$

Reducing the block diagram and substituting the impedance relationships yields the following transfer function.

$$y = \frac{\frac{KBD}{MD^2}}{1 + BD + K} x = \frac{KB}{MBD^2 + KMD + KB} x = \frac{B}{\frac{MB}{K} D^2 + MD + B} x$$

As the spring stiffness is made very large, the transfer function approaches the expected transfer function as

$$y = \frac{B}{MD + B} x$$

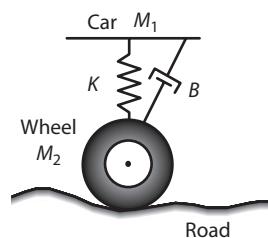
Problems of this nature are often found in real systems and with proper attention, integral causality can be maintained.

EXAMPLE 2-15 Automobile Suspension System

The suspension system of a car can be modeled on a per-wheel basis as a two-mass system: the car mass and the wheel mass. The tire behaves as a spring, and the connection between the tire and the car is a spring-shock absorber (damper) assembly. The road roughness provides the input to the system as a displacement. The outputs are the axle displacement and the vehicle displacement.

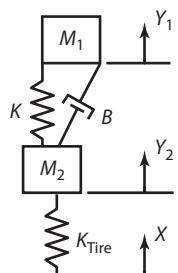
An illustration of the suspension system is shown in Figure 2-64.

FIGURE 2-64 SUSPENSION SYSTEM ILLUSTRATION



The mechanical diagram is shown in Figure 2-65. Since the input and output signals are displacements, displacement is selected as the potential variable and force as the flow variable.

FIGURE 2-65 SUSPENSION MECHANICAL DIAGRAM



Solution

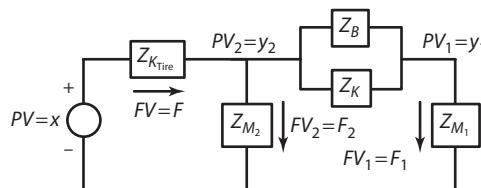
Step 1. Create/simplify the impedance diagram.

The impedances that will be used in the impedance diagram are listed here.

$$\begin{aligned} Z_{K_{\text{tire}}} &= \frac{1}{K_{\text{tire}}} \\ Z_{M_1} &= \frac{1}{M_1 D^2} \quad Z_{M_2} = \frac{1}{M_2 D^2} \\ Z_B &= \frac{1}{BD} \quad Z_K = \frac{1}{K} \\ \frac{1}{Z_{KB}} &= BD + K \end{aligned}$$

The impedance diagram is presented in Figure 2-66, and forces also have been labeled. It can be clearly seen how much force is drawn by each of the two masses. This feature of the impedance diagram is especially useful when losses need to be calculated.

FIGURE 2-66 SUSPENSION SYSTEM IMPEDANCE DIAGRAM



Step 2. Identify all independent nodes (FV and PV) in the impedance diagram and label all signals.

The impedance diagram may be reduced by first combining the parallel spring-damper into an equivalent impedance defined as Z_{KB} . With this reduction, the impedance diagram has one FV node at y_2 and two PV nodes over $Z_{K_{\text{tire}}}$ and Z_{KB} . The node equations are summarized here.

$$\text{FV node at } y_2: FV - FV_1 - FV_2 = 0$$

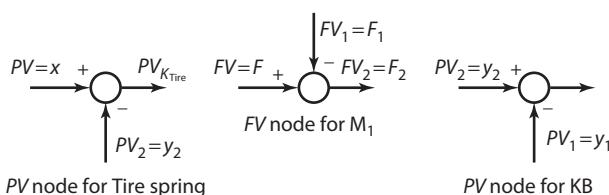
$$\text{PV node for } Z_{K_{\text{tire}}}: PV - PV_2 = PV_{K_{\text{tire}}}$$

$$\text{PV node for } Z_{KB}: PV_2 - PV_1 = PV_{KB}$$

Step 3. Represent select nodes as a summing junction and select the output of the summing junction such that (when it is connected to its associated impedance blocks) either gain or integral causality results.

These summing junction representations of the node equations are shown in Figure 2-67.

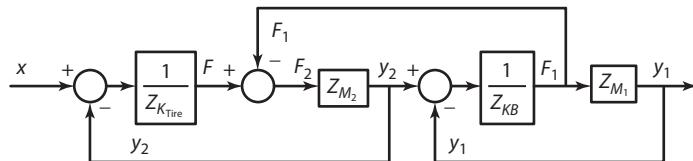
FIGURE 2-67 SUSPENSION SYSTEM BLOCK DIAGRAM SUMMING JUNCTIONS



Step 4. Add the impedance blocks; connect and create all necessary intermediate and output signals to complete the block diagram.

For integral causality, the inputs to the Z_{M_1} and Z_{M_2} blocks must be *FV* signals. Since only one *FV* node equation is present, we must use the Z_{KB} block to produce the additional *FV* signal required for Z_{M_2} . For brevity, the general *PV* and *FV* notation is dropped, and the completed block diagram shown in Figure 2-68 uses the problem variables.

FIGURE 2-68 SUSPENSION SYSTEM BLOCK DIAGRAM



The system equations may be derived by manipulating the block diagram. Several transfer functions are computed and presented in Table 2-6.

TABLE 2-6 TRANSFER FUNCTIONS FROM BLOCK DIAGRAM MANIPULATION

y_1 from y_2 :	$y_1 = \frac{BD + K}{M_1 D^2 + BD + K} y_2$
y_2 from x :	$y_2 = \frac{K_{\text{tire}}(M_1 D^2 + BD + K)}{M_1 M_2 D^4 + (M_1 + M_2)BD^3 + [(M_1 + M_2)K + M_1 K_{\text{tire}}]D^2 + (BD + K)K_{\text{tire}}} x$
Wheel mass force, F_1 :	$F_1 = \frac{y_1}{Z_{M_1}}$
Car mass force, F_2 :	$F_2 = \frac{y_2}{Z_{M_2}}$
y_1 can be computed directly from x by multiplying the two transfer functions:	$\begin{aligned} \frac{y_1}{x} &= \frac{y_1}{y_2} \frac{y_2}{x} \\ &= \left(\frac{K_{\text{tire}}(BD + K)}{M_1 M_2 D^4 + (M_1 + M_2)BD^3 + [(M_1 + M_2)K + M_1 K_{\text{tire}}]D^2 + (BD + K)K_{\text{tire}}} \right) \end{aligned}$

The system equations represented as differential equations are listed here as

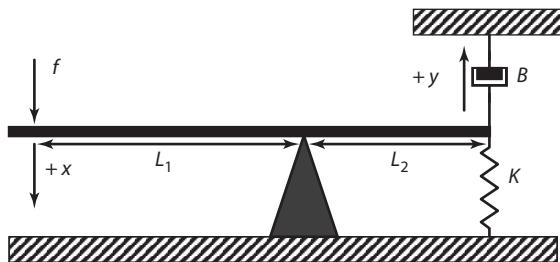
$$\ddot{Y}_2 M_1 M_2 + \ddot{Y}_2 (M_1 + M_2)B + \ddot{Y}_2 [(M_1 + M_2)K + M_1 K_{\text{tire}}] + \ddot{Y}_2 B K_{\text{tire}} + Y_2 K K_{\text{tire}} = \dot{X} M_1 K_{\text{tire}} + \dot{X} B K_{\text{tire}} + X K K_{\text{tire}}$$

and

$$\ddot{Y}_1 M_1 M_2 + \ddot{Y}_1 (M_1 + M_2)B + \ddot{Y}_1 [(M_1 + M_2)K + M_1 K_{\text{tire}}] + \ddot{Y}_1 B K_{\text{tire}} + Y_1 K K_{\text{tire}} = \dot{X} B K_{\text{tire}} + X K K_{\text{tire}}$$

EXAMPLE 2-16 Mechanical Lever System

This final example illustrates the application of the transformer analogy to a mechanical system which utilizes a lever arm. An illustration of the lever system is presented in Figure 2-69.

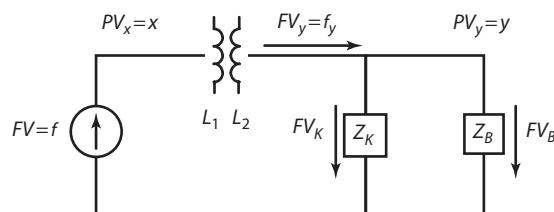
FIGURE 2-69 MECHANICAL LEVER SYSTEM EXAMPLE

An input force, f , is applied to one end of a lever arm resulting in a vertical deflection, x . The arrow directions signify the positive direction of all signals. The location of the lever arm pivot is selected to produce a force amplification, which is then applied to a spring–damper load connected to ground.

Solution

Step 1. Create/simplify the impedance diagram.

The impedance diagram using displacement as the potential variable is presented in Figure 2-70.

FIGURE 2-70 MECHANICAL LEVER SYSTEM IMPEDANCE DIAGRAM

The transformer relates the PV from the primary to the secondary by the relationship

$$PV_2 = \frac{N_2}{N_1} PV_1$$

where N_1 and N_2 are analogous to the lever ratios, L_1 and L_2 , respectively.

Since displacement is the potential variable, the impedances of the spring and damper are

$$Z_K = \frac{1}{K} \quad \text{and} \quad Z_B = \frac{1}{BD}$$

Step 2. Identify all independent nodes (FV and PV) in the impedance diagram and label all signals.

The impedance diagram has one FV node whose equation is

$$FV_y - FV_K - FV_B = 0$$

The auxiliary equation for the lever ratio is

$$FV_y = \frac{L_1}{L_2} FV$$

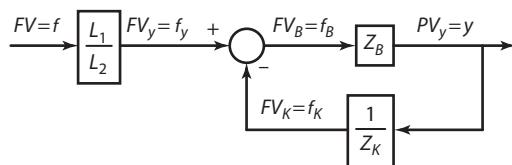
Step 3. Represent select nodes as a summing junction and select the output of the summing junction such that (when it is connected to its associated impedance blocks) either gain or integral causality results.

For integral causality, the Z_B block must have FV as input. Since the Z_K block has no causality, restrictions of the FV node equation summing junction should be constructed to have FV_B as its output.

Step 4. Add the impedance blocks; connect and create all necessary intermediate and output signals to complete the block diagram.

Adding the damper and spring impedances along with the lever ratio to the summing junction created in step 3 produces the final block diagram shown in Figure 2-71.

FIGURE 2-71 MECHANICAL LEVER SYSTEM BLOCK DIAGRAM



Going a step further, we can manipulate the block diagram to compute some internal characteristics of the system. For example, the force applied to the load, f_y , is computed using the transformer relationship:

$$f \ L_1 = f_y \ L_2 \Rightarrow f_y = \frac{L_1}{L_2} f$$

The vertical displacement at the load, y , is computed by computing the closed loop transfer function as

$$y = \frac{\frac{L_1}{L_2}}{BD + K} f$$

The vertical displacement at the source is computed from the displacement at the load using the transformer characteristic:

$$x \ L_2 = y \ L_1 \Rightarrow x = \frac{L_1}{L_2} y = \frac{\frac{L_1^2}{L_2}}{BD + K} f$$

The overall system equations relating input force to load force and displacement are

$$f_y = \frac{L_1}{L_2} f \text{ and } \dot{y}B + yK = \frac{L_1}{L_2} f$$

In more complex applications, the linear spring and damper models used thus far may not provide sufficient accuracy to describe the overall system behavior. In these situations, we may resort to nonlinear models for these components.

2.7 Mechanical Rotational Systems

Mechanical rotational system analysis also is based on Newton's Law; however, the law is slightly modified to account for rotation instead of translation. The law states:

The vector sum of all moments applied to a body equals the product of the vector angular acceleration of the body times its inertia.

A rotational system obeys Equation 2-5.

$$T = J\ddot{\theta} \quad (2-5)$$

In the SI system, the units are defined as

T = total torque, N-m

J = body inertia about its center of mass, kg-m²

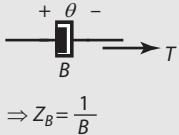
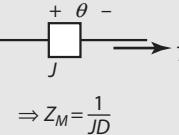
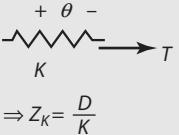
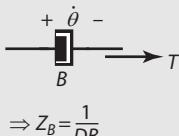
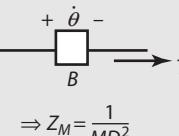
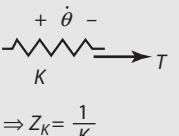
$\ddot{\theta}$ = angular acceleration, $\frac{\text{rad}}{\text{s}^2}$

Two elements typically encountered in mechanical rotational systems are the linear torsional damper and the linear torsional spring. The damper produces a torque proportional to the applied angular velocity, and the spring produces a torque proportional to the applied angle.

An analogy similar to that used for translation systems exists for rotational systems—except angle replaces displacement, angular speed replaces velocity, and torque replaces force. Also, mass becomes inertia, the translational spring constant becomes a torsional spring constant, and translational damping becomes rotational damping. The impedance analogies are identical in form to those used in translational systems. The flow variable is defined as torque, and the potential variable is defined as either angular velocity or angle.

The analogies and impedances for rotational systems are summarized in Table 2-7.

TABLE 2-7 IMPEDANCE ANALOGIES FOR ROTATIONAL SYSTEMS

Analogy		Component		
$PV = \text{Velocity}, \dot{\theta}$	$FV = \text{Torque}, T$	Damper:  $\Rightarrow Z_B = \frac{1}{B}$	Inertia:  $\Rightarrow Z_M = \frac{1}{JD}$	Spring:  $\Rightarrow Z_K = \frac{D}{K}$
$PV = \text{Displacement}, \theta$	$FV = \text{Torque}, T$	Damper:  $\Rightarrow Z_B = \frac{1}{DB}$	Mass:  $\Rightarrow Z_M = \frac{1}{MD^2}$	Spring:  $\Rightarrow Z_K = \frac{1}{K}$

The SI Units used for mechanical rotational systems are summarized in Table 2-8.

TABLE 2-8 ROTATIONAL SYSTEM UNITS

System	T	$\dot{\theta}$	K	B	J
SI	nt – meter	$\frac{\text{rad}}{\text{sec}^2}$	$\frac{\text{kg} - \text{meter}^2}{\text{sec}^2}$	$\frac{\text{kg} - \text{meter}^2}{\text{sec}}$	$\text{kg} - \text{meter}^2$

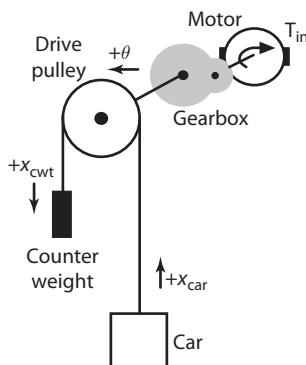
The remainder of this section presents an example modeling application for a complex rotational system which illustrates how gear ratios as well as inertias, springs, and gravity forces are modeled in the mechanical rotational discipline. The steps are very similar to those used for modeling mechanical translation systems.

EXAMPLE 2-17 Elevator System

A cable-driven elevator hoistway system consists of a drive pulley (drive sheave) attached to a gearbox powered by an electric motor. The drive sheave is wrapped (usually six or more times to prevent slippage) with a cable—one end of which is attached to a counterweight and the other end to the elevator cab.

An illustration of the elevator hoistway system is shown in Figure 2-72. The cable is assumed to act as a spring with no damping. For modeling, the cable weight on either side of the pulley is halved. One half is lumped as part of the pulley weight, and the other half lumped into the car weight and counterweight, respectively.

FIGURE 2-72 GEARED ELEVATOR HOISTWAY SYSTEM ILLUSTRATION



The radius of the drive sheave is designated as r , and the gear ratio as $1:N$ (N motor revolutions to 1 drive sheave revolution).

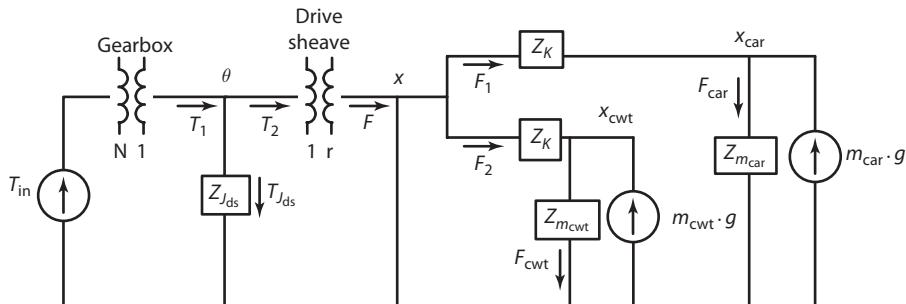
Since the hoistway system contains springs, the logical choice for the potential variable is displacement.

Solution

Step 1. Create/simplify the impedance diagram.

The following impedance diagram (Figure 2-73) is constructed.

FIGURE 2-73 GEARED ELEVATOR HOISTWAY SYSTEM IMPEDANCE DIAGRAM



The force due to gravity has been included on both the counterweight and the car and on the direction results from the definition of the car and counterweight directions. The variable x denotes the linear displacement of the drive sheave and is related to θ by

$$\frac{2\pi r}{2\pi} \theta = x \text{ or } x = r\theta$$

The impedances in Figure 2-73 are listed as

$$Z_{J_{ds}} = \frac{1}{J_{ds}D^2}; Z_K = \frac{1}{K}; Z_{m_{cwt}} = \frac{1}{m_{cwt}D^2}; Z_{m_{car}} = \frac{1}{m_{car}D^2}$$

Step 2. Identify all independent nodes (FV and PV) in the impedance diagram and label all signals.

The impedance diagram has six nodes. Four of these nodes are FV nodes, and two are PV nodes. The node equations are given as

$$\text{FV node at } \theta: T_{J_{ds}} = T_1 - T_2$$

$$\text{FV node at } x: F = F_1 + F_2$$

$$\text{FV node at } x_{\text{car}}: F_1 = F_{\text{car}} + m_{\text{car}} \cdot g$$

$$\text{FV node at } x_{\text{cwt}}: F_2 = F_{\text{cwt}} - m_{\text{cwt}} \cdot g$$

$$\text{PV node across } Z_K \text{ at the car: } x - x_{\text{car}} = F_1 \cdot Z_K$$

$$\text{PV node across } Z_K \text{ at the cwt: } x - x_{\text{cwt}} = F_2 \cdot Z_K$$

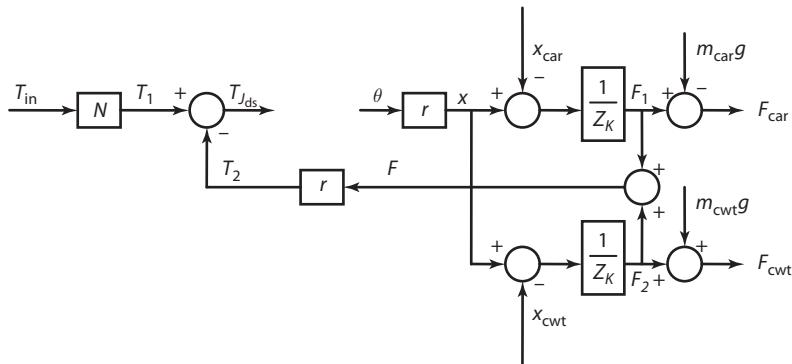
Several auxiliary equations pertaining to the gear ratios are also necessary and listed as

$$\text{Gear ratio: } T_1 = N T_{\text{in}}$$

$$\text{Drive sheave ratio: } F = \frac{T_2}{r}$$

Step 3. Represent select nodes as a summing junction and select the output of the summing junction such that (when it is connected to its associated impedance blocks) either gain or integral causality results.

Construction of the block diagram begins by implementing the FV and PV equations as summing junctions. We also include the auxiliary equations. The initial block diagram is presented in Figure 2-74.

FIGURE 2-74 GEARED ELEVATOR SYSTEM SUMMING-JUNCTION BLOCK DIAGRAM

Step 4. Add the impedance blocks; connect and create all necessary intermediate and output signals to complete the block diagram.

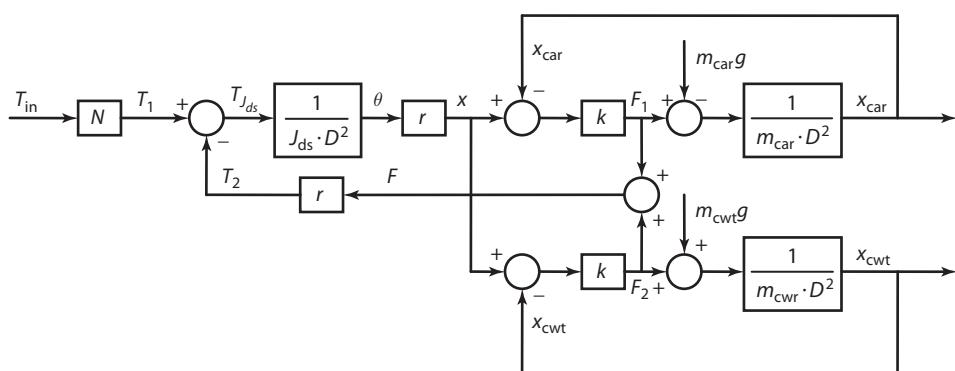
Substitution of the three mass impedances give

$$Z_{J_{ds}} = \frac{1}{J_{ds}D^2}$$

$$Z_{m_{cwt}} = \frac{1}{m_{cwt}D^2}$$

$$Z_{m_{car}} = \frac{1}{m_{car}D^2}$$

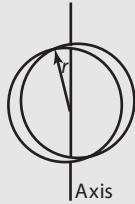
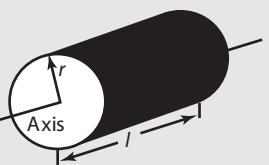
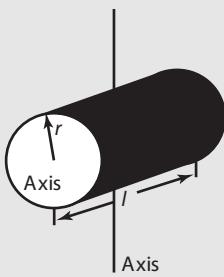
Replacing the spring impedances, $Z_K = \frac{1}{K}$ allows us to complete the block diagram. The completed block diagram is presented in Figure 2-75.

FIGURE 2-75 GEARED ELEVATOR SYSTEM BLOCK DIAGRAM

In Example 2-17, the reaction torque of the car and counterweight to the motor have been excluded. The effect is important, as it models the effect of load or reaction torque on the motor. The effect can be added easily once two fundamental electromechanical relationships—Lorentz's law and Faraday's law—are presented. These relationships are described in the next section.

In some mechanical rotational applications the inertia values may not be given and must be calculated. Table 2-9 presents inertia calculations for several common geometric shapes.

TABLE 2-9 INERTIAS FOR SEVERAL COMMON GEOMETRIC SHAPES

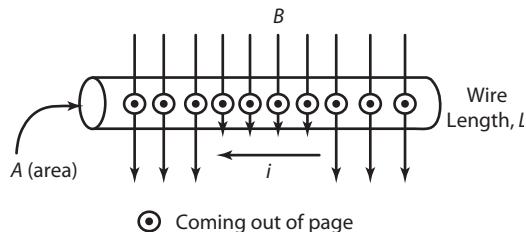
 <p>Horizontal Ring: $J = mr^2$</p>	 <p>Vertical Ring: $J = \frac{1}{2}mr^2$</p>
 <p>Horizontal Solid Cylinder: $J = \frac{1}{2}mr^2$</p>	 <p>Vertical Solid Cylinder: $J = \frac{1}{4}mr^2 + \frac{1}{12}ml^2$</p>
 <p>Solid Sphere: $J = \frac{2}{5}mr^2$</p>	 <p>Hollow Sphere: $J = \frac{2}{3}mr^2$</p>

2.8 Electrical–Mechanical Coupling

Motors, generators, and various sensors couple electrical systems with mechanical systems. The electromagnetic coupling is based on two laws: Lorentz's Law which describes electrical to mechanical coupling and Faraday's Law describing mechanical to electrical coupling.

2.8.1 Lorentz's Law—Electrical to Mechanical Coupling

Lorentz's force law, Figure 2-76, is used to relate current traveling through a wire in a magnetic field to force exerted on the wire.

FIGURE 2-76 LORENTZ'S FORCE LAW

$$F = ilB \text{ (Lorentz's Law)}$$

where

F = force, newtons

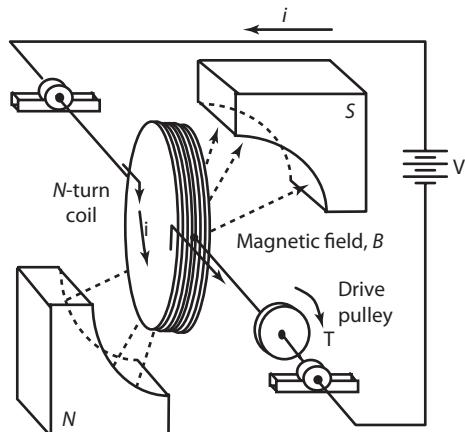
$$i = \text{current, amps} = \frac{\text{coulombs}}{\text{s}}$$

l = length of wire, meters

$$B = \text{magnetic field, Tesla} = \frac{\text{newtons}}{\text{Amp} \cdot \text{meter}}$$

The force exerted on the wire is coming up out of the page, obeying the right-hand rule: pointer finger in the direction of the current, middle finger in the direction of the magnetic field, and thumb in the direction of the force. The magnetic field is oriented at right angles to the current traveling through the wire. In situations where an angle other than 90° exists, the force is computed using the orthogonal component of the magnetic field, $F = ilB \sin \varphi$.

Lorentz's law relates current through a wire in a magnetic field to the mechanical translation force on the wire. A more useful form of the law, Figure 2-77, relates current in a coil to the

FIGURE 2-77 CURRENT-TORQUE RELATIONSHIP OF A COIL

mechanical torque exerted by the coil. Torque exerted by a current loop is the basic operating principle of many devices, including the electric motor and most electric meters.

$$T = NiAB \sin \varphi \text{ (Lorentz's law)}$$

where

T = torque, newton – meters

N = number of turns in the coil

i = current, amps

A = coil area, meters²

B = magnetic field, Tesla = $\frac{\text{newtons}}{\text{amp} \cdot \text{meter}}$

φ = angle between B and current

An external voltage supply, V , is used to create the current flowing through the coil. The torque exerted by the coil, T , (accessible through the drive pulley) is in the clockwise direction obeying the right-hand rule: pointer finger in the direction of the current, middle finger in the direction of the magnetic field, and thumb in the direction of the force. The magnet is curved to follow the radius swept by the rotating coil for a fraction of the complete revolution. During that fraction, the angle between the current direction and the magnetic field, φ , is 90°; however, as the coil rotates further, only the orthogonal component of the magnetic field, $B \sin \varphi$, is used, hence the reason for the $\sin \varphi$ term.

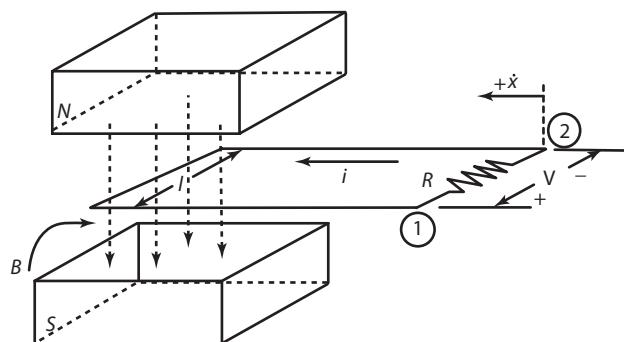
2.8.2 Faraday's Law—Mechanical to Electrical Coupling

Faraday's law of induction (Figure 2-78) relates the velocity of a wire loop as it is moved through a magnetic field to induce voltage (and current since the loop is closed) in the wire loop.

$$V = B l \dot{x} \text{ (Faraday's law)}$$

$$i = \frac{V}{R} \text{ (Ohm's law)}$$

FIGURE 2-78 FARADAY'S INDUCTION LAW



where

V = Induced voltage, volts

B = Magnetic field, Tesla = $\frac{\text{newton}}{\text{amp} \cdot \text{meter}}$

l = length of wire, meters

\dot{x} = horizontal velocity of loop, $\frac{\text{meters}}{\text{sec}}$

i = current, amps = $\frac{\text{coulombs}}{\text{s}}$

R = wire loop resistance, Ω

According to Faraday's law, current and voltage are induced in the closed loop only when motion exists ($\dot{x} \neq 0$), otherwise the induced current and voltage are zero. In situations where an angle, φ , other than 90° exists between the magnetic field and the current direction, the induced voltage and current are computed using the orthogonal component of the magnetic field, $B \sin \varphi$. Two directions of motion, $\pm \dot{x}$, are examined to determine the direction of the induced current and polarity of the induced voltage.

First, when $\dot{x} > 0$, the wire loop moves to the left into the magnetic field, B . Let us assume that the movement of the loop is due to an externally applied force, f_{in} , applied in the $+\dot{x}$ direction at point 2 in Figure 2-78. Faraday's law tells us there is a current induced in the loop, and Lenz's law tells us that this induced current will have a direction such that the net reaction force, f_{reaction} , it creates (by virtue of Lorentz's Law) opposes the applied force (the reaction forces on the sides of the loop are always equal and opposite in sign resulting in a zero net force contribution). Lenz's law requires that f_{reaction} be in the $+\dot{x}$ direction (rightward). Because the magnetic field, B , is directed downward, the induced current must travel in the counterclockwise direction (as shown), obeying the right-hand rule. The resulting induced voltage across the resistor becomes $V = B l \dot{x}$ (according to Faraday's Law) with the voltage drop going from point 1 to point 2 following the direction of the current.

Second, when $\dot{x} < 0$, the wire loop moves to the right out of the magnetic field, B . As before, it is assumed that the movement of the loop is due to an externally applied force, f_{in} , applied in the $+\dot{x}$ direction at the point 2 in Figure 2-78. Lenz's law requires that the reaction force, f_{reaction} , due to the induced current be in the $+\dot{x}$ direction (leftward). Because the magnetic field, B , is directed downward, the induced current must travel in the clockwise direction, obeying the right-hand rule. The resulting induced voltage across the resistor becomes $V = B l \dot{x}$ (according to Faraday's Law) with the voltage drop going from point 2 to point 1 following the direction of the current.

Faraday's law relates motion of a closed wire loop through a constant magnetic field to the electrical current in the wire. If a load impedance is inserted in the loop (such as a resistor), a voltage will also appear. A more useful form of the law relates rotational motion of a coil to electrical current flowing in the coil. This is the basic operating principal of the electric AC generator shown in Figure 2-79.

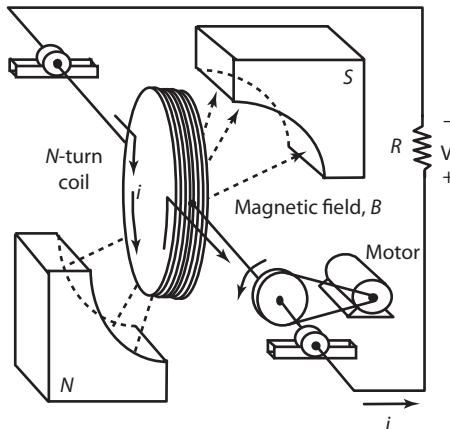
$$V = NAB \dot{\theta} \sin \theta + (\text{Faraday's Law})$$

$$i = \frac{V}{R} (\text{Ohm's Law})$$

where

V = induced voltage, volts

N = number of turns in coil

FIGURE 2-79 BASIC OPERATING PRINCIPLE OF THE ELECTRIC AC GENERATOR

$$A = \text{area of coil, meters}^2$$

$$B = \text{magnetic field, Tesla} = \frac{\text{newtons}}{\text{amp} \cdot \text{meter}}$$

$$\dot{\theta} = \text{angular velocity of coil, } \frac{\text{radians}}{\text{s}}$$

$$i = \text{current, amps} = \frac{\text{coulombs}}{\text{s}}$$

$$R = \text{load resistance, } \Omega$$

The motor in Figure 2-79 is providing a counterclockwise input torque, T_{in} , moving the coil in a counterclockwise direction. The reaction torque, $T_{reaction}$ (produced by the induced current in the coil) will occur in a clockwise direction, requiring the induced current to travel in a counterclockwise direction (as shown) when viewed from the left face of the coil. With the coil windings connected to a load resistance, the induced voltage will equal the voltage drop across the resistance with the indicated polarity. As the coil rotates, the angle between the magnetic field and the current direction differs from 90° . The orthogonal contribution of the magnetic field becomes $B \sin \dot{\theta} t$, as indicated. The result is a sinusoidally varying (AC) current and voltage.

2.8.3 Electrical–Mechanical Coupling Linear Relationships

Normally, motors and generators are constructed with enough poles and are wide enough magnets such that the sinusoidal component is *smoothed* to the point where it can be neglected. In these situations, Lorentz's and Faraday's laws can be linearly approximated. The linear relationships are summarized in Equations 2-6 and 2-7.

Lorentz's Electrical to Mechanical Linear Relationship

$$T = K_t i \text{ where } K_t \equiv NAB \frac{\text{newton} - \text{meters}}{\text{amp}} \quad (2-6)$$

In Equation 2-6, the magnetic field is usually given in teslas units, Tesla $\equiv \frac{\text{newton}}{\text{meter} \cdot \text{amp}}$. The units for area are square-meters.

Faraday's Mechanical to Electrical Linear Relationship

$$V = K_{\text{bemf}} \dot{\theta} \quad \text{where } K_{\text{bemf}} \equiv NAB \quad \frac{\text{volts}}{\frac{\text{rad}}{\text{sec}}} \quad (2-7)$$

(Note: “bemf” refers to “back electro motive force.”)

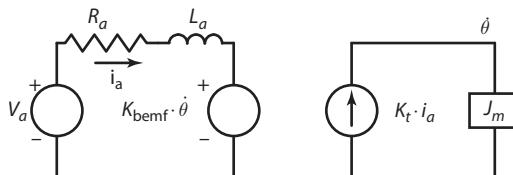
In Equation 2-7, the magnetic field is usually given in units of $\frac{\text{volt} \cdot \text{sec}}{\text{meter}^2}$ which are different units from those used in Equation 2-6. However, since voltage $\equiv \frac{\text{joules}}{\text{coulomb}} \equiv \frac{\text{newton} \cdot \text{meter}}{\text{coulomb}}$ and current $= \frac{\text{coulomb}}{\text{second}}$, it easily is seen that the two magnetic field units are consistent.

EXAMPLE 2-18 DC Motor

The DC motor is an actuator which converts electrical energy to mechanical energy. It is capable of producing high torque and accurate speed regulation. The motor is controlled by application of a DC voltage to its armature windings, which results in an armature current. The armature current creates an electromagnetic torque at the rotor according to Lorentz’s law. To prevent the rotor speed from going to infinity as the result of a constant torque input, an electrical damping term is present which produces a *back-emf* according to Faraday’s law. The effect of the back-emf is to reduce the voltage drop across the armature windings, thus reducing the current and the torque.

The electrical circuit diagram for the DC motor (including its inertia, J_m) is presented in Figure 2-80. The block diagram model is constructed following the same four-step procedure that has been used for all analogy-based modeling.

FIGURE 2-80 DC MOTOR MODEL CIRCUIT DIAGRAM



Solution

Step 1. Create/simplify the impedance diagram.

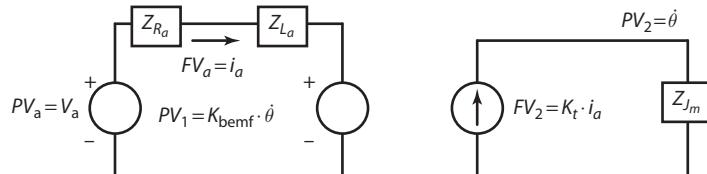
Construction of the impedance diagram for the DC motor is straightforward and presented in Figure 2-81.

Step 2. Identify all independent nodes (FV and PV) in the impedance diagram and label all signals.

The impedance diagram has two PV nodes whose equations are listed as

$$\text{PV node equation: } PV_a - PV_{R_a} - PV_{L_a} - PV_1 = 0$$

$$\text{PV node equation: } PV_2 = FV_2 \cdot Z_{J_m}$$

FIGURE 2-81 DC MOTOR MODEL IMPEDANCE DIAGRAM

The motor and generator auxiliary equations are also required.

$$\begin{aligned} PV_1 &= K_{\text{bemf}} \cdot PV_2 \\ FV_2 &= K_t \cdot FV_a \end{aligned}$$

Step 3. Represent select nodes as a summing junction and select the output of the summing junction such that (when it is connected to its associated impedance blocks) either gain or integral causality results.

Causality is an issue for the Z_{L_a} and Z_{J_m} blocks; the Z_{R_a} block is not affected. For integral causality, the Z_{L_a} block must have a PV input. To achieve this, the first PV node equation is represented as a summing junction whose output is PV_{L_a} . The Z_{J_m} block must have a FV input for integral causality. The second PV node equation is solved for FV_2 to accomplish this.

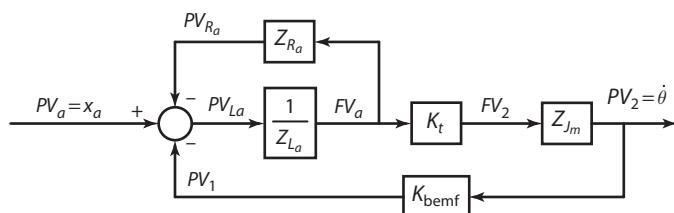
Step 4. Add the impedance blocks; connect and create all necessary intermediate and output signals to complete the block diagram.

The impedances used in the diagram are summarized as

$$\begin{aligned} Z_{L_a} &= L_a \cdot D \\ Z_{R_a} &= R_a \\ Z_{J_m} &= \frac{1}{J_m \cdot D} \end{aligned}$$

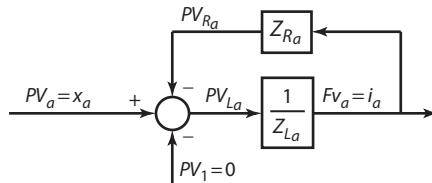
Adding these to the results, the complete block diagram is presented in Figure 2-82. From this DC motor block diagram, several characteristic can be investigated. The transfer function relating input voltage, x_a , to rotational velocity, $\dot{\theta}$, is derived from the block diagram by substituting the actual impedances for the impedance blocks. The resulting transfer function becomes

$$\dot{\theta} = \frac{K_t}{J_m L_a D^2 + J_m R_a D + K_t K_{\text{bemf}}} x_a$$

FIGURE 2-82 DC MOTOR BLOCK DIAGRAM

The armature loop of the motor consists of a resistance and an inductance, R_a and L_a , respectively. The time constant of the armature loop is computed from the subdiagram shown in Figure 2-83.

FIGURE 2-83 ARMATURE SUBDIAGRAM OF THE DC MOTOR



The PV_1 signal is set to zero for this calculation, and the transfer function from x_a to i_a becomes

$$i_a = \frac{1}{L_a D + R_a} x_a = \frac{\frac{1}{R_a}}{\frac{L_a}{R_a} D + 1} x_a$$

For small motors (under 1 or 2 hp), the time constant of the armature coil is usually in the vicinity of $\frac{L_a}{R_a} \approx .01$ seconds.

The back-emf constant, K_{bemf} , is approximated from the motor nameplate data as $\frac{\text{rated voltage, volts}}{\text{rated speed, rpm}}$ which may be converted to units of $\frac{\text{V-s}}{\text{rad}}$ by a $\frac{60}{2\pi}$ multiplication.

The motor torque constant, K_t , may be set to the back-emf constant unless the more accurate *blocked rotor* data is available. This information relates various armature currents to rotor torque and includes all dynamic losses and the effect of saturation.

2.9 Fluid Systems

A fluid is a substance which flows. It can be either a liquid or a gas. Gases (such as air) are often treated as compressible, since they expand to fit their container, while liquids (such as water and oil) are usually considered incompressible.

A force applied to a fluid produces a reaction force which is exerted by the fluid to the surface it is in contact with. A force may be applied to a fluid in either of two ways:

1. An externally applied pressure on an area.
2. The weight of the overhead fluid, called the head (height).

Pressure is related to head by the relationship

$$P = \rho g H$$

where

P = Pressure, Pascals

ρ = Mass density, kg/m³

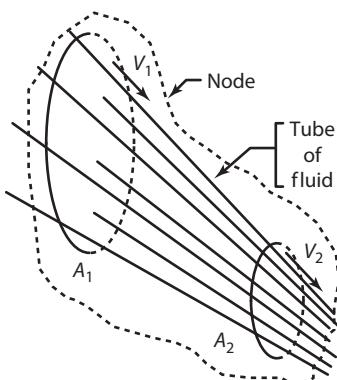
g = Gravity acceleration, $\frac{\text{m}}{\text{s}^2}$

H = Fluid head, m

Conservation of mass is analogous to Kirchoff's current law in electric circuits provided the fluid flow is steady, irrotational, and nonviscous. Conservation of mass is represented by the continuity equation, in which the total ingoing mass flow rate equals the total outgoing mass flow rate.

In Figure 2-84 mass flow rate is denoted as m , fluid velocity as v , fluid density as ρ , and the cross-sectional area of the tube of flow at location i as A_i .

FIGURE 2-84 PRINCIPLE OF CONSERVATION OF MASS—CONTINUITY EQUATION



Mass flow rates:

Into the node at A_1 : $m_1 = \rho_1 A_1 v_1$

Out of the node at A_2 : $m_2 = \rho_2 A_2 v_2$

Continuity requirement:

$$m_1 = m_2 \Rightarrow \rho_1 A_1 v_1 = \rho_2 A_2 v_2$$

Or, in general:

$$\rho A v = \text{constant}$$

For incompressible fluids, the density, ρ , is constant, and the continuity equation is written in terms of the fluid velocity as

$$q_1 = q_2$$

where

$$q \equiv Av = \text{volume flow rate, } \text{m}^3/\text{s}$$

For compressible fluids, the density, ρ , varies, and the continuity equation must be written in terms of the fluid mass flow rate as

$$m_1 = m_2$$

where:

$$m \equiv \rho A v = \text{mass flow rate, kg/s}$$

In practice, the weight flow rate, w , in units of N/s, often is used in place of the mass flow rate. We will use the weight flow rate for the remainder of this section.

Conservation of energy is a second important principle in fluid systems. Its application to steady, incompressible, and nonviscous fluid flow results in an energy equation called Bernoulli's equation. Bernoulli's equation states that the energy between two locations in a streamline differs by the net energy added (energy supplied minus energy lost).

With reference to Figure 2-84, Bernoulli's equation may be written for locations 1 and 2 as

$$\frac{P_1}{w} + \frac{v_1^2}{2g} + H_1 + E_{\text{net}} = \frac{P_2}{w} + \frac{v_2^2}{2g} + H_2$$

where

$$H_i \equiv \text{height of location } i, \text{ m}$$

$$E_{\text{net}} \equiv \text{energy added} - \text{energy lost}$$

Neglecting the E_{net} term, each side of the equation consists of two categories: a velocity dependent part ($v^2/2g$) and a static part ($P/w + H$). These categories lead to the definitions of dynamic and static pressures.

$$P_{\text{dynamic}} \equiv \frac{v^2}{2g} \cdot \rho g = \frac{\rho v^2}{2}$$

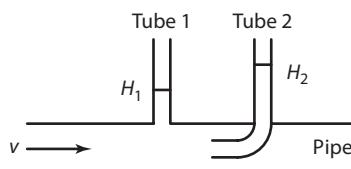
$$P_{\text{static}} \equiv \frac{P}{w} \cdot w + H \cdot \rho g = P + \rho g H$$

A third pressure frequently encountered is the stagnation pressure. The stagnation pressure is the sum of the static plus dynamic pressures.

EXAMPLE 2-19 Pitot Tube

Bernoulli's equation is used to determine the velocity of a fluid moving through a tube. A common application of the principle is the pitot tube (Figure 2-85), which is a device for measuring the speed of an incompressible fluid.

FIGURE 2-85 PITOT SYSTEM



With reference to Figure 2-85, the height of the fluid in tube 1 is the static head, and the height of the fluid in tube 2 is the static plus dynamic (stagnation) head. Assuming the net energy

contribution between locations 1 and 2 is zero and further assuming a level pipe, Bernoulli's equation simplifies to

$$\frac{P_1}{w} + \frac{v_1^2}{2g} + 0 = \frac{P_2}{w} + \frac{v_2^2}{2g} + 0$$

The velocities v_1 and v_2 are the fluid velocities at the entrances to the tubes. The velocity at the entrance to tube 1 is the fluid velocity, $v_1 = v$, and the velocity at the entrance to tube 2 is $v_2 = 0$ (called a stagnation point). Substituting these velocities into the Bernoulli equation produces an equation relating the fluid velocity to the head pressure.

$$v = \sqrt{2g\left(\frac{P_2}{w} - \frac{P_1}{w}\right)} = \sqrt{2g(H_2 - H_1)}$$

Fluid systems can be modeled as block diagrams using the analogy procedure discussed previously in this chapter. The following analogies in Table 2-10 for the PV and FV are used for fluid systems.

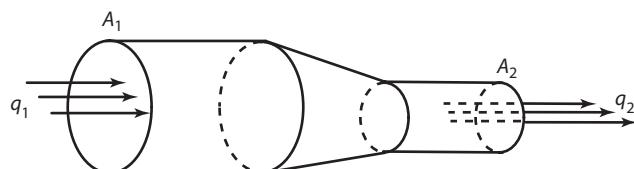
TABLE 2-10: IMPEDANCE ANALOGIES FOR FLUID SYSTEMS

Compressible Fluids	Incompressible Fluids
PV = Pressure or head	PV = Pressure
FV = Volume flow rate, q	FV = Weight flow rate, w

Most fluid systems consist of ducts, restrictions (or orifices), and tanks. Restrictions can be viewed as ducts with changes in their cross-sectional area and include orifices, valves, and nozzles. The impedances of restrictions and tanks are considered next.

A restriction in a fluid system is analogous to a resistance in an electrical system. A restriction can be viewed as a duct section with a change in its cross-sectional area, as in Figure 2-86.

FIGURE 2-86 FLUID RESTRICTION



For an incompressible fluid, continuity requires that $q_1 = q_2$ or $A_1 v_1 = A_2 v_2$. Assuming the restriction to be level, Bernoulli's equation becomes

$$P_1 + \frac{\rho v_1^2}{2} = P_2 + \frac{\rho v_2^2}{2}$$

Solving the continuity equation for v_1 and substituting the result into Bernoulli's equation, the velocity of the fluid at location 2 in Figure 2-86 becomes

$$v_2 = \sqrt{\frac{2(P_1 - P_2)}{\rho(1 - A_2/A_1)^2}}$$

or

$$q_2 = A_2 \cdot \sqrt{\frac{2(P_1 - P_2)}{\rho(1 - A_2/A_1)^2}}$$

The restriction equation is nonlinear and is often written in a more general form valid for incompressible fluids as

$$q = C_d A \sqrt{\frac{2(P_1 - P_2)}{\rho}} \quad (2-8)$$

or

$$q = C_d A \sqrt{2g\Delta H}$$

where

C_d ≡ discharge coefficient, ($0 < C_d \leq 1$)

A ≡ restriction area, A_2

A similar equation exists when dealing with compressible fluids; however, instead of using volume flow rate, the weight flow rate is used.

The general restriction equation for compressible fluids is

$$w = w_s KAY \sqrt{\frac{2p(P_1 - P_2)}{w_s}} \quad (2-9)$$

where

w ≡ weight flow rate

w_s ≡ specific weight of fluid

Y ≡ expansion factor; $\begin{cases} = 1; \text{incompressible} \\ < 1; \text{compressible} \end{cases}$

$$K \equiv \frac{C_d}{\sqrt{1 - (A_2/A_1)^2}}$$

With a basic understanding of the fluid resistance behavior, we can now establish the impedance relationship for the component using pressure as the *PV* and volume flow as the *FV*. Since the resistance obeys a nonlinear relationship, calculation of a linear impedance will require linearization. We'll begin with the general restriction equation for an incompressible fluid repeated here.

$$q = C_d A \sqrt{\frac{2(P_1 - P_2)}{\rho}}$$

The functional form for this equation is

$$q = q(A, P_1, P_2)$$

At a specified operation condition, (A_0, P_{10}, P_{20}) , the incompressible fluid resistance equation may be approximated by the following linearization.

$$\Delta q = \frac{\partial q}{\partial A} \Big|_{\substack{A_0 \\ P_{10} \\ P_{20}}} \cdot \Delta A + \frac{\partial q}{\partial P_1} \Big|_{\substack{A_0 \\ P_{10} \\ P_{20}}} \cdot \Delta P_1 + \frac{\partial q}{\partial P_2} \Big|_{\substack{A_0 \\ P_{10} \\ P_{20}}} \cdot \Delta P_2$$

The partials are evaluated as

$$\frac{\partial q}{\partial A} \Big|_{\substack{A_0 \\ P_{10} \\ P_{20}}} \equiv K_a$$

$$\frac{\partial q}{\partial P_1} \Big|_{\substack{A_0 \\ P_{10} \\ P_{20}}} \equiv K_p$$

$$\frac{\partial q}{\partial P_2} \Big|_{\substack{A_0 \\ P_{10} \\ P_{20}}} \equiv -K_p$$

The impedance of the restriction equation is the ratio of the *PV* to the *FV*, which is the inverse of the third partial:

$$Z_R \equiv \frac{\partial P_2}{\partial q} = \frac{\rho q_o}{(C_d A)^2}$$

2.9.1 Tanks

A tank in a fluid system is analogous to a capacitance in an electrical system. The tank impedance takes either of two forms, depending on the compressibility of the fluid. In order to simplify the analysis, we will use the volume flow rate as the flow variable and approximate compressibility effects through use of the bulk modulus of elasticity. It also will be useful to view total volume flow rate as consisting of two terms:

$$q \equiv q_{\text{com}} + q_{\text{inc}}$$

where

q_{com} ≡ compressible component

q_{inc} ≡ incompressible component

For an incompressible fluid, the total flow is $q = q_{\text{inc}}$. Feeding this into a tank with cross-sectional area as A , the rate of change in the tank head, \dot{H} , is determined by the equation

$$\dot{H} = \frac{1}{A} (q_{\text{in}} - q_{\text{out}})$$

or using operator notation as

$$H = \frac{1}{AD} \Delta q$$

The impedance of the tank, using pressure as the PV and volume flow for as the FV , becomes

$$Z_T \equiv \frac{1}{AD}$$

For a compressible fluid, the volume flow rate is, $q = q_{\text{com}}$. The compressibility effect is represented using bulk modulus of elasticity of the fluid, β . The bulk modulus, or fluid stiffness, is defined as

$$\beta(\text{Pa}) \equiv \frac{\Delta P(\text{Pa})}{\Delta V(\text{m}^3)/V(\text{m}^3)}$$

Solving for the change in volume yields

$$\Delta V = \frac{V}{\beta} \cdot \Delta P$$

Taking the derivative of both sides (recognizing V and β as constants) and substituting $q = \Delta \dot{V} = \dot{V}$ produces the volume flow rate relationship:

$$q = \frac{V}{\beta} \cdot \Delta \dot{P}$$

or using operator notation, we have

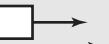
$$q = \frac{VD}{\beta} \cdot \Delta P$$

The impedance of the tank, using pressure as the PV and volume flow for as the FV , becomes

$$Z_T = \frac{\beta}{VD}$$

Table 2-11 summarizes the fluid restriction and tank impedances for both compressible and incompressible fluids.

TABLE 2-11: FLUID SYSTEM ANALOGIES

Analogy		Component	
Compressible: $PV =$ Pressure, P	$FV =$ Volume or Weight Flow Rate, q, w	<p>Restriction:</p>  $\Rightarrow Z_R = \frac{\rho q_0}{(C_d A)^2}$	<p>Tank:</p>  $\Rightarrow Z_T = \frac{\beta}{VD}$
Incompressible: $PV =$ Pressure, P	$FV =$ Volume Flow Rate, q	<p>Restriction:</p>  $\Rightarrow Z_R = \frac{\rho q_0}{(C_d A)^2}$	<p>Tank:</p>  $\Rightarrow Z_T = \frac{1}{VD}$

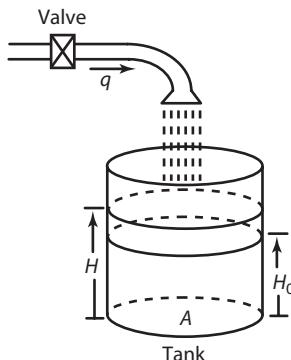
In the remainder of this section, several examples are presented which illustrate how the modified analogy approach is used to construct block diagram models for various fluid systems. For brevity, we will use problem specific variables instead of the generalized *PV* and *FV* notation in each of these examples.

EXAMPLE 2-20 Water Tank Block Diagram Model

A tank is filled with water from a faucet whose flow is controlled by an on-off valve. The fluid volume flow rate, q , in units of volume/time is the flow variable, and the height of the fluid in the tank, H , is the potential variable.

An illustration of the tank system is shown in Figure 2-87. The objective of this example is to develop the block diagram model for the tank system using the analogy approach.

FIGURE 2-87 FLUID TANK SYSTEM ILLUSTRATION



The tank is cylindrical with a cross-sectional area A . The height of water in the tank is represented by

$$H = H_0 + \frac{1}{A} \int q = H_0 + \frac{1}{DA} q$$

The term H_0 is the initial height of water in the tank. The impedance of the tank is the ratio of the *PV* across it to the *FV* through it and represented by

$$Z_{\text{tank}} = \frac{\Delta PV}{FV} = \frac{H - H_0}{q} = \frac{1}{DA}$$

Solution

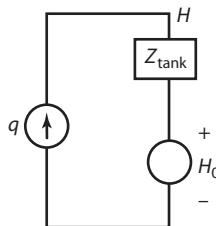
Step 1. Create/simplify the impedance diagram.

In going from the illustration to the impedance diagram, the input is selected as the volume flow rate. This is represented in the impedance diagram as a *FV* source. All flow from the source goes into the tank with no splitting or leakage. The tank fluid height then accumulates beginning at the initial height. The impedance diagram is presented in Figure 2-88.

Step 2. Identify all independent nodes (*FV* and *PV*) in the impedance diagram and label all signals.

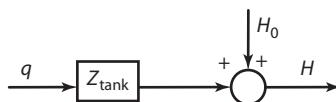
The impedance diagram has one *PV* node establishing the tank height as a function of the initial tank height and the flow into the tank.

$$H = H_0 + \frac{1}{DA} q$$

FIGURE 2-88 FLUID TANK SYSTEM IMPEDANCE DIAGRAM

Step 3. Represent select nodes as a summing junction and select the output of the summing junction such that (when it is connected to its associated impedance blocks) either gain or integral causality results.

The PV summing junction is presented in Figure 2-89.

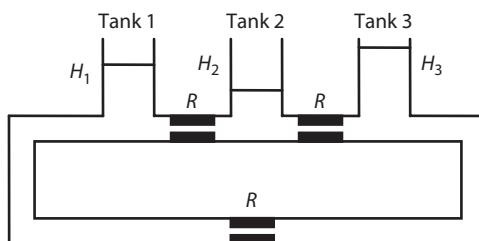
FIGURE 2-89 COMPLETED BLOCK DIAGRAM REPRESENTATION OF THE TANK SYSTEM

Step 4. Add the impedance blocks; connect and create all necessary intermediate and output signals to complete the block diagram.

The tank impedance has been included in the PV node equation—no other impedances are present—and the final block diagram is presented in Figure 2-89.

EXAMPLE 2-21 Three Tank Liquid System

This example is representative of the behavior of a tanking system filled with an incompressible fluid with no active source of pressure (such as a pump). All pressures are due to fluid head and atmosphere. The system consists of three cylindrical tanks all connected in series by pipes. Systems of this type may be applied to modeling the “slosh” of fluid in large baffled tanks, such as those found in ship and aircraft tankers. The three-tank system illustration is presented in Figure 2-90.

FIGURE 2-90 THREE-TANK SYSTEM ILLUSTRATION

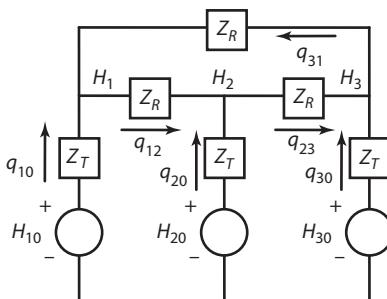
Each tank is cylindrical with a cross-sectional area A ; and the initial fluid height in each tank is different and given as H_{10} , H_{20} , H_{30} . The pipes connecting the tanks have identical resistances, R . The objective of this example is to develop the block diagram model for the three-tank system using the analogy approach.

Solution

Step 1. Create/simplify the impedance diagram.

Using the volume flow rate as the flow variable and head as the potential variable, the impedance diagram can be written as shown in Figure 2-91.

FIGURE 2-91 THREE-TANK SYSTEM IMPEDANCE DIAGRAM



Step 2. Identify all independent nodes (FV and PV) in the impedance diagram and label all signals.

The impedance diagram has three FV nodes and three PV nodes. These equations are summarized as

$$\text{FV node at } H_1: q_{10} + q_{31} - q_{12} = 0$$

$$\text{FV node at } H_2: q_{12} + q_{20} - q_{23} = 0$$

$$\text{FV node at } H_3: q_{23} + q_{30} - q_{31} = 0$$

The impedances are summarized as

$$Z_R \equiv R = \frac{\rho q_0}{(C_d A)^2} \quad \text{and} \quad Z_T \equiv \frac{1}{VD}$$

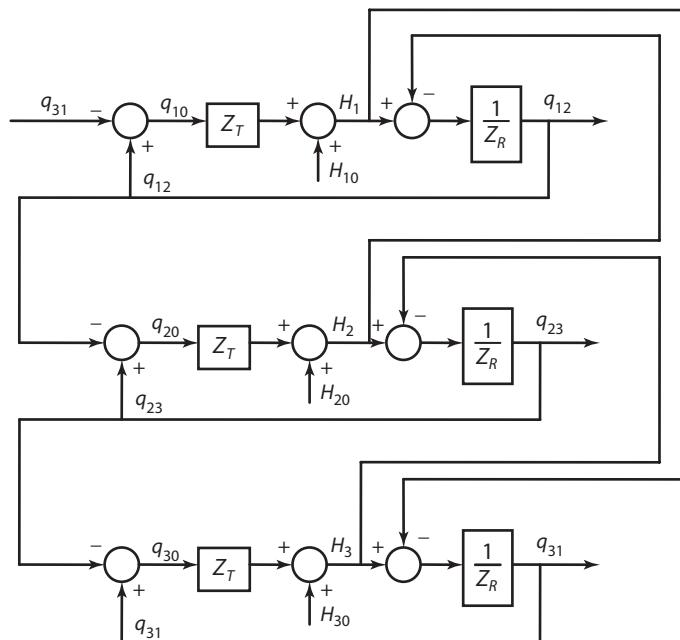
Step 3. Represent select nodes as a summing junction and select the output of the summing junction such that (when it is connected to its associated impedance blocks) either gain or integral causality results.

Causality is only an issue with the tank impedances, because for these elements, the input must be a flow variable (volume flow rate) for integral causality. This means that the three summing junctions used for the three FV node equations must have q_{10} , q_{20} , and q_{30} as outputs.

Step 4. Add the impedance blocks; connect and create all necessary intermediate and output signals to complete the block diagram

Including the impedances from step 2 and the node equations from step 3, the complete block diagram is presented in Figure 2-92.

This block diagram may initially appear complex, but after some examination, it will be evident that it is a collection of copies based on one simple feedback system pattern connected in a *daisy chain* manner. This configuration, also present in the three-mass hoistway model, is commonly encountered in multi mass or multi volume systems. The feedback paths in the daisy chain interconnections are reaction signals, and the forward paths are the forcing signals.

FIGURE 2-92 THREE-TANK SYSTEM BLOCK DIAGRAM**EXAMPLE 2-22 Hydraulic Pressure Regulator**

Hydraulic systems are powerful and extremely fast. They often use oil as the working fluid; however, due to the response speed, compressibility of the oil becomes an issue and must be included in the model.

In this example, we consider a pressure regulating valve whose function is to maintain constant pressure at the load despite fluctuations in the oil flow to the load. The regulator could be applied to many liquids, including oil and water. Regulators of this type are often found in domestic-oil heating systems and used to regulate the water pressure developed during heating. An illustration of the pressure regulator is shown in Figure 2-93.

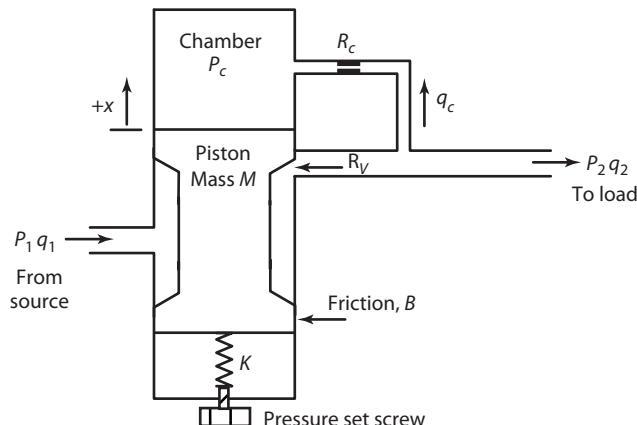
The pressure regulator operates as follows. For a disturbance increase in the load, P_2 , the chamber pressure, P_c , increases and pushes the piston down, thus reducing the valve opening, flow, and load pressure. For a disturbance decrease in the load, the opposite happens, resulting in an increase in the valve opening, flow from the source, and load pressure.

The chamber volume is equivalent to a tank and is assumed negligible compared to the piping volume, also a tank, which is connected to the load. Therefore, fluid compressibility only will be included in the load piping tank. The chamber tank will model the incompressible portion.

Solution**Step 1. Create/simplify the impedance diagram.**

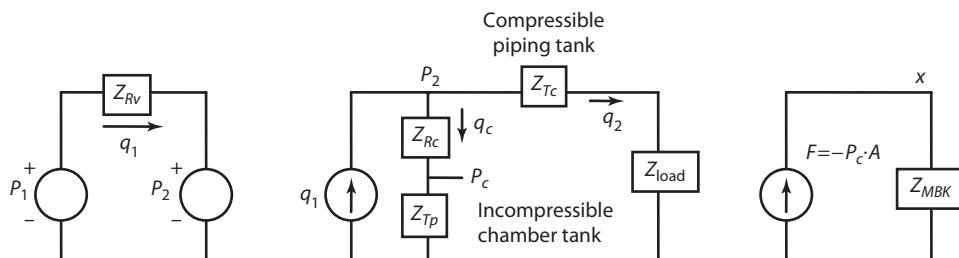
The impedance diagram consists of three subdiagrams,

- A valve flow-rate subdiagram
- A tank subdiagram
- A force balance subdiagram

FIGURE 2-93 PRESSURE REGULATOR SYSTEM ILLUSTRATION

The valve flow-rate subdiagram was created from the system illustration by observing that the flow rate through the valve is proportional to the potential (pressure) difference across the valve. Using a general impedance for the actual valve characteristics (we will substitute either the second order approximation or the nonlinear characteristic later), the flow-rate subcircuit is written according to Ohms law.

The impedance diagram for the pressure regulator system is shown in Figure 2-94.

FIGURE 2-94 PRESSURE REGULATOR SYSTEM IMPEDANCE DIAGRAM

The tank subdiagram was written from the flow splitter downstream of the valve. At this point, some of the flow through the valve goes to the chamber, and some goes (splits) to the load. The value of the flow in each branch of the split is determined by the branch impedance. The flow traveling to the chamber first encounters a resistance (restriction) followed by a tank (the chamber). Although the volume of the chamber tank varies according to the piston displacement, it is constant at any instant. The remainder of the flow from the splitter goes to the load—first encountering the piping followed by the load impedance. The piping is modeled as a compressible fluid tank, because the piping volume is known to be much greater than the chamber volume. The load impedance is unknown; however, any reasonable value could be used.

The force-balance subdiagram transfers fluid energy into mechanical energy as a normal force on the piston in the chamber side. This force (pressure times area) is applied to the piston mass, damping, and the spring between the piston and the casing. Since the mass, damper, and spring all have a common

ground (the casing), their impedance circuit consists of three parallel branches each to ground. For the spring to be attached to ground, we have assumed that the initial position of the piston is defined as zero, so that the delta displacement of the spring equals the displacement. The resulting piston displacement from the force-balance subdiagram is used in the valve impedance in the first subcircuit to create a feedback effect from the load.

The impedance's of the two tanks and the piston are given as

$$Z_{Tc} = \frac{1}{AD}$$

$$Z_{Tp} = \frac{\beta}{VD}$$

$$Z_{MBK} = \frac{1}{MD^2 + BD + K}$$

For integral causality, the tank impedances must have flow inputs and the mechanical impedance must have a force input.

Step 2. Identify all independent nodes (FV and PV) in the impedance diagram and label all signals.

Leaving the two restriction impedances in general form, Z_{Rv} and Z_{Rc} , the system equation are derived from the impedance diagram as

$$\text{Flow through valve: } q_1 = (P_1 - P_2)/Z_{Rv}$$

$$\text{Flow to chamber tank: } q_c = q_1 - q_2 = (P_2 - P_c)/Z_{Rc}$$

$$\text{Chamber tank pressure: } P_c = q_c Z_{Tc} = q_c \frac{1}{AD}$$

$$\text{Flow to load: } P_2 = q_2 Z_{Tp} = q_2 \frac{\beta}{VD}$$

$$\text{Piston force balance: } F = x Z_{MBK} = -P_c A = x(MD^2 + BD + K)$$

Since the area of the restriction to the chamber is fixed, the impedance of this resistance could be taken as a constant. On the other hand, the area of the valve varies depending on the piston location according to the relationship $q_1 = C_d A \sqrt{2(P_1 - P_2)/\rho}$. Applying the second-order approximation and assuming that P_1 remains constant, the linearized relationship becomes $\Delta q_1 = K_x \Delta x - K_p \Delta P_2$. Substituting this into the system equations and simplifying results in the following set of linearized equations, we have

$$\text{Flow through valve: } \Delta q_1 = K_x \Delta x - K_p \Delta P_2$$

$$\text{Flow to chamber tank: } \Delta q_c = \Delta q_1 - \Delta q_2 = (\Delta P_2 - \Delta P_c)/R_c$$

$$\text{Chamber tank pressure: } \Delta P_c = \Delta q_c \frac{1}{AD}$$

$$\text{Flow to load: } \Delta P_2 = \Delta q_2 \frac{\beta}{VD}$$

$$\text{Piston force balance: } \Delta F = -\Delta P_c A = \Delta x(MD^2 + BD + K)$$

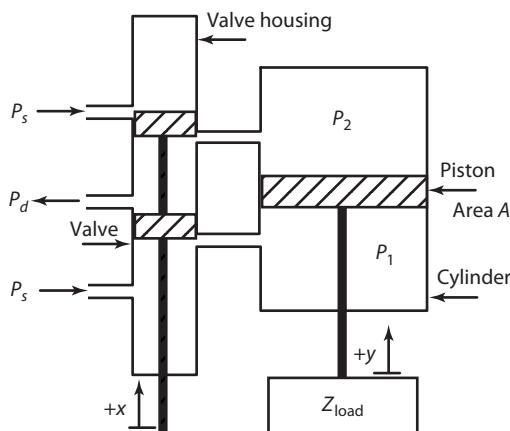
The remaining steps to complete the construction of the block diagram are left as an exercise.

EXAMPLE 2-23 Hydraulic Actuator

Hydraulic actuators are used in applications which require high actuating forces. Actuators of this type are found in commercial airliners for aerodynamic surface control, construction equipment, machine tools, large guns, and vehicular power steering. The main advantages of hydraulics are their large power-to-size ratio, rapid response, and high torque. Disadvantages include the need to install and maintain high-pressure hydraulic lines, line leakage as a fire hazard, and the adverse effect of temperature on the working fluid viscosity (resulting in drastic changes in the control gain).

An illustration of the hydraulic actuator is presented in Figure 2-95.

FIGURE 2-95 HYDRAULIC ACTUATOR SYSTEM ILLUSTRATION

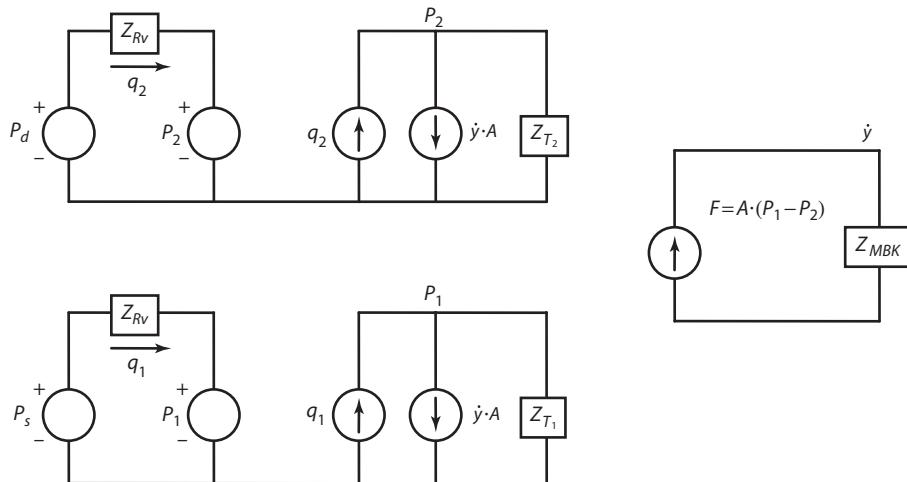


High-pressure oil is supplied through two lines to the valve housing with the remaining center line acting as a drain back to the oil source. Low-gain command signals are applied to the valve rod, resulting in motion in the $+/-x$ direction. Depending on the direction, the valve ports flow to either the top or bottom of the piston, resulting in high-gain motion of the load in the $+/-y$ direction. We will derive the impedance diagram for $a + x$ motion resulting in flow to the bottom chamber of the cylinder.

Solution

Step 1. Create/simplify the impedance diagram.

The impedance diagram (Figure 2-96), is constructed with the same structure as the valve, with chamber 2 on top, chamber 1 on the bottom, and the piston to the right. The impedance diagram for chamber 1 is constructed as follows. The pressure difference between the supply and chamber 1 creates a flow through the valve impedance, Z_{Rv} . The resulting flow goes into chamber 1. Chamber 1 is solid on all surfaces—except the piston can move up and down. Upward movement of the piston, y , is equivalent to a negative flow rate, $\dot{y}A$, (an outflow) provided the fluid is under compression. Denoting the chamber 1 impedance as $Z_{T_1} = \beta/V_1 D$, the net flow into the chamber is $q_1 - \dot{y}A$, which when passed through the chamber 1 produces the pressure P_1 .

FIGURE 2-96 HYDRAULIC ACTUATOR IMPEDANCE DIAGRAM

We will leave the formulation of the block diagram model as an exercise.

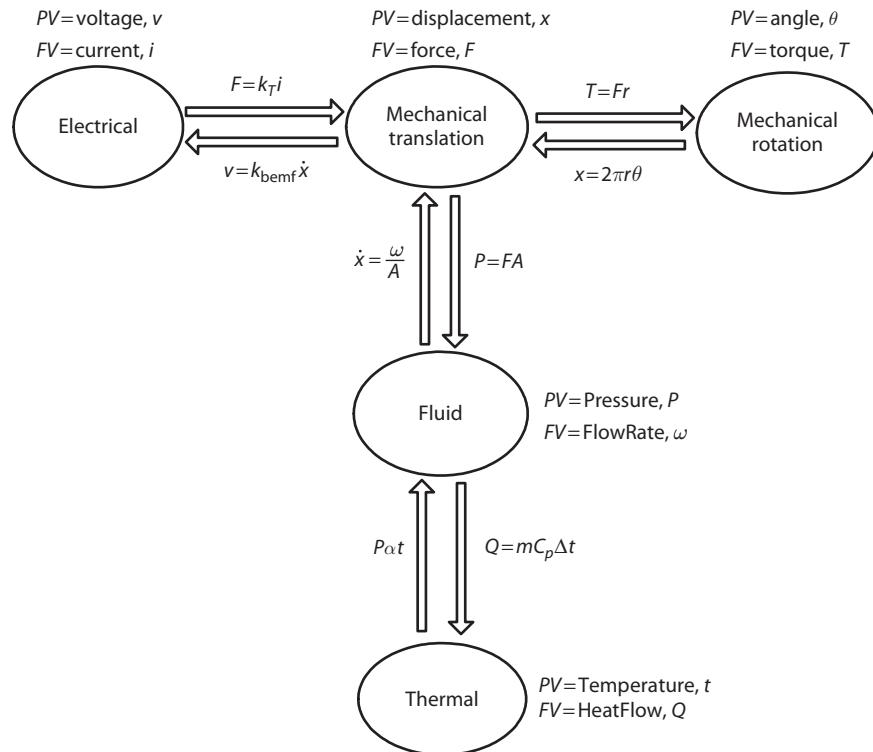
Appendix to Chapter 2 describes the systems with more than one input and/or output and are known as Multi-Input Multi-Output (MIMO) systems. An example of a MIMO system using State Space Method is also provided.

2.10 Summary

During the design stage of a mechatronics system, it is necessary to understand the performance characteristics of individual system components in various disciplines as well as the overall combined system performance. Component and system modeling play a critical role in the mechatronics development process, allowing functionality and complexity to be traded between disciplines to iteratively obtain an optimal system architecture.

This chapter has introduced two block-diagram based modeling approaches: the direct approach and the modified analogy approach. The direct approach is most suitable for single discipline modeling, while the modified analogy approach is more suitable for modeling multidisciplinary (mechatronic) applications. Figure 2-97 summarizes the basic PV and FV coupling equations that exists between five disciplines.

Figure 2-97 has shown coupling between select disciplines—in practice, other coupling paths may also exist. For example, if we were observing the thermal operation of a printed circuit board during various stress conditions, the electrical–thermal discipline may be directly coupled with one another.

FIGURE 2-97 BASIC MULTI-DISCIPLINARY COUPLING

REFERENCES

- Kuo, Benjamin C., *Automatic Control Systems, Third Edition*. Prentice-Hall Inc., New Jersey, 1975.
- D' Azzo, John J. and Constantine, Houpis H., *Linear Control System Analysis and Design Conventional and Modern, Third Edition*. McGraw-Hill Book Co., New York, 1988.
- Raven, Francis H., *Automatic Control Engineering, Third Edition*. McGraw-Hill Book Co., New York, 1978.
- Haliday, David and Resnick, Robert, *Fundamentals of Physics*. John Wiley & Sons, Inc. New York, 1970.
- Rizzoni, Giorgio, *Principles and Applications of Electrical Engineering, Third Edition*. McGraw-Hill Book Co., New York, 2000.
- Schwarz, Steven and Oldham, William. *Electrical Engineering—An Introduction*. Holt, Rinehart, and Winston, New York, 1984.
- U.S. Navy Bureau of Naval Personnel, *Basic Electronics*. Dover Publications, Inc. New York, 1973.
- Irwin, J. David, *Basic Engineering Circuit Analysis, Fourth Edition*. Prentice-Hall Inc., New Jersey, 1994.
- Lennart Ljung, *System Identification Theory for the User*. Prentice-Hall Inc., New Jersey, 1987.
- http://en.wikibooks.org/wiki/Control_Systems/MIMO_Systems
- Underwood, C. P., *HVAC Control Systems*. Taylor and Francis Group, 1998.

Romagnoli, Jose A. and Palazoglu, Ahmet,
Introduction to Process Control. Taylor and Francis Group, 2005.

[http://en.wikipedia.org/wiki/State space \(controls\)](http://en.wikipedia.org/wiki/State_space_(controls))

Bugeja, M., “Non-linear swing-up and stabilizing control of an inverted pendulum system,” 2003.

Proceeding of IEEE International Conference, EUROCON 2003. Computer as a Tool The IEEE Region 8, 22-24 Sept. 2003, Vol.2, pp. 437–441.

PROBLEMS

- 2.1. Write the following differential equations in D -operator form;
- $\dot{x}(t) + r(t) = 2x(t)$
 - $\dot{x}(t) + x(t) = 0$
 - $\dot{x}(t) + \int x(\tau)d\tau = x(t)$
 - $\ddot{x}(t) + 2\dot{x}(t) + x(t) = \dot{r}(t) + 3r(t)$
- 2.2. The following equations represent systems with input $r(t)$ and output $x(t)$. Compute the transfer function, $T(D) \equiv \frac{x(t)}{r(t)}$, for each system. Present your results in monic form using D operator notation.
- $3\dot{x}(t) + x(t) = 2r(t)$
 - $\ddot{x}(t) + \dot{x}(t) = 7\dot{r}(t)$
 - $2\dot{x}(t) + \int x(\tau)d\tau = r(t)$
 - $4\ddot{x}(t) + 7\dot{x}(t) - x(t) = 4\dot{r}(t) + r(t)$
- 2.3. Compute the loop transfer function, LTF , the closed-loop transfer function, $CLTF$, and the return difference, RD , for the following block diagrams.

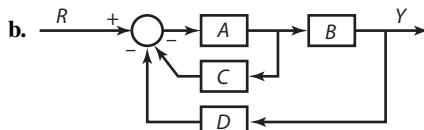
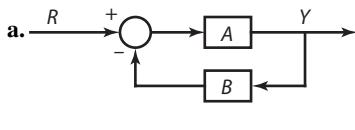


FIGURE P2-3

- 2.4. To illustrate how feedback is used to attenuate the effect of parameter disturbances on the controlled variable, compute the transfer functions, $\frac{Y}{R}$, for the following open- and closed-loop control systems. The control is the K block and the plant is G . The parameter variation is represented as the additive perturbation, ΔG .

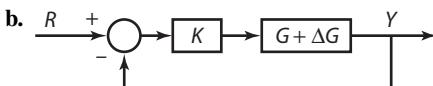
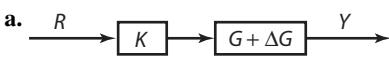
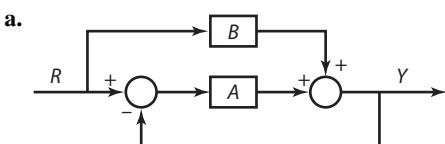


FIGURE P2-4

- 2.5. Use block diagram manipulations to compute the transfer functions for the following block diagrams.



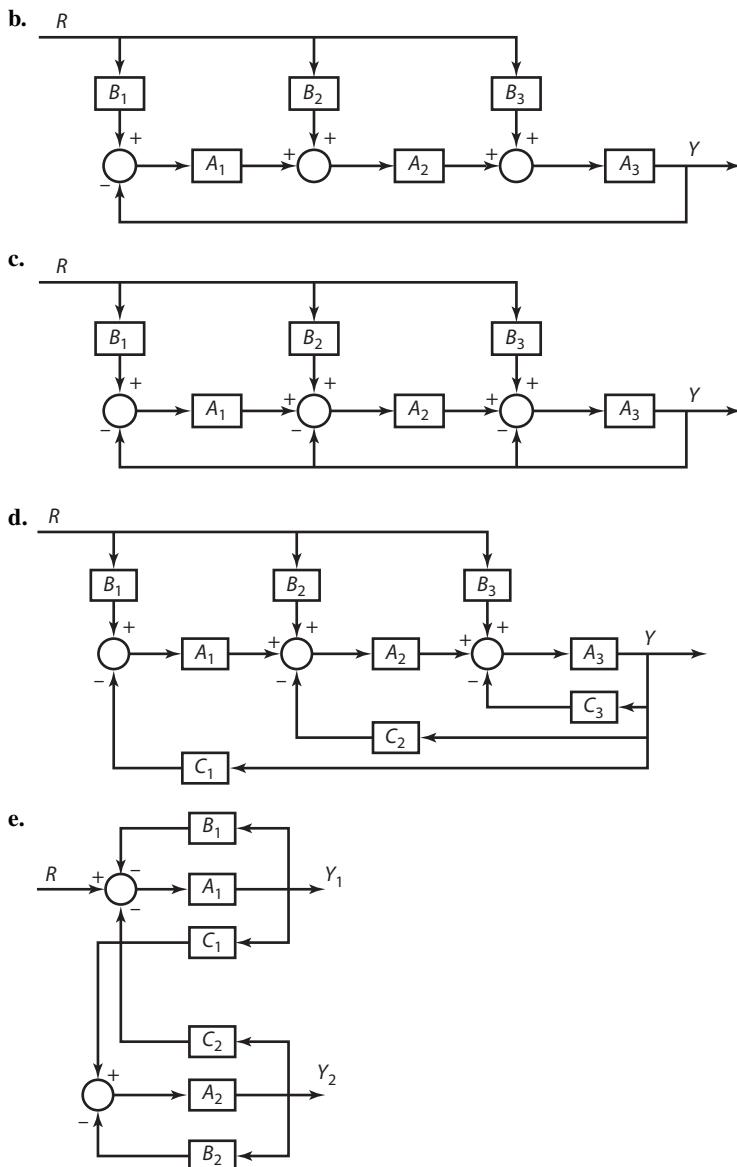


FIGURE P2-5

- 2.6. For the following mechanical system, construct the block diagram model and find the transfer function $\frac{x}{F}$.

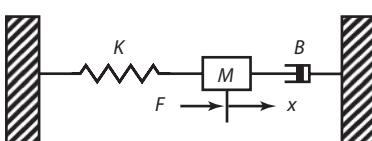


FIGURE P2-6

- 2.7. For the following mechanical system, construct the block diagram model and find the transfer function $\frac{x}{F}$.

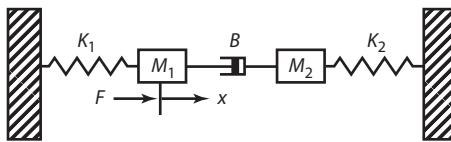


FIGURE P2-7

- 2.8. For the following mechanical lever system, construct the block diagram model and find the transfer function $\frac{x}{F}$.



FIGURE P2-8

- 2.9. The following mechanical system may be used to measure acceleration. Construct the block diagram model and find the transfer functions $\frac{x_1}{F}$, $\frac{x_2}{F}$, and $\frac{x_2}{x_1}$.

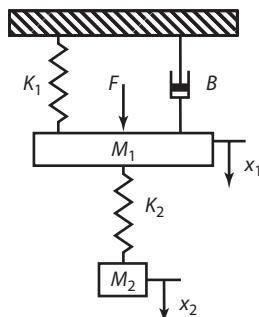


FIGURE 2-9

- 2.10. Compute the block diagram representation for the following electrical circuit.

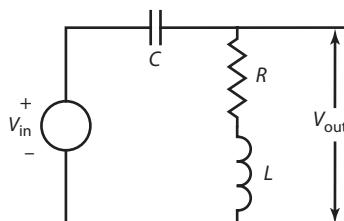
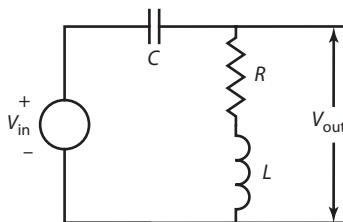
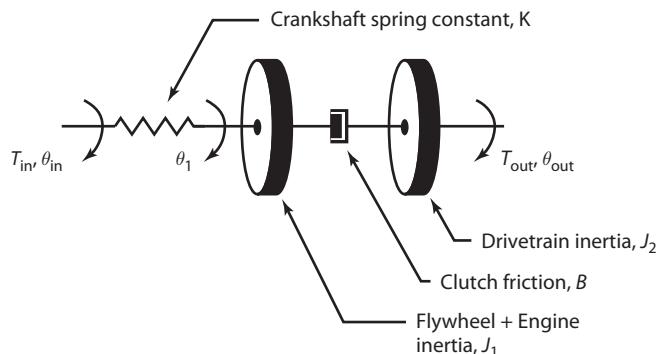


FIGURE P2-10

- 2.11. Compute the block diagram representation and the transfer functions for $\frac{V_C}{V_{in}}$, $\frac{V_R}{V_{in}}$, $\frac{V_L}{V_{in}}$, and $\frac{V_{out}}{V_{in}}$ for the following electrical circuit.

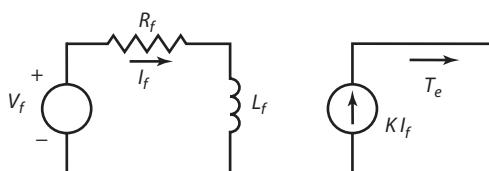
**FIGURE P2-11**

- 2.12. The “dry” plate clutch is often used in automobile drivetrain applications to transmit power from the engine to the driving wheels. An illustration of the clutch is shown in Figure P2-12.

**FIGURE P2-12**

The input to the clutch is torque, T_{in} , and the output is speed, $\dot{\theta}_{out}$. The impedances are based on torque as the flow variable and angle as potential variable. Speed is found by differentiating the potential variable.

- Construct the impedance diagram and label all signals.
 - Compute the transfer function $\dot{\theta}_{out}/\dot{\theta}_{in}$
- 2.13. The armature-controlled DC motor discussed in this chapter has an inherent “back emf” feedback loop present. Another control configuration is called field control. In this configuration, the “back emf” feedback is absent. The circuit diagram for a field controlled DC motor is shown in Figure P2-13.

**FIGURE P2-13**

The input to the motor is the field voltage, V_f , and the output is the electromagnetic torque, T_e . The impedance's are based on current (FV) and voltage (PV) on the electrical side and torque (FV) and angle (PV) on the mechanical side. K is the current-torque constant.

- Construct the impedance diagram for the field-controlled DC motor and label all signals.
 - Attach a load consisting of an inertia, J , plus friction, B , to the mechanical side and compute the transfer function $\dot{\theta}_{\text{out}}/V_f$ where $\dot{\theta}_{\text{out}}$ is the speed of the inertia, J .
- 2.14. An illustration of a simple propeller system in water is shown in Figure P2-14.

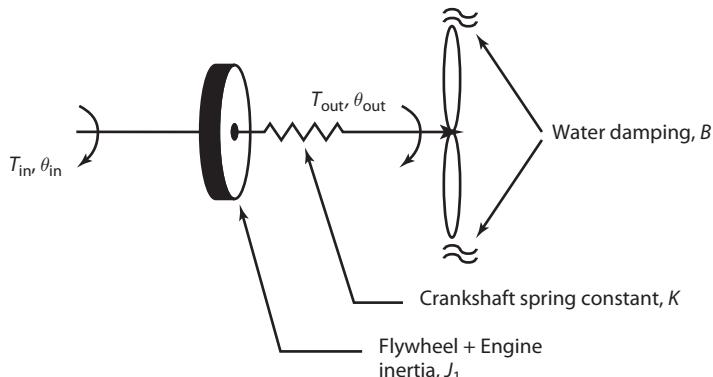


FIGURE P2-14

The input to the propeller system is torque, T_{in} , and the output is the prop speed, $\dot{\theta}_{\text{out}}$. The impedances are based on torque, as the flow variable and angle as potential variable.

- Construct the impedance diagram and label all signals.
 - Compute the transfer function $\dot{\theta}_{\text{out}}/T_{\text{in}}$
- 2.15. A transformer circuit which accounts for magnetization and core losses is presented in Figure P2-15.

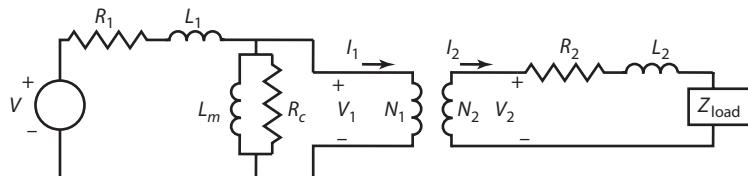


FIGURE P2-15

Voltage, V_{in} , is applied to the transformer primary side coil, which consists of a series resistance and inductance, R_1 and L_1 . The secondary side coil of the transformer is also modeled as a series resistance and inductance, R_2 and L_2 . The magnetization and core losses in the core of the transformer are modeled with L_c and R_m . The impedance diagram for the transformer is created by replacing each element of the circuit with its associated impedance. The impedance of each resistance, R_i , is denoted as Z_{R_i} , and the impedance of each inductance, L_i , as Z_{L_i} .

- Draw the impedance diagram for the transformer system.
- Draw the block diagram from the impedance diagram.
- From the block diagram compute the system equation relating the input, V_{in} , to the output, I_2 .

- 2.16. A DC motor is used to power a geared elevator system. The modified impedance diagram is presented in Figure P2-16.

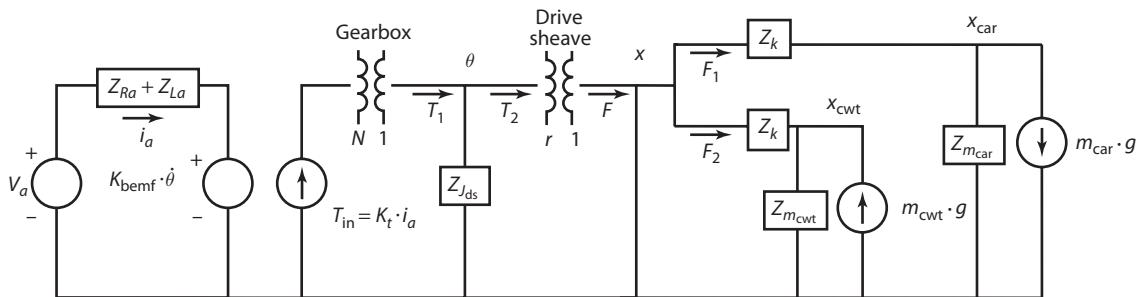


FIGURE P2-16

The impedances in the diagram are defined, using angle as the potential variable, as follows;

$$Z_{J_{ds}} = \frac{1}{J_{ds}D^2}; \quad Z_k = \frac{1}{k}; \quad Z_{m_{cwt}} = \frac{1}{m_{cwt}D^2}; \quad Z_{m_{car}} = \frac{1}{m_{car}D^2}$$

- Compute the block diagram system.
- From the block diagram, compute the following relationships:
 - Motor armature current to back emf.
 - Electromagnetic torque to armature current.
 - Gearbox torque transfer.
 - Force on the drive sheave.

Appendix to Chapter 2

Multi Input Multi Output Systems

Systems with more than one input and/or more than one output are known as Multi-Input Multi-Output systems, or they are frequently known by the abbreviation MIMO. The inputs and outputs of a MIMO system are generally interacting. An example of a MIMO system would be simultaneous control of both temperature and humidity in a close control air conditioning.

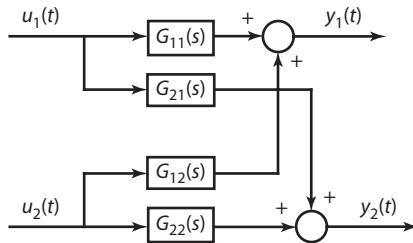
In a MIMO system we have a vector of inputs and a vector of outputs. The matrix that relates the Laplace transform of the output vector to that of the input vector is called the Transfer Function Matrix (TFM). Let us consider a MIMO system that has two inputs and two outputs as shown in the Figure 2-74.

Based on the Figure 2-74, the relationship between the inputs and the outputs are given by

$$Y_1(s) = G_{11}(s)U_1(s) + G_{12}(s)U_2(s)$$

and,

$$Y_2(s) = G_{21}(s)U_1(s) + G_{22}(s)U_2(s)$$

FIGURE 2-74 A SIMPLE MIMO SYSTEM BLOCK DIAGRAM

The above equations can be written in matrix form as

$$\begin{bmatrix} Y_1(s) \\ Y_2(s) \end{bmatrix} = \begin{bmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{bmatrix} \begin{bmatrix} U_1(s) \\ U_2(s) \end{bmatrix}$$

or

$$Y(s) = G(s)U(s)$$

where $G(s)$ is the TFM of the MIMO system under consideration.

MIMO systems that are lumped and linear can be described easily with state-space equations. This form is better suited for computer simulation than nth order input-output differential equations.

State Space Model

A state space representation is a mathematical model of a physical system as a set of input, output and state variables related by first-order differential equations. Let's say that we have two outputs, y_1 and y_2 , and two inputs, u_1 and u_2 . These are related in our system through the following system of differential equations:

$$\ddot{y}_1 + a_1\dot{y}_1 + a_0(y_1 + y_2) = u_1(t)$$

and;

$$\ddot{y}_2 + a_2(y_2 - y_1) = u_2(t)$$

Let us now assign our state variables and produce our first-order differential equations. As seen we have two second order differential equations and we would need two state variables for each of the differential equations (four in all) to take a first order form as explained further.

Let,

$$\begin{aligned} x_1 &= y_1 \\ x_2 &= \dot{x}_1 = \dot{y}_1 \\ x_3 &= y_2 \\ x_4 &= \dot{x}_3 = \dot{y}_2 \end{aligned}$$

now,

$$\dot{x}_2 = \ddot{y}_1 = -a_1\dot{y}_1 - a_0(y_1 + y_2) + u_1(t) = -a_1x_2 - a_0(x_1 + x_3) + u_1(t)$$

and;

$$\dot{x}_4 = \ddot{y}_2 = -a_2(y_2 - y_1) + u_2(t) = -a_2(x_3 - x_1) + u_1(t)$$

And finally we can assemble our state space equations as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -a_0 & -a_1 & -a_0 & 0 \\ 0 & 0 & 0 & 1 \\ a_2 & 0 & -a_2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

or

$$\dot{X} = AX + BU$$

and

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

or

$$Y = CX + DU$$

Thus, the general state space representation of a linear system with ‘ p ’ inputs ‘ q ’ outputs and ‘ n ’ state variables is

$$\dot{X} = AX + BU$$

and

$$Y = CX + DU$$

where,

X = State Vector of ‘ n ’ elements;

U = Input Vector of ‘ p ’ elements;

Y = Output Vector of ‘ q ’ elements;

A = State Matrix of the order ‘ $n \times n$ ’;

B = Input Matrix of the order ‘ $n \times p$ ’;

C = Output Matrix of the order ‘ $q \times n$ ’;

D = Feed forward Matrix of the order ‘ $q \times p$ ’;

Note: In this general formulation, all matrices are supposed to be time-invariant, i.e., none of their elements can depend on time. Also, for simplicity, D is often chosen to be the zero matrix, i.e., the system is chosen not to have direct feed through. Direct feed through is the case when a function output ‘ y ’ requires an input ‘ u ’ in order to execute i.e., ‘ u ’ has direct feed through to ‘ y ’.

EXAMPLE

Let us consider an example of a MIMO system and try to model it using State Space Method. The inverted pendulum is a classic problem in dynamics and control theory and widely used as a benchmark for testing control algorithms (PID controllers, neural networks, fuzzy control, genetic algorithms, etc). The non-linear inverted pendulum model considers the force on the cart as the input, and the angle of the pendulum and cart displacement as the outputs.

A Single-rod Inverted Pendulum (SIP) consists of a freely pivoted rod mounted on a cart as shown in the Figure 2-75. Figure 2-76 (a) and (b) shows the free-body diagrams of the system

FIGURE 2-75 SINGLE-ROD INVERTED PENDULUM SYSTEM

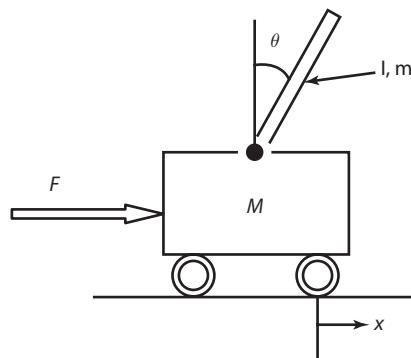
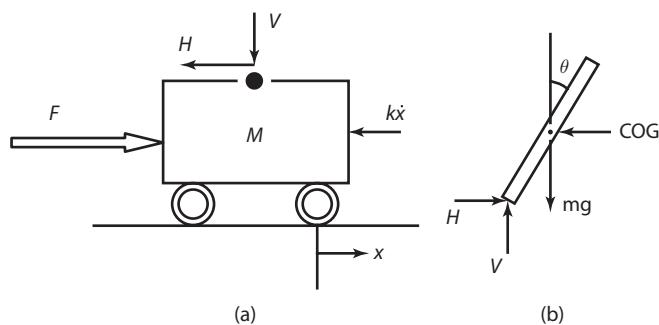


FIGURE 2-76 SINGLE-ROD INVERTED PENDULUM SYSTEM FREE-BODY DIAGRAM



Where,

m is the mass at the centre of gravity (COG) of the pendulum;

M is the mass of the cart;

L is the distance from the COG of the pendulum to the pivot;

x is the horizontal displacement of the cart;

g is the gravitational acceleration;

θ is the rod angular displacement;

k is the cart viscous friction coefficient;

c is the pendulum viscous friction coefficient;

I is the moment of inertia of the pendulum about the COG;

V and H are the vertical and horizontal reaction forces on the rod and F is the horizontal force on the cart.

The position vector of the pendulum COG with respect to the pivot is $(L \sin \theta i + L \cos \theta j)$, where i and j are the direction vector in x and y direction, respectively. However, the pivot point is also translating in x direction and hence, the resultant position vector of the pendulum COG is $(x + L \sin \theta)i + L \cos \theta j$. Applying Newton's second law at the center of gravity of the pendulum along the horizontal and vertical components yields

$$V - mg = m \frac{d^2}{dt^2} (L \cos \theta)$$

$$H = m \frac{d^2}{dt^2} (x + L \sin \theta)$$

Taking moments about the center of gravity yields the torque equation.

$$I\ddot{\theta} + c\dot{\theta} = VL \sin \theta - HL \cos \theta$$

Applying Newton's second law for the cart yields

$$F - H = M\ddot{x} + k\dot{x}$$

By combining above equations, the non-linear mathematical model of the cart and pendulum system is obtained and is given by

$$\begin{aligned}\ddot{\theta} &= \frac{1}{1 + L^2m} Lm(g \sin \theta - \ddot{x} \cos \theta) - c\dot{\theta} \\ \ddot{x} &= \frac{1}{M + m} [F - Lm(\ddot{\theta} \cos \theta - \dot{\theta}^2 \sin \theta) - k\dot{x}]\end{aligned}$$

However, as mentioned earlier only linear systems can be described with state-space equations, hence, we would need to linearize these equations. The inverted position of the pendulum corresponds to the unstable equilibrium point $(\theta, \dot{\theta}) = (0, 0)$. This corresponds to the origin of the state space. In the neighborhood of this equilibrium point, both θ and $\dot{\theta}$ are very small. In general, for small angles of θ and $\dot{\theta}$, $\sin \theta \approx \theta$, $\cos \theta \approx 1$ and $\dot{\theta}^2 \theta \approx 0$. Using these approximations, the mathematical model linearized around the unstable equilibrium point of the inverted pendulum is obtained, and given by

$$\begin{aligned}\ddot{\theta} &= \frac{1}{I + L_m^2} [Lm(g\theta - \ddot{x}) - c\dot{\theta}] \\ \ddot{x} &= \frac{1}{M + m} [F - Lm\ddot{\theta} - k\ddot{x}]\end{aligned}$$

To get these equations into valid state space matrix form both $\ddot{\theta}$ and \ddot{x} must be functions of lower order terms only. Hence, substituting \ddot{x} in $\ddot{\theta}$ and $\ddot{\theta}$ in \ddot{x} the above equations are further solved and presented as

$$\begin{aligned}\ddot{\theta} &= \frac{1}{I + L^2m} \left[Lm \left(g\theta - \frac{1}{M+m} \left[F - Lm\ddot{\theta} - k\dot{x} \right] \right) - c\dot{\theta} \right] \\ \ddot{\theta} &= \frac{Lmg}{I + L^2m} \theta - \frac{Lm}{(I + L^2m)(M+m)} F + \frac{(Lm)^2}{(I + L^2m)(M+m)} \ddot{\theta} + \frac{Lmk}{(I + L^2m)(M+m)} \dot{x} \\ &\quad - \frac{c}{I + L^2m} \dot{\theta} \\ \ddot{\theta} - \frac{(Lm)^2}{(I + L^2m)(M+m)} \ddot{\theta} &= \frac{Lmk}{(I + L^2m)(M+m)} \dot{x} - \frac{c}{I + L^2m} \dot{\theta} + \frac{Lmg}{I + L^2m} \theta \\ &\quad - \frac{Lm}{(I + L^2m)(M+m)} F \\ \left[\frac{(I + L^2m)(M+m) - (Lm)^2}{(I + L^2m)(M+m)} \right] \ddot{\theta} &= \frac{Lmk}{(I + L^2m)(M+m)} \dot{x} - \frac{(M+m)c}{(I + L^2m)(M+m)} \dot{\theta} \\ &\quad + \frac{(M+m)Lmg}{(I + L^2m)(M+m)} \theta - \frac{Lm}{(I + L^2m)(M+m)} F \\ \left[\frac{I(M+m) + L^2Mm + (Lm)^2 - (Lm)^2}{(I + L^2m)(M+m)} \right] \ddot{\theta} &= \frac{Lmk}{(I + L^2m)(M+m)} \dot{x} - \frac{(M+m)c}{(I + L^2m)(M+m)} \dot{\theta} \\ &\quad + \frac{(M+m)Lmg}{(I + L^2m)(M+m)} \theta - \frac{Lm}{(I + L^2m)(M+m)} F \\ [I(M+m) + L^2Mm] \ddot{\theta} &= Lmk\dot{x} - (M+m)c\dot{\theta} + (M+m)Lmg\theta - LmF \\ \ddot{\theta} &= \frac{Lmk}{I(M+m) + L^2Mm} \dot{x} - \frac{(M+m)c}{I(M+m) + L^2Mm} \dot{\theta} + \frac{(M+m)Lmg}{I(M+m) + L^2Mm} \theta \\ &\quad - \frac{Lm}{I(M+m) + L^2Mm} F\end{aligned}$$

Let,

$$\nu_1 = \frac{(M+m)}{I(M+m) + L^2Mm}$$

Hence,

$$\begin{aligned}\ddot{\theta} &= \frac{Lmk\nu_1}{(M+m)} \dot{x} - c\nu_1 \dot{\theta} + Lmg\nu_1 \theta - \frac{Lmv_1}{(M+m)} F \\ \ddot{\theta} &= \frac{Lmk\nu_1}{(M+m)} \dot{x} + Lmg\nu_1 \theta - c\nu_1 \dot{\theta} - \frac{Lmv_1}{(M+m)} F\end{aligned}$$

Similarly,

$$\begin{aligned}
 \ddot{x} &= \frac{1}{M+m} \left\{ F - Lm \left[\frac{1}{I+L^2m} [Lm(g\theta - \dot{x}) - c\dot{\theta}] \right] - k\dot{x} \right\} \\
 \ddot{x} &= \frac{1}{M+m} F - \frac{(Lm)^2 g}{(I+L^2m)(M+m)} \theta + \frac{(Lm)^2}{(I+L^2m)(M+m)} \dot{x} + \frac{Lmc}{(I+L^2m)(M+m)} \dot{\theta} \\
 &\quad - \frac{k}{M+m} \dot{x} \\
 \ddot{x} - \frac{(Lm)^2}{(I+L^2m)(M+m)} \ddot{x} &= \frac{1}{M+m} F - \frac{(Lm)^2 g}{(I+L^2m)(M+m)} \theta + \frac{Lmc}{(I+L^2m)(M+m)} \dot{\theta} \\
 &\quad - \frac{k}{M+m} \dot{x} \\
 \left[\frac{(I+L^2m)(M+m) - (Lm)^2}{(I+L^2m)(M+m)} \right] \ddot{x} &= \frac{(I+L^2m)}{(I+L^2m)(M+m)} F - \frac{(Lm)^2 g}{(I+L^2m)(M+m)} \theta \\
 &\quad + \frac{Lmc}{(I+L^2m)(M+m)} \dot{\theta} - \frac{(I+L^2m)k}{(I+L^2m)(M+m)} \dot{x} \\
 [I(M+m) + L^2Mm] \ddot{x} &= (I+L^2m)F - (Lm)^2 g\theta + Lmc\dot{\theta} - (I+L^2m)k\dot{x} \\
 \ddot{x} &= \frac{(I+L^2m)}{[I(M+m) + L^2Mm]} F - \frac{(Lm)^2 g}{[I(M+m) + L^2Mm]} \theta + \frac{Lmc}{[I(M+m) + L^2Mm]} \dot{\theta} \\
 &\quad - \frac{(I+L^2m)k}{[I(M+m) + L^2Mm]} \dot{x}
 \end{aligned}$$

Let

$$v_2 = \frac{(I+L^2m)}{I(M+m) + L^2Mm}$$

Hence,

$$\begin{aligned}
 \ddot{x} &= v_2 F - \frac{(Lm)^2 g v_2}{(I+L^2m)} \theta + \frac{Lmc v_2}{(I+L^2m)} \dot{\theta} - k v_2 \dot{x} \\
 \ddot{x} &= -\frac{(Lm)^2 g v_2}{(I+L^2m)} \theta + \frac{Lmc v_2}{(I+L^2m)} \dot{\theta} - k v_2 \dot{x} + v_2 F
 \end{aligned}$$

Now, our state variables are θ , $\dot{\theta}$, x and \dot{x} and hence, the two linear differential equations can be presented in state space from as

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{\theta} \\ \ddot{\theta} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -kv_2 & -\frac{(Lm)^2 gv_2}{(I + L^2 m)} & \frac{Lmcv_2}{(I + L^2 m)} \\ 0 & 0 & 0 & 1 \\ 0 & \frac{Lmkv_1}{(M + m)} & Lmgv_1 & -cv_1 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ \theta \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} 0 \\ v_2 \\ 0 \\ -\frac{Lmv_1}{(M + m)} \end{bmatrix} F$$

and

$$\begin{bmatrix} x \\ \theta \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ \theta \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} F$$

CHAPTER 3

SENSORS AND TRANSDUCERS

- 
- 3.1 An Introduction to Sensors and Transducers
 - 3.1.1 Sensor Classification
 - 3.1.2 Parameter Measurement in Sensors and Transducers
 - 3.1.3 Quality Parameters
 - 3.1.4 Errors and Uncertainties in Mechatronic Modeling Parameters
 - 3.2 Sensitivity Analysis—Influence of Component Variation
 - 3.3 Sensors for Motion and Position Measurement
 - 3.3.1 Resistance Transducers
 - 3.3.2 Inductive Transducers
 - 3.3.3 LVDT
 - 3.3.4 RVDT
 - 3.3.5 Capacitance Transducers
 - 3.4 Digital Sensors
 - 3.4.1 Digital Encoders
 - 3.4.2 Encoder Principle
 - 3.4.3 Incremental Encoders
 - 3.4.4 Absolute Encoders
 - 3.4.5 Linear Encoder
 - 3.4.6 Moire Fringe Transducer
 - 3.4.7 Applications
 - 3.5 Force, Torque, and Tactile Sensors
 - 3.5.1 Sensitivity of Resistive Transducers
 - 3.5.2 Strain Gauges
 - 3.5.3 Offset Voltage
 - 3.5.4 Tactile Sensors
 - 3.6 Vibration and Acceleration Sensors
 - 3.6.1 Piezoelectric Transducers
 - 3.6.2 Active Vibration Control
 - 3.6.3 Magnetostrictive Transducer
 - 3.7 Sensors for Flow Measurement
 - 3.7.1 Solid Flow
 - 3.7.2 Liquid Flow
 - 3.7.3 Sensors Based on Differential Pressure
 - 3.7.4 Ultrasonic Flow Transducers for Flow Measurement
 - 3.7.5 Drag Force Flow Meter
 - 3.7.6 Turbine Flow Meter
 - 3.7.7 Rotor Torque Mass Flow Meter
 - 3.7.8 Fluid Measurement using Laser Doppler Effect
 - 3.7.9 Hot Wire anemometers
 - 3.7.10 Electromagnetic Flow Meters
 - 3.8 Temperature Sensing Devices
 - 3.8.1 Thermistors
 - 3.8.2 Thermocouple
 - 3.8.3 Radiative Temperature Sensing
 - 3.8.4 Temperature Sensing using Fiber Optics
 - 3.8.5 Temperature Sensing using Interferometrics
 - 3.9 Sensor Applications
 - 3.9.1 Eddy Current Transducers
 - 3.9.2 Hall Effect
 - 3.9.3 Pneumatic Transducers
 - 3.9.4 Ultrasonic Sensors
 - 3.9.5 Range Sensors
 - 3.9.6 Laser Interferometric Transducer
 - 3.9.7 Fiber Optic Devices in Mechatronics
 - 3.10 Summary
 - References
 - Problems

Instrumentation plays a key role in the modern technological world. An essential component in mechatronic systems which is integrally linked to instrumentation is the sensor, whose function is to

Provide a mechanism for collecting different types of information about a particular process.

Sensors are used to inspect work, evaluate the conditions of work under progress, and facilitate the higher-level monitoring of the manufacturing operation by the main computing system. They can be

used during pre-process, in-process and post-process operations. In some situations, sensors are used to translate a physical phenomenon into an acceptable signal that can be analyzed for decision making. Intelligent systems use sensors to monitor particular situations influenced by a changing environment and to control them with corrective actions.

In virtually every application, sensors transform real-world data into electrical signals. A sensor is defined as

A device that produces an output signal for the purpose of sensing of a physical phenomenon.

Sensors are also referred to as **transducers**. They cover a broader range of activities, which provide them with the ability to identify environmental inputs that can extend beyond the human senses. A transducer is defined as

A device that converts a signal from one physical form to a corresponding signal, which has a different physical form.

In a transducer, the quantities at the input level and the output level are different. A typical input signal could be electrical, mechanical, thermal, and optical. Signal detection is normally handled by electrical transducers in manufacturing industries involving certain process automation. A transducer is an element or device used to convert information from one form to another. The change in information is measured easily.

A spring is a simple example of a transducer. When a certain force is applied to a spring, it stretches, and the force information is translated to displacement information, as shown in Figure 3-1. Different quantities of force produce differential movements, which are a measure of the force.

Displacement y is proportional to force F , which can be expressed as

$$F = k \cdot y$$

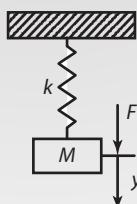
where k is constant

F = applied force

y = deflection

k = constant

FIGURE 3-1 PRIMARY TRANSDUCER



3.1 Introduction to Sensors and Transducers

The extent to which sensors and transducers are used is dependent upon the level of automation and the complexity of the control system. The modeling requirements of the complex control systems have introduced a need for fast, sensitive, and precise measuring devices. Due to these demands, sensors are being miniaturized and implemented in a microscale by combining several sensors and

data-processing mechanisms. Many microsystems have been built on the “lab-on-a-chip” concept. The entire unit can be contained in a silicon chip of the size less than 0.5×0.5 mm.

Selection of a sensor or a transducer depends on

- Variables measured and application.
- The nature of precision and the sensitivity required for the measurement.
- Dynamic range.
- Level of automation.
- Complexity of the control system and modeling requirements.
- Cost, size, usage, and ease of maintenance.

Two important components in modern control systems (whether electrical, optical, mechanical, or fluid) are the system’s sensors and transducers. The sensor elements detect measurands (variables to be measured) and convert them into acceptable form, generally as electrical signals. The maximum accuracy of the total system is controlled by the sensitivity of the individual sensors and the internally generated noise of the sensor itself. In a control system used for measurement and control, any parameter change either in measurands (variables to be measured) or in signal conditioning, has a direct effect on the sensitivity of the model.

Figure 3-2 shows elements of a sensor-based measurement system. The function of the sensor is to sense the information of interest and to convert this information into an acceptable form by a signal conditioner. The function of the signal conditioner is to accept the signal from the detector and to modify it in a way acceptable to the display unit. The function of the display-readout is to accept the signal from the signal conditioner and to present it in a displayable fashion. The output can be in the form of an output display, or a printer, or it may be passed on to a controller. It also can be manipulated and fed back to the source from which the original signal was measured.

FIGURE 3-2 A MECHATRONICS MEASUREMENT SYSTEM WITH AUXILIARY ENERGY SOURCE

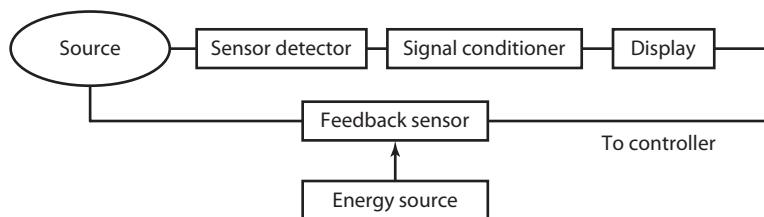
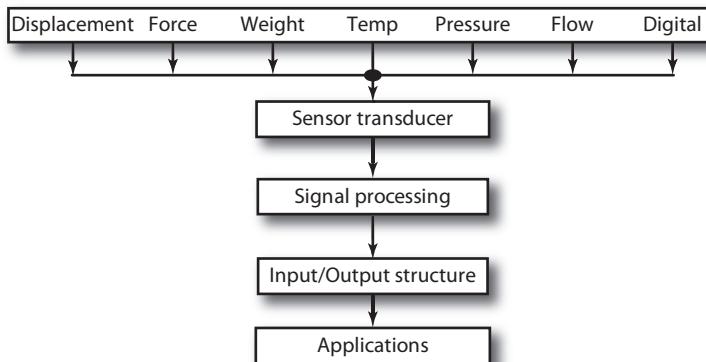


Figure 3-3 presents the components of an instrumentation system used for a general sensing application. A typical system consists of primary elements that sense and convert information into a more suitable form to be handled by the measurement system—signal conditioning stage for processing and modifying the information, an input/output stage for interface, and control with the external processes.

FIGURE 3-3 GENERAL INSTRUMENTATION SYSTEM AND ITS COMPONENTS

3.1.1 Sensor Classification

In the design of a mechatronics system, selection of a suitable sensor is very important. Table 3-1 summarizes some general sensor classifications.

Sensors are classified into two categories based on the output signal, power supply, operating mode and the variables being measured.

- **Analog sensors:** Analog is a term used to convey the meaning of a continuous, uninterrupted, and unbroken series of events. Analog sensors typically have an output, which is proportional to the variable being measured. The output changes in a continuous way, and this information is obtained on the basis of amplitude. The output is normally supplied to the computer using an analog-to-digital converter.
- **Digital sensors:** Digital refers to a sequence of discrete events. Each event is separate from the previous and next events. The sensors are digital if their logic-level outputs are of a digital nature. Digital sensors are known for their accuracy and precision, and do not require any converters when interfaced with a computer monitoring system.

TABLE 3-1 SENSOR CLASSIFICATION SCHEMES

Classification	Sensor Type
Signal Characteristics	Analog Digital
Power Supply	Active Passive
Mode of Operation	Null type Deflection type
Subject of Measurement	Acoustic Biological Chemical Electric Mechanical Optical Radiation Thermal Others

Another form of classification, active or passive, is based on the power supply.

- **Active sensors:** Active sensors require external power for their operation. The external signal is modified by the sensor to produce the output signal. Typical examples of devices requiring an auxiliary energy source are strain gauges and resistance thermometers.
- **Passive sensors:** In a passive sensor, the output is produced from the input parameters. The passive sensors (self generating) produce an electrical signal in response to an external stimulus. Examples of passive types of sensors include piezoelectric, thermoelectric, and radioactive.

Based on the operating and display mode of an instrumentation system, sensors are classified as deflection type or null type.

- **Deflection sensors:** Deflection sensors are used in a physical setup where the output is proportional to the measured quantity that is displayed.
- **Null sensors:** In null-type sensing, any deflection due to the measured quantity is balanced by the opposing calibrated force so that any imbalance is detected.

A final classification of sensors is based on the subject of measurement. Such subjects include acoustic, biological, chemical, electric, magnetic, mechanical, optical, radiation, thermal, and others.

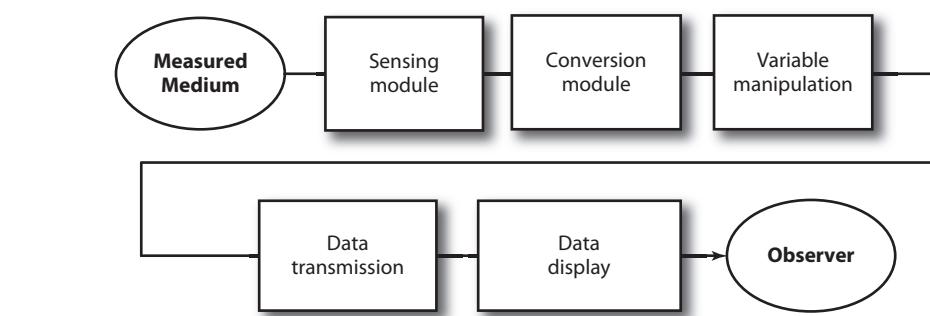
3.1.2 Parameter Measurement in Sensors and Transducers

Let us examine the instrumentation system model from the viewpoint of its functional elements in a generalized way. The elements contribute to the sensing and measurement of an instrumentation system and also influence the quality of the device.

Figure 3-4 shows a block diagram of elements of a typical instrumentation system. The basic subsystems include the following modules.

- Sensing module
- Conversion module
- Variable manipulation module
- Data transmission
- Presentation module

FIGURE 3-4 ELEMENTS OF AN INSTRUMENTATION SYSTEM



The integrated effect of all the functional modules results in a useful measurement system. A description of each module is given here.

Sensing Module The first element to receive a signal from the measured medium and produces an output depending on the measured quantity. During the process of sensing, some energy gets extracted from the measured medium. In fact, the measured quantity gets disturbed by the act of measurement, making a perfect measurement theoretically impossible. Good instruments are normally designed to minimize the error of measurement.

Conversion Module Converts one physical variable to another. It is also known as a transducing element. In certain cases, the transduction of the input signal may take place progressively in stages, such as primary, secondary, and tertiary transduction.

Variable Manipulation Module Usually, this involves signal conditioning. Some examples of variable manipulation element are amplifiers, linkage mechanisms, gearboxes, magnifiers, etc. An electronic amplifier accepts a small voltage signal as an input signal and generates a signal that is many times larger than the input signal.

Data Transmission Module This sends a signal from one point to another point. For example, the transmission element could be a simple device such as a shaft and bearing assembly or could be a complicated device, such as a telemetry system for transmitting signals from ground to satellites.

Data Display Module Produces information about the measured quantity in a form that can be recognized by one of the human senses.

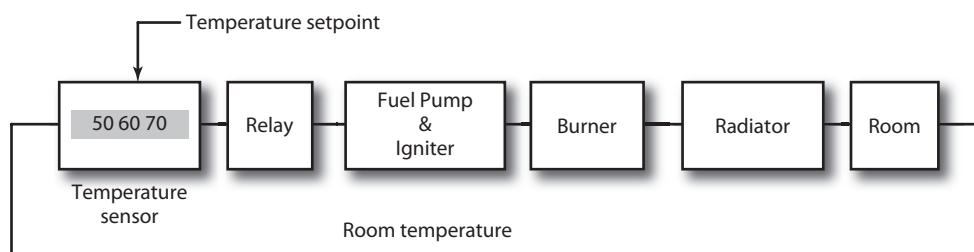
EXAMPLE 3.1 Home Heating System

The functional elements of a typical home heating system are shown in Figure 3-5.

Solution

The block diagram represents the six major system components and their interconnections. The interconnections completely define the inputs and outputs for each of the six major blocks. For instance, the thermostat block processes two input signals (a room temperature and a temperature set point,) to produce one output signal, which is sent to a mechanical relay switch. The thermostat acts as a primary sensor and transducer.

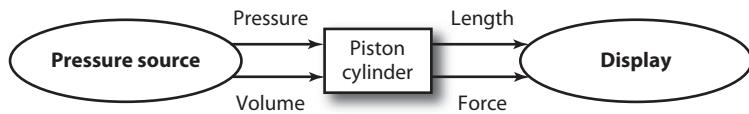
FIGURE 3-5 HOME HEATING SYSTEM EXAMPLE



EXAMPLE 3.2 Pressure Sensor

An example of a pressure sensor in the form of a spring-loaded piston and a display mechanism is shown in Figure 3-6. This pressure sensing instrument can be broken down into functional elements. The source is connected to a pneumatic cylinder. The pressure acts on the piston and spring mechanism. The spring works as a primary sensor and variable conversion element. The deflection of the spring is transferred to the display as a movement of the dial indicator.

FIGURE 3-6 SCHEMATIC OF A PRESSURE SENSOR



3.1.3 Quality Parameters

Sensors and transducers are often used under different environmental conditions. Like human beings, they are sensitive to environmental inputs such as pressure, motion, temperature, radiation, and magnetic fields.

Sensor characteristics are described in terms of seven properties discussed and illustrated in the following subsections.

- Sensitivity
- Resolution
- Accuracy
- Precision
- Backlash
- Repeatability
- Linearity

Sensitivity Sensitivity is the property of the measuring instrument to respond to changes in the measured quantity. It also can be expressed as the ratio of change of output to change of input as shown in Figures 3-7 and 3-8.

FIGURE 3-7 BASIC TRANSDUCER MODEL

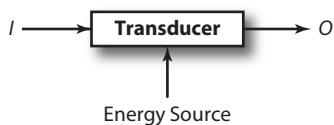
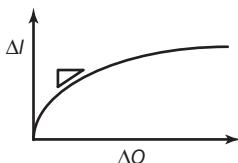


FIGURE 3-8 INPUT-OUTPUT RELATIONSHIP

Sensitivity is measured by

$$S = \frac{\Delta O}{\Delta I}$$

where S is the sensitivity, ΔO represents change in output, and ΔI represents the change in input. For example, in an electrical measuring instrument if a movement of 0.001 mm causes an output voltage change of 0.02 V, the sensitivity of the measuring instrument is $S = \frac{0.02}{0.001} = 20 \text{ V/mm}$

Resolution Resolution is defined as the smallest increment in the measured value that can be detected. It is also known as the degree of fineness with which measurements can be made. For example, if a micrometer with a minimum graduation of 1 mm is used to measure to the nearest 0.5 mm, then by interpolation, the resolution is estimated as 0.5 mm.

Accuracy Accuracy is a measure of the difference between the measured value and actual value. Accuracy depends on the inherent instrument limitations. An experiment is said to be accurate if it is unaffected by experimental error. An accuracy of ± 0.001 means that the measured value is within 0.001 units of actual value. In practice, the accuracy is defined as a percentage of the true value.

$$\text{Percentage of true value} = \frac{\text{measured value} - \text{true value}}{\text{true value}} (100)$$

If a precision balance reads 1 g with error of 0.001 g, then the accuracy of the instrument is specified as 0.1%. The difference between the measured value and true value is called bias (error).

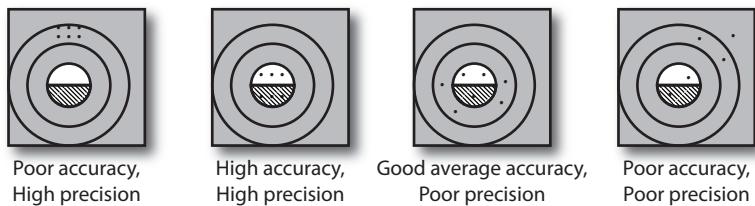
Precision Precision is the ability of an instrument to reproduce a certain set of readings within a given accuracy. Precision is dependent on the reliability of the instrument.

EXAMPLE 3.3 Target Shooting

Figure 3-9 presents an illustration of degree of accuracy and precision in a typical target-shooting example.

Solution

The “high precision, poor accuracy” situation occurs when the person hits all the bullets on a target plate on the outer circle and misses the bull’s eye. In the second case, “high accuracy, high precision”, all the bullets hit the bull’s eye and are spaced closely enough. In the third example, “good accuracy, poor precision”, the bullet hits are placed symmetrically with respect to the bull’s eye but are spaced apart. In the last case, “poor accuracy, poor precision”, the bullets hit the target in a random order.

FIGURE 3-9 TARGET SHOOTING EXAMPLE

Backlash Backlash is defined as the maximum distance or angle through which any part of a mechanical system can be moved in one direction without causing any motion of the attached part. Backlash is an undesirable phenomenon and is important in the precision design of gear trains.

Repeatability Repeatability is the ability to reproduce the output signal exactly when the same measurand is applied repeatedly under the same environmental conditions.

Linearity The characteristics of precision instruments are that the output is a linear function of the input. However, linearity is never completely achieved, and the deviations from the ideal are termed linearity tolerances. The linearity is expressed as the percentage of departure from the linear value (i.e., maximum deviation of the output curve from the best-fit straight line during a calibration cycle). The nonlinearity is normally caused by nonlinear elements, such as mechanical hysteresis, viscous flow or creep, and electronic amplifiers.

3.1.4 Errors and Uncertainties in Mechatronic Modeling Parameters

Modern mechatronic technology relies heavily on the use of sensors and measurement technology. The control of industrial processes and automated systems would be very difficult without accurate sensors and measurement systems. The economical production of a mechatronic instrument requires the proper choice of sensors, material, and hardware and software design. To a large degree, the final choice of an instrument for any particular application depends upon the accuracy desired. If a low degree of accuracy is acceptable, it is not economical to use expensive sensors and precise sensing components. If, however, the instrument is used for high-precision applications, the design tolerances must be small.

Any system which relies on a measurement system will involve some amount of uncertainty. The uncertainty may be caused by the individual inaccuracy of sensors, random variations in measurands, or environmental conditions. The accuracy of the total system depends on the interaction of the components and their individual accuracy. This is true for measurement instruments as well as production systems, which depend on many subsystems and components. A typical instrument may consist of many components that have complex interrelations, and each component may contribute to the overall error. The errors and inaccuracies in each of these components can have a large cumulative effect.

3.2 Sensitivity Analysis—Influence of Component Variation

The accuracy and precision of a complex die mechanism in a manufacturing environment depends upon its design and on the design tolerances of its interrelated parts. Similarly, if an experiment has a number of component sources—each being measured individually using independent instruments—a procedure to compute the total accuracy is necessary. From the point of view of the total system, this

procedure must also account for individual variations in component tolerances. The *error analysis* method helps us to identify the contribution of component error in accuracy calculations. The procedure also helps to allocate individual design tolerances or variations if the total design tolerance or variation is known. An illustrative example is presented next.

Let us consider the problem of computing a quantity N that is a known function of n independent variables, $x_1, x_2, x_3, \dots, x_n$ which are the measured quantities of one instrument (or component output of different instruments contributing to one system).

$$N = f(x_1, x_2, \dots, x_n) \quad (3-1)$$

Let $\pm\Delta x_1, \pm\Delta x_2, \dots, \pm\Delta x_n$ be the individual errors in each of the quantities. These errors will cause total error in the computed result N shown in Equation 3-1.

$$N \pm \Delta N = f(x_1 \pm \Delta x_1, x_2 \pm \Delta x_2, \dots, x_n \pm \Delta x_n) \quad (3-2)$$

We obtain ΔN by subtracting N from $N \pm \Delta N$. Since the procedure is time consuming, approximate solutions can be obtained using Taylor's series. Expanding Equation 3-2 in a Taylor series produces

$$\begin{aligned} f(x_1 \pm \Delta x_1, x_2 \pm \Delta x_2, \dots, x_n \pm \Delta x_n) &= f(x_1, x_2, \dots, x_n) + \Delta x_1 \frac{\partial f}{\partial x_1} \\ &\quad + \Delta x_2 \frac{\partial f}{\partial x_2} + \frac{1}{2} (\Delta x_1)^2 \frac{\partial^2 f}{\partial x_1^2} + \dots + \dots \end{aligned} \quad (3-3)$$

All partial derivatives in the series are evaluated at the known values of $x_1, x_2, x_3, \dots, x_n$. Since the measurements have been taken, the x_i 's are all known values, which can be substituted into the expressions for the partial derivatives to produce appropriate values.

In practice, the Δx 's will be small quantities, hence Δx^2 terms are negligible. Equation 3-3 then reduces to

$$\begin{aligned} f(x_1 \pm \Delta x_1, x_2 \pm \Delta x_2, \dots, x_n \pm \Delta x_n) &= f(x_1, x_2, \dots, x_n) + \Delta x_1 \frac{\partial f}{\partial x_1} + \\ &\quad \Delta x_2 \frac{\partial f}{\partial x_2} + \dots + \Delta x_n \frac{\partial f}{\partial x_n} \end{aligned} \quad (3-4)$$

The absolute error, E_a , is defined by

$$E_a = \Delta N = \Delta x_1 \frac{\partial f}{\partial x_1} + \Delta x_2 \frac{\partial f}{\partial x_2} + \dots + \Delta x_n \frac{\partial f}{\partial x_n} \quad (3-5)$$

The absolute value is used because some of the partial derivatives may be negative and would have a canceling effect. Equation 3-5 is useful because it illustrates which of the variables exert the strongest influence on the accuracy of overall results.

For example, if the term $\frac{\partial f}{\partial x_3}$ were high compared with other partial derivative terms then a small Δx_3 would have a large effect on the total error E_a .

$$\text{Percentage error } E_r = \frac{\Delta N}{N} \times 100 = \frac{100 E_a}{N}$$

$$\text{Computed results} = N \pm \frac{\Delta N}{N} \times 100$$

In certain situations, the limitation on the total accuracy is known, but the design limits on the accuracy of individual components is not known. In such cases, if the overall accuracy is known and if one wishes to find the individual component accuracies that are needed, the method *of equal effects* is employed. This assumes that each source of error would contribute an equal amount to the total error.

$$\Delta N = \frac{\partial f}{\partial x_1} \Delta x_1 + \frac{\partial f}{\partial x_2} \Delta x_2 + \dots + \frac{\partial f}{\partial x_n} \Delta x_n$$

Assuming each term to be of equal importance, we may write

$$\frac{\partial f}{\partial x_1} \Delta x_1 = \frac{\partial f}{\partial x_2} \Delta x_2 = \dots = \frac{\partial f}{\partial x_n} \Delta x_n = \frac{\Delta N}{n} \quad (3-6)$$

Now that the allowable overall error ΔN is known, and since $x_1, x_2, x_3, \dots, x_n$ are also known, we may write

$$\therefore \frac{\partial f}{\partial x_i} \Delta x_i = \frac{\Delta N}{n}$$

The allowable error Δx_i in each measurement is calculated by solving for Δx_i as

$$\Delta x_i = \frac{\Delta N}{n \left(\frac{\partial f}{\partial x_i} \right)} \text{ where } i = 1, 2, 3, \dots, n \quad (3-7)$$

The method of equal effects, summarized in Equation 3-6, considers the absolute values of all variables and gives an estimate of the maximum uncertainty of the measured variable in terms of N .

Another method known as the *square root of sum of squares* (RSS) is based on the fact that all uncertainties are evaluated at the same confidence level. This is shown in Equation 3-8. Whenever the RSS method is applied, the confidence level of the uncertainty in the result N will be the same as the confidence levels of the uncertainties in the x_i 's.

$$\Delta N = \sqrt{\sum_{i=1}^{i=n} \left(\Delta x_i \frac{\partial f}{\partial x_i} \right)^2} \quad (3-8)$$

Three examples illustrating the uncertainty calculations previously discussed are presented in the following sections.

EXAMPLE 3.4 Speed Control System Example

A mechatronic speed control system is used where the relationship between the angular velocity and the force applied is given by the expression:

$$\omega = \sqrt{\frac{F}{mr}}$$

where

F is the force applied in newtons

r = radius of rotation

m = mass of the rotating weight

If $m = 200 \pm 0.01$ g, $r = 25 \pm 0.01$ mm, and $F = 500 \pm 0.1\%$ (N), determine the uncertainty in the rotational speed.

Solution

The speed is computed using the formula,

$$\omega = \sqrt{\frac{F}{mr}}$$

$$\omega = \sqrt{\frac{500}{(0.2)(0.025)}} = 316.23$$

Consider each component of error contributing to the measurement of the angular velocity.

$$E_a = \Delta N = \left[\Delta x_1 \frac{\partial f}{\partial x_1} \right] + \left[\Delta x_2 \frac{\partial f}{\partial x_2} \right] + \dots$$

Computing various partial derivatives,

$$\frac{\partial \omega}{\partial m} = \frac{-0.5\sqrt{F}}{m^{\frac{3}{2}}\sqrt{r}} = \frac{-0.5\sqrt{500}}{(0.2)^{\frac{3}{2}}\sqrt{0.025}} = -790.57$$

$$\frac{\partial \omega}{\partial F} = \frac{1}{2\sqrt{F}} \cdot \frac{1}{\sqrt{mr}} = \frac{1}{2\sqrt{500}} \cdot \frac{1}{\sqrt{(0.025)(0.2)}} = 0.3162$$

$$\frac{\partial \omega}{\partial r} = -\frac{1}{2} \sqrt{\frac{F}{m}} \cdot \frac{1}{r^{\frac{3}{2}}} = -\frac{1}{2} \sqrt{\frac{500}{0.2}} \cdot \frac{1}{(0.025)^{\frac{3}{2}}} = 6324.56$$

$$E_a = \Delta N = (0.5)(0.316) + (1)(10^{-5})(790.57) + (1)(10^{-5})(6324.56) = 0.229$$

$$\text{Error} = \frac{\Delta N}{N} = \frac{0.229}{316.23} = 0.000725 \approx 0.072\%$$

EXAMPLE 3.5 RLC Circuit

The impedance of the RLC circuit operating on alternating current is given by the equation

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$

If the uncertainty in each of R , L , and C is 5%, calculate the uncertainty in the measurement of Z . The resistance R is given as 2 kΩ, the inductance L is 0.8 H, and the capacitance C is 5 μF.

Solution

The impedance equation is

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$

where

$$X_L = \omega L = 2\pi fL$$

$$X_C = \frac{1}{2\pi fC}$$

$$R = 2^k \Omega \pm 5\% = 2000 \pm 100 \Omega$$

$$L = 0.8H \pm 5\% = 0.8 \pm 0.04 H$$

$$\begin{aligned} C &= 5\mu F \pm 5\% \text{ or } (5)(10^{-6}) \pm (0.25)(10^{-6})F \\ &= (5)(10^{-6}) \pm (250)(10^{-9})F \end{aligned}$$

$$f = 60 \text{ Hz}$$

$$X_L = 2\pi fL = 2(\pi)(60)(0.8) = 301.6$$

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2(\pi)(60)(5)(10^{-6})} = 530.52$$

$$Z = \sqrt{R^2 + (X_L - X_C)^2} = \sqrt{2000^2 + (301.6 - 530.52)^2} = 2013$$

Partial derivatives,

$$\frac{\partial Z}{\partial R} = \frac{R}{\sqrt{R^2 + (X_L - X_C)^2}} = \frac{2013}{\sqrt{2000^2 + (301.6 - 530.52)^2}} = 0.99$$

$$\frac{\partial Z}{\partial X_L} = \frac{X_L - X_C}{\sqrt{R^2 + (X_L - X_C)^2}} = \frac{301.6 - 530.52}{\sqrt{2000^2 + (301.6 - 530.52)^2}} = -0.114$$

$$\frac{\partial Z}{\partial X_C} = \frac{X_C - X_L}{\sqrt{R^2 + (X_L - X_C)^2}} = \frac{530.52 - 301.6}{\sqrt{2000^2 + (301.6 - 530.52)^2}} = 0.114$$

$$\Delta N = 0.999(100) + 0.114(0.04) + 0.114(250)10^{-9} = 99.9$$

which is 4.96%.

EXAMPLE 3.6 Resistance Measurement

Constantan is an alloy (with 55% copper and 45% nickel), which is used in the construction of strain gauges. It has a resistivity of $49 \times 10^{-8} \Omega \cdot \text{m}$. The length of the constantan wire is calculated using the formula,

$$L = \frac{RA_c}{\rho_c}$$

where

$$R = 90 \Omega,$$

$$A_c = 7.85 \times 10^{-7} \text{ m}^2$$

If the uncertainty in the measurement of R , A , and ρ is about 10% in each case, calculate the absolute error in the measurement of length of the wire. If the total error is to be limited to half of the calculated value above, how do you allocate the accuracy to individual measurements?

$$L = \frac{RA_c}{\rho_c} = \frac{(90)(7.85 \times 10^{-7})}{49 \times 10^{-8}} = 144.18$$

where

$$R = 90\Omega \pm 9$$

$$A_C = 7.85 \times 10^{-7}\text{m}^2 \pm 7.85 \times 10^{-8}\text{m}^2$$

$$\rho_C = 49 \times 10^{-8}\Omega\cdot\text{m} \pm 4.9 \times 10^{-8}\Omega\cdot\text{m}$$

Partial derivatives are

$$\frac{\partial L}{\partial R} = \frac{A_C}{\rho_C} = 1.602$$

$$\frac{\partial L}{\partial A_C} = \frac{R}{\rho_C} = 1.84 \times 10^8$$

$$\frac{\partial L}{\partial \rho_C} = -\frac{RA_C}{\rho_C^2} = -2.94 \times 10^8$$

$$\Delta N = (1.602)(9) + (1.84)(10^8)(7.85)(10^{-8}) + (2.94)(10^8)(4.9)(10^{-8}) = 43.25$$

$$\text{Percentage error} = \frac{43.25}{144.18} \times 100 = 30\%$$

If the error is limited to 15%, what accuracies will be allocated to individual measurement?

Solution

Error is limited to 15%; variation permitted in the parameters can be calculated using Equation 3-7.

$$R = \frac{(0.15)(144.18)}{(1.602)(3)} = \pm 4.50\Omega$$

$$A_C = \frac{(0.15)(144.18)}{(1.84)(10^8)(3)} = \pm 3.92(10)^{-8}\text{m}^2$$

$$\rho_C = \frac{(0.15)(144.18)}{(2.94)(10^8)(3)} = \pm 2.45(10)^{-8}\Omega\cdot\text{m}$$

3.3 Sensors for Motion and Position Measurement

An integrated manufacturing environment typically consists of

- Machining centers/manufacturing cells
- Inspection stations
- Material handling
- Devices
- Packaging centers
- Areas where the raw material and finished products are handled

The integrated system monitors the environment to understand the progress of the product in the production scheme. The sensors interact with the controllers and provide a detailed account of

status of the process as well as environmental conditions. The controller sends signals to the actuators, which respond according to the functions.

Sensor-based manufacturing systems consist of data measurement by a plurality of sensors, sensor integration, signal processing, and pattern recognition. Motion measurement (especially the measure of displacement, position, and velocity of physical objects) is essential for many feedback control applications (especially those used in robotics, process, and automotive industries). Motion transducers are a class of transducers used for the measurement of mechanical quantities that include:

- Displacement
- Force
- Pressure
- Flow rate
- Temperature

Primary and Secondary Transducers Sometimes the transducer measures one phenomenon in order to measure another variable. The primary transducer senses the preliminary data and converts it into another form, which is again converted into some usable form by a secondary transducer. As an example, measurement of force is performed using a spring element, and the resulting displacement of the spring is measured using another electrical transducer. The force causes the spring to extend and the mechanical displacement is proportional to the force. The spring is considered to be the *primary transducer*, which converts force into displacement. The end of the spring is connected to another electrical transducer, which senses its displacement and transmits it as an electrical signal. This electrical transducer is called a *secondary transducer*. In most measurement systems, it is common to have such combinations of transducer elements in which a primary transducer is the mechanical element, and an electrical transducer (acting in the secondary stage) is the secondary unit.

Selection Criteria for a Transducer

- The range of the measurement
- Suitability of the transducer for such measurement
- Required resolution
- Material of the measured object
- Available space
- Environmental conditions
- Power available for sensing
- Cost
- Production volume

Transducers of the electrical, electromechanical, optical, pneumatic, and piezoelectric types are commonly used in motion measurement.

Transducer Classification Based on the Principle of Transduction

- **Potentiometric:** Potentiometric transducers apply the principle of *change in resistance* of material in the sensor.
- **Capacitance:** Capacitance transducers apply the principle of *capacitance variation* between a set of plate assemblies.
- **Inductance:** Inductance transducers are based on the principle of *variation of inductance* by the insertion of core material into an inductor. Inductance variations serve as a measure of displacement.
- **Piezoelectric:** Piezoelectric transducers are based on the principle of *charge generation*. Whenever certain piezoelectric crystals are subjected to mechanical motion, an electric voltage is induced. This effect can be reversed by applying an electric voltage and deforming the crystal.

3.3.1 Resistance Transducers

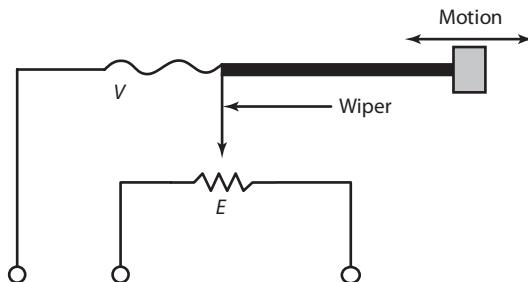
Potentiometric Principle A displacement transducer using variable resistance transduction principle can be manufactured with a rotary or linear potentiometer.

A potentiometer is a transducer in which a rotation or displacement is converted into a potential difference.

As shown in Figure 3-10, the displacement of the wiper of a potentiometer causes the output potential difference obtained between one end of the resistance and the slider. This device converts linear or angular motion into changing resistance, which may be converted directly to a voltage or current signal. The position of the slider along the resistance element determines the magnitude of the electrical potential. The voltage across the wiper of linear potentiometer is measured in terms of the displacement, d , and given by the relationship

$$V = E \frac{d}{L}$$

FIGURE 3-10 POTENTIOMETER TRANSDUCER PRINCIPAL



Here E is the voltage across the potentiometer, and L is the full-scale displacement of the potentiometer.

If the movement of the slider is in a circular path along a resistance element, then rotational information is converted into information in the form of a potential difference. The output of the rotary

transducer is proportional to the angular movement. If there is any loading effect from the output terminal, the linear relationship between the wiper position and the output voltage will change.

The error, which is called the *loading error*, is caused by the input impedance of the output devices. To reduce the loading error, a voltage source, which is not seriously affected by load variations (e.g., stabilized power source) and signal-conditioning circuitry with high-input impedance should be used. It is also advisable to isolate the wiper of the potentiometer from the sensing shaft.

The disadvantage of the potentiometric transducer is its slow dynamic performance, low resolution, and susceptibility to vibration and noise. However, displacement transducers with a relatively small traverse length have been designed using strain-gauge-type resistance transducers.

SUMMARY

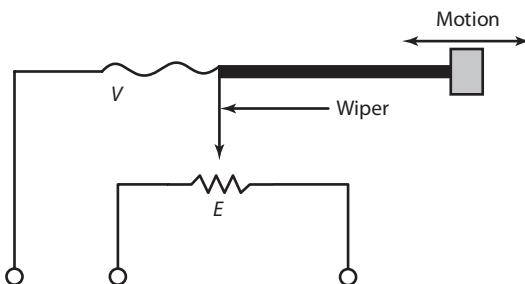
Potentiometric Principle

A transducer in which a rotation or displacement is converted into a potential difference.

This type of transducer (Figure 3-11) converts linear or angular motion into changing resistance, which is converted directly to a voltage or current signal. The position of the slider along the resistance element determines the magnitude of the electrical potential. The voltage across the wiper of linear potentiometer is measured in terms of the displacement, d , and given by the relationship

$$V = E \frac{d}{L}$$

FIGURE 3-11



where E is the voltage across the potentiometer and L is the full-scale displacement of the potentiometer.

Rotary Potentiometer

If the movement of the slider is in a circular path along a resistance element, rotational information is converted into information in the form of a potential difference. The output of the rotary transducer is proportional to the angular movement.

Features

- Linear potentiometers are often considered when an electrical signal proportional to displacement is required, but also where cost should be kept low and high accuracy is not critical.
- Typical rotary potentiometers have a range of $\pm 170^\circ$. Their linearity varies from 0.01 to 1.5%.

Applications

- Used for position monitoring of products on assembly lines and checking dimensions of the product in quality control systems.
- Rotary potentiometers are used in applications involving rotational measurement for applications ranging from machine tools to aircraft.

3.3.2 Inductance Transducers

Inductive transducers are used for proximity sensing and also for motion position detection, motion control, and process control applications.

Inductive transducers are based on the Faraday's law of induction in a coil. Faraday's law of induction specifies that the induced voltage, or *electromotive force* (EMF), is equal to the rate at which the magnetic flux through the circuit changes. If varying magnetic flux is applied to a coil, then electromotive force appears at every turn of the coil. If the coil is wound in such a manner that each turn has the same area of cross section, the flux through each turn will be the same. The induced voltage equation is shown in Equation 3-9.

$$V = N \frac{d\phi}{dt} \quad (3-9)$$

Here, N is the number of turns, and $\phi = BA$, where B is the magnetic field and A is the area of the coil. It follows that the voltage output can be changed by changing the flux enclosed by the circuit. This can be done by changing the amplitude of the magnetic field B or area of the coil A .

The equation can also be expressed as

$$V = N \frac{d(BA)}{dt} \quad (3-10)$$

Rewriting Equation 3-10 as

$$V = \frac{dN(\phi)}{dt} = \frac{d\psi}{dt} \quad (3-11)$$

where $\psi = N\phi$.

Here, N is the number of turns in the circuit, and ψ is the total flux linkages of the circuit. It is concluded that the voltage generated is equal to the rate of change of flux linkages. It is also known that the magnetic field B , produced by a current i in any circuit, is proportional to the current and geometry of the coil.

The total flux linkages of the circuit can be expressed in terms of a constant L , which is the inductance of the circuit. Inductance of the circuit is defined as the flux linkage per unit current, as given in Equation 3-12.

$$L = \frac{\psi}{i} = \frac{N\phi}{i} \quad (3-12)$$

Flux is defined as

$$\phi = \frac{Ni}{R} \quad (3-13)$$

In this equation, R is the reluctance of the flux path. The reluctance in a magnetic circuit is analogous to resistance in electrical circuits. Self-inductance of a coil is expressed by Equation 3-14 as

$$L = \frac{N}{i} \left(\frac{Ni}{R} \right) = \frac{N^2}{R} \quad (3-14)$$

where

N = number of turns

R = reluctance of the magnetic circuit

The reluctance is expressed as

$$R = \frac{l}{\mu A}$$

where

μ is the effective permeability of the medium in and around the coil

l is the length of the coil, m

A is the area of the cross section of the coil, m^2

The unit of inductance is called the *Henry* (H). Equation 3-15 shows that a change in self-inductance of the coil can be caused by changing the number of turns, the geometric configuration, or by a change of permeability of the magnetic material.

$$L = N^2 \mu \left(\frac{A}{l} \right) = N^2 \mu G \quad (3-15)$$

where $G = \frac{A}{l}$ = geometric factor.

The inductance change can be caused by variations in any of the following:

- Geometry of the coil by changing the number of turns in the coil.
- Effective permeability of the medium in and around the coil.
- Change of reluctance of the magnetic path or by variation of the air gap.
- Change of mutual inductance (by change of coupling between coils 1 and 2 with aiding or opposing field).

The change in self-inductance caused by the geometric configuration is the result of the coil arrangement. There are two parts of the coil mounted on iron cores. One part is stationary, and the other movable. The displacement changes the position of the movable part of the coil, which produces a change in the self-inductance of the coil.

Transducers also can be designed which utilize variations in the number of turns. The output relationship becomes

$$L \propto N^2 \propto (\text{displacement})^2 \quad (3-16)$$

Change in Mutual Inductance Inductive transducers based on the principle of variation of mutual inductance use multiple coils. The presence of an induced emf in a circuit due entirely to a change of current in another circuit is called *mutual induction*.

To illustrate, consider two coils, 1 and 2, with turns N_1 and N_2 , respectively. The current i_1 flowing in coil 1, produces a flux φ . If R is the reluctance of the magnetic path, the induced emf in coil 2 due to current in coil 1 is

$$e_2 = N_2 \frac{d(\varphi)}{dt} = N_2 \frac{d(N_1 i_1 / R)}{dt} \quad (3-17)$$

$$\begin{aligned} e_2 &= \frac{N_1 N_2}{R} \frac{di_1}{dt} \\ e_2 &= M \frac{di_1}{dt} \end{aligned} \quad (3-18)$$

where mutual inductance is

$$M = \frac{N_1 N_2}{R}$$

In the same fashion, emf induced in coil 2 due to change in current in coil 1 is

$$e_1 = M \frac{di_2}{dt} \quad (3-19)$$

The expression of mutual inductance is modified by the factor K , which represents the loss in flux linkages between two coils:

$$\text{Mutual inductance; } M = \frac{N_1 N_2}{R} K \quad (3-20)$$

From Equation 3-14, we know that

$$\begin{aligned} L_1 &= \frac{N_1^2}{R}, \quad L_2 = \frac{N_2^2}{R} \\ L_1 L_2 &= \frac{N_1^2 N_2^2}{R^2} \end{aligned} \quad (3-21)$$

Using Equations 3-20 and 3-21, the mutual inductance is expressed as

$$M = K \sqrt{L_1 L_2} \quad (3-22)$$

In Equation 3-22, K is known as the *coefficient of coupling* between the two coils. Thus, mutual inductance between the coils can be changed by variations in either of the self-inductances or the coefficient of coupling.

Inductance transducers for measuring displacement use the principle of change in mutual inductance of a coil at varying core positions. When the core is centrally located, the voltage induced in each secondary is the same. When the core is displaced, the change in flux linkage causes one secondary voltage to increase and the other to decrease. The secondary windings are generally connected in series opposition, so the voltage induced in each are out of phase with the other. The output voltage is zero when a core is centrally located and increases as the core is moved either in or out. The voltage amplitude is linear with core displacement over some range of core

travel. The signal-conditioning circuit produces a voltage output, which is proportional to the displacement. The polarity of the output voltage derivative is relative to the direction of core motion.

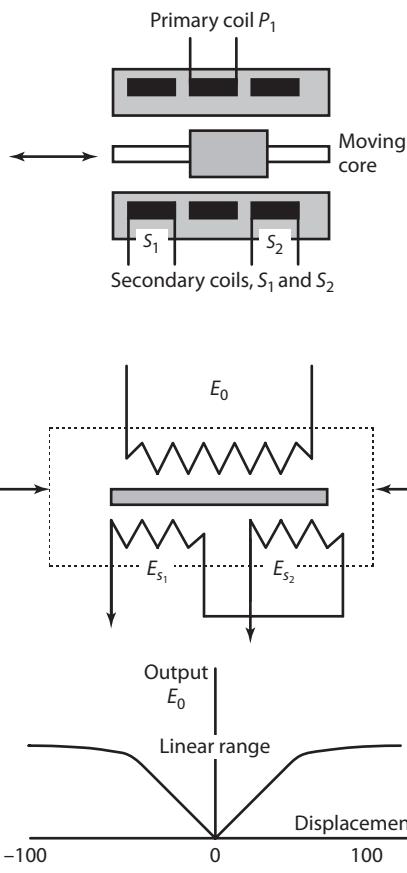
3.3.3 Linear Variable Differential Transformer (LVDT)

LVDTs are the most widely used transducers. They are used to measure displacement directly as a sensing element in a number of situations involving motion. LVDTs can resolve very small displacements. Their high resolution, high accuracy, and good stability make them an ideal device for applications involving short displacement measurements.

LVDTs consist of one primary winding, P_1 , and two secondary windings, S_1 and S_2 . Each is wound on a cylindrical former with rod-shaped magnetic cores positioned centrally inside the coil assemblies. This provides a dedicated path for the magnetic flux linking the coils. An oscillating excitation voltage is applied to the primary coil. The current through the primary creates voltages in secondary windings. The ferromagnetic core concentrates the magnetic field. If the core is closer to one of the secondary coils, the voltage in that coil will be higher.

Let the output of the secondary winding S_1 be E_{S1} and that of S_2 be E_{S2} . When the core is at its normal null position, equal voltages are induced in each coil. When these two outputs are connected in phase opposition, as shown in Figure 3-12, the magnitude of the resultant voltage will be zero.

FIGURE 3-12 SCHEMATIC OF LINEAR VARIABLE DIFFERENTIAL TRANSFORMER (LVDT)



This is known as the null position, and the output E_{s1} will be equal to E_{s2} . As the moving core is displaced, the mutual inductance between the fixed coils changes. The LVDT outputs a bipolar voltage proportional to displacement. The output voltage is positive and gives no indication of the direction in which the core has been moved. Proper signal conditioners can be designed to give indication of the direction.

LVDTs have limitations when used for dynamic measurements. They are not well suited for frequencies greater than 1/10 of the excitation frequency. In addition, the mass of the core introduces some amount of mechanical loading error. Proper selection of a LVDT depends on the range of displacement measurement. The voltage versus displacement is linear up to a certain point, but nonlinear beyond that region. The sensitivity of the transducer is also dependent on the excitation signal frequency, f , and the primary current, I_p . For good results, the linearity range of travel should be limited to the width of the primary coil. Typical LVDT range is from ± 2 to ± 400 mm with non-linearity errors of about $\pm 0.25\%$.

The signal output E_0 , in relation to the other characteristics of the coil, is given by Equation 3-23.

$$E_0 = \frac{16\pi^3 f I_p n_p n_s}{10^9 \ln\left(\frac{r_0}{r_1}\right)} \frac{2bx}{3w} \left(1 - \frac{x^2}{2b^2}\right) \quad (3-23)$$

where

f = excitation signal frequency

I_p = primary current

n_p = number of turns in primary

n_s = number of secondary turns

b = width of primary coil

w = width of secondary coil

x = core displacement

r_0 and r_i = the outer and inner radius of the coil

3.3.4 Rotary Variable Differential Transformer (RVDT)

The RVDT can be used wherever precision angular rotations are measured. The RVDT uses the same principle as LVDT, except it has a rotating magnetic core. Some RVDTs have a typical range of $\pm 40^\circ$ with a linearity error around $\pm 0.5\%$ of the range. Although LVDTs and RVDTs are used as primary transducers, they also can be used as a secondary transducer in areas of measurement of force, weight, pressure, and flow.

Typical applications of inductance transducers include the following.

- Measurement of the thickness of plates.
- Detection of dimensional changes in parts after they are manufactured.
- Angular speed measurement of a rotating device.
- Precise detection of specimen size.
- Liquid level applications.
- Measurement of precision gap in welding applications.

SUMMARY**Linear Variable Differential Transformer****Principle:**

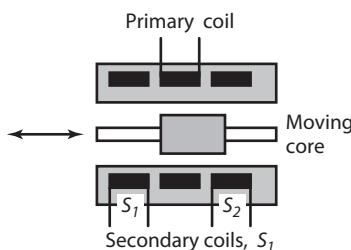
Based on the Faraday's law of induction in a coil, which specifies that the induced voltage, or electromotive force (EMF), is equal to the rate at which the magnetic flux through the circuit changes

$$V = N \frac{d\phi}{dt}$$

Here, N is the number of turns, and $\phi = BA$, where B is the magnetic field and A is the area of the coil.

Description:

Figure 3-13 consists of one primary winding P_1 and two secondary windings S_1 and S_2 , where each is wound on a cylindrical former with rod-shaped magnetic cores positioned centrally inside the coil assemblies. This provides a dedicated path for the magnetic flux linking the coils. An oscillating excitation voltage is applied to the primary coil. The current through the primary creates voltages in secondary windings. The ferromagnetic core concentrates the magnetic field. If the core is closer to one of the secondary coils, the voltage in that coil will be higher.

FIGURE 3-13**Rotary Variable Differential Transformer**

The RVDT uses the same principle as LVDT, except it has a rotating magnetic core.

Features

- High resolution, high accuracy, and good stability make them an ideal for applications involving short displacement measurements.
- Sensitive transducers provide resolution down to about 0.05 mm. They have operating ranges from about ± 0.1 to ± 300 mm.
- Accuracy is ± 0.5 mm of full-scale reading.
- Less sensitive to wide ranges in temperature than potentiometers.

Applications

- Measurement of precision gap between weld torch and work surface in welding applications.
- Measurement of the thickness of plates in rolling mills.
- Detection of surface irregularity of parts after they are machined.

- Angular speed measurement of a rotating device.
 - Precise detection of specimen size.
 - Liquid level applications.
-

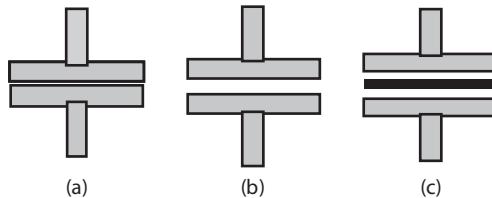
3.3.5 Capacitance Transducers

The variation in capacitance between two separated members (electrodes) is used for the measurement of many physical phenomenon. Capacitance is a function of the effective area of the conductors, the separation between the conductors, and the dielectric strength of the material. A change in capacitance can be brought about by varying any one of the three parameters. These variations are summarized here.

- Changing the distance between the two parallel electrodes.
- Changing the dielectric constant, permittivity, of dielectric medium ϵ .
- Changing the area of the electrodes, A .

Figure 3-14 illustrates the variable capacitance principle for displacement measurement utilizing the parallel-plate capacitor. In Figures 3-14(a) and 3-14(b), the gap is varied, and Figure 3-14(c) presents the situation where a dielectric material is inserted between the parallel plates.

FIGURE 3-14 PRINCIPLE OF VARIABLE CAPACITANCE



The ratio of the amount of charge stored on one of the plates to the amount of voltage across the capacitor is the *capacitance*. The capacitance is directly proportional to the area of plates and inversely proportional to the distance between them. The governing equation is given in Equation 3-27.

$$C = \frac{\epsilon A}{d} \quad (3-27)$$

The constant of proportionality ϵ , known as the *permittivity*, is a function of the type of material separating the plates. For a capacitance transducer with insulating material, the capacitance between the plates is defined as

$$C = \frac{\epsilon_r \epsilon_0 A}{d} \text{ farads, F} \quad (3-28)$$

where

ϵ_r = dielectric constant of the insulating medium (for air $\epsilon_r = 1$)

ϵ_0 = permittivity of air or free space (in a vacuum), which is $8.85 \times 10^{-12} \text{ F/m}$,

$$8.85 \text{ pF/m, or } \frac{1}{36\pi(10^9)} \text{ F/m}$$

A = overlapping area in plates, m^2

d = distance between electrodes or plates, m

This equation establishes a relationship between the plate area and the distance between the plates. Varying either of them linearly changes the capacitance, which can be measured by a circuit. The equation is valid for parallel-plate capacitors. However, if the geometry of the electrodes changes, the equation must be modified.

Variable capacitance transducers have applications in the area of liquid level measurement, chemical plants, and in situations where non-conductors are required. Let ΔA , Δd , and ΔC represent the changes in area, position, and capacitance, respectively. ΔC can be represented as

$$\frac{\Delta C}{C} = -\frac{\Delta d}{d} \quad (3-29)$$

$$\frac{\Delta C}{C} = \frac{\Delta A}{A}$$

Capacitance Transducers Using Change in Distance Between Plates Figure 3-15 illustrates a typical arrangement of a capacitance transducer that employs plate distance variations causing a change in capacitance. The right plate is fixed, and the left plate is movable by the displacement which is to be measured. The capacitance is computed as

$$C = \frac{\epsilon_r \epsilon_0 A}{d}$$

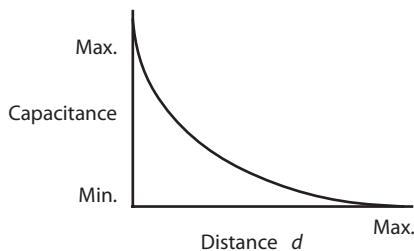
FIGURE 3-15 CAPACITANCE CHANGE DUE TO PLATE SEPARATION



If air is the dielectric medium, $\epsilon_r = 1$. The capacitance is inversely proportional to the distance between the plates. The overall response of the transducer is not linear, as shown by the distance versus capacitance plot of Figure 3-16; however, transducers of this type are used for the measurement of extremely small displacements where the relationship is approximately linear.

The sensitivity factor is expressed as

$$S = \frac{\partial C}{\partial d} = \frac{-\epsilon_r \epsilon_0 A}{d^2}$$

FIGURE 3-16 VARIATION OF CAPACITANCE WITH DISTANCE

Capacitance Transducers Using Change in Area of Plates For parallel-plate capacitors, the capacitance is

$$C = \frac{\epsilon_r \epsilon_0 A}{d} = \frac{\epsilon_r \epsilon_0 Lw}{d} \quad (3-30)$$

where

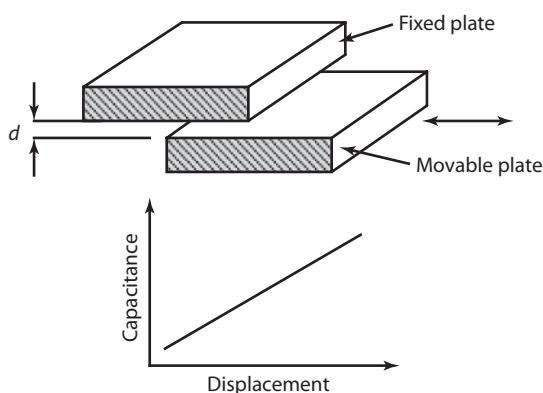
L = the length of overlapping part of plates

w = the width of overlapping part of plates

The sensitivity of the capacitance transducer becomes

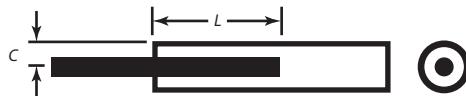
$$S = \frac{\partial C}{\partial l} = \frac{\epsilon_r \epsilon_0 Lw}{d} \text{ F/m} \quad (3-31)$$

There is a linear relationship between displacement and the capacitance. The preceding equations show that the capacitance is directly proportional to the area of the plates and varies linearly with changes in the displacement between the plates. Transducers of this type are used for the measurement of relatively large displacements (Figure 3-17).

FIGURE 3-17 CAPACITANCE VARIATION BY CHANGE IN AREA

Capacitance Transducers Using Change in Area (Cylindrical Shapes) A cylindrical-shaped capacitor consists of two coaxial cylinders with the outer diameter of the inner cylinder defined as D_1 , the inner diameter of outside cylinder as D_2 , and the length as L . Consider an example involving overlapping conductors, in which the inner cylinder can be moved with respect to the outer cylinder, causing a change in capacitance (Figure 3-18).

FIGURE 3-18 CHANGE IN AREA BASED ON CYLINDRICAL SHAPES

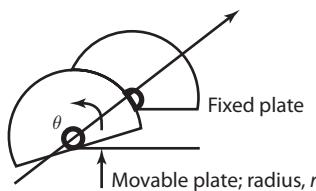


The capacitance is computed as

$$C = \frac{2\pi\epsilon_r\epsilon_0 L}{\ln \frac{D_2}{D_1}} \quad (3-32)$$

Capacitance Transducers for Angular Rotation The basic principle of change in area also can be used for rotational measurement. As shown in Figure 3-19, one plate is fixed and the other is movable. The angular displacement to be measured is applied to the movable plate. This angular displacement changes the effective area between plates and, thus, changes the capacitance. The capacitance is maximum when the plates completely overlap each other.

FIGURE 3-19 ANGULAR ROTATION OF PLATES

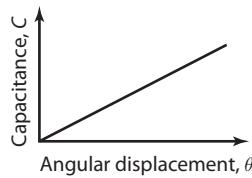


The maximum value of the capacitance is computed as

$$C = \frac{\epsilon A}{d} = \epsilon_0\epsilon_r \frac{\Pi r^2/2}{d} \quad (3-33)$$

The capacitance at angle θ (Figure 3-20) is computed as

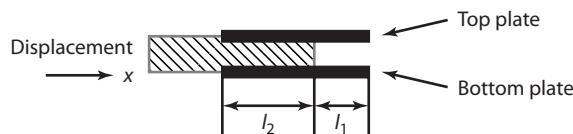
$$C = \epsilon_r\epsilon_0 \left(\frac{\theta}{2}\right) \frac{r^2}{d} \text{ F} \quad (3-34)$$

FIGURE 3-20 CAPACITANCE VARIATION ON ROTATION

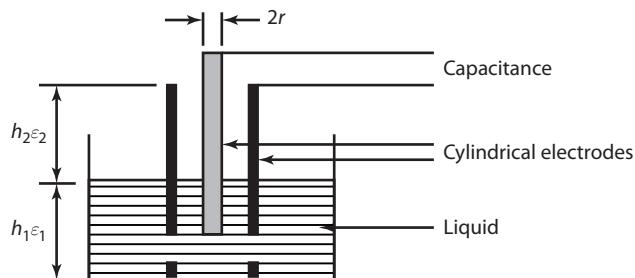
where angular displacement θ is in radians. The relationship is linear and the maximum angular displacement is 180° . The sensitivity is calculated as

$$S = \frac{\partial C}{\partial \theta} = \frac{\epsilon_r \epsilon_0}{2d} r^2$$

Capacitance Transducers Using Variation of Dielectric Constant The change in capacitance caused by a change in the dielectric constant of the separating material is another principle which can be used in capacitance transducers. Figure 3-21 shows an arrangement of two plates separated by a material of different dielectric constant. As this material is moved, it causes a variation of dielectric constant in the region separating the two electrodes, resulting in a change in capacitance.

FIGURE 3-21 TWO PLATES SEPARATED BY A MATERIAL OF DIFFERENT DIELECTRIC CONSTANT

As shown in Figure 3-22, the top plate and bottom plate are partially separated by the dielectric material. As the material moves a distance x as shown, the distance l_1 decreases and l_2 increases.

FIGURE 3-22 VARIATION OF CAPACITANCE BY DIELECTRIC CONSTANT

The initial value of the capacitance, assuming a dielectric material of thickness d and width w , can be described as

$$C = \frac{\epsilon_0 \epsilon_r w l_1}{d} + \frac{\epsilon_0 \epsilon_r w l_2}{d} \quad (3-35)$$

$$C = \frac{\epsilon_0 w}{d} \{l_1 + \epsilon_r l_2\}$$

Equation 3-35 has two terms. One represents the capacitance of the two electrodes separated by air, and the other represents the capacitance of the dielectric material between the electrodes.

If the dielectric material is moved through a distance x , as shown in Figure 3-22, the capacitance increases from C to $C + \Delta C$, and the change in capacitance is shown as

$$C + \Delta C = \frac{\epsilon_0 w}{d} \{l_1 - x + \epsilon_r(l_2 + x)\} \quad (3-36)$$

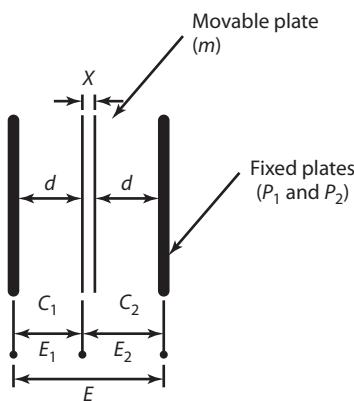
$$C + \Delta C = \frac{\epsilon_0 w}{d} \{\{l_1 + \epsilon_r l_2 + x(\epsilon_r - 1)\}\}$$

$$\Delta C = \frac{\epsilon_0 w x (\epsilon_r - 1)}{d}$$

The change in capacitance is proportional to the displacement x . This principle is also used in devices for measuring levels in nonconducting liquids. As shown in Figure 3-22, the electrodes are two concentric cylinders and the nonconducting liquid provides a dielectric medium between them. At the lower end of the outer cylinder, there are holes which allow passage of liquid. As the fluid level changes, the dielectric constant between the electrodes changes, which subsequently results in a change in capacitance.

Capacitance Transducers Based on Differential Arrangement Differential capacitance transducers are also used for precision displacement measurement. Figure 3-23 shows two fixed plates and a movable plate to which the displacement is applied.

FIGURE 3-23 DIFFERENTIAL ARRANGEMENT OF PLATES



Let C_1 and C_2 be the capacitances of two plates which are fixed. Plate m is midway between the two plates. An alternating voltage, E , is applied across the plates, P_1 and P_2 , and the potential differences across the two capacitors is measured. Assuming $\varepsilon = \varepsilon_o \varepsilon_r$ the following equations are written,

$$C_1 = \frac{\varepsilon A}{d}, C_2 = \frac{\varepsilon A}{d} \quad (3-37)$$

$$\text{Voltage across } C_1: E_1 = \frac{EC_2}{C_1 + C_2} = \frac{E}{2}$$

$$\text{Voltage across } C_2: E_2 = \frac{EC_1}{C_1 + C_2}$$

At midway point, $E_1 - E_2$ is zero. If x is the displacement of movable plate,

$$C_2 = \frac{\varepsilon A}{d - x}, C_1 = \frac{\varepsilon A}{d + x}$$

The differential output voltage is

$$\Delta E = E_1 - E_2 = \frac{(d + x)}{2d} E - \frac{(d - x)}{2d} E = \frac{x}{d} E$$

The output voltage varies linearly with displacement x . Capacitance transducers based on differential arrangement are used for measurement applications in the range of 0.001 to 10 mm and provide accuracy up to 0.05%. The sensitivity of the transducer is

$$S = \frac{\Delta E}{x} = \frac{E}{d} \quad (3-38)$$

A capacitive transducer is a displacement-sensitive transducer. A suitable processing circuit is necessary to generate a voltage corresponding to the capacitance change. General losses in the capacitance are attributed to

- DC leakage resistance
- Dielectric losses in the insulators
- Losses in the dielectric gap

Capacitance transducers have several advantages. They require extremely small forces to operate, are very sensitive, and require low power to operate. Their frequency response is good up to 50 kHz, making them good candidates for applications involving dynamics. Disadvantages include the need to insulate metallic parts from each other and loss of sensitivity due to error sources associated with the cable connecting the transducer to the measuring point.

Other Arrangements

1. Three material configuration:

$$C = \frac{A}{(36)(10^9) \cdot \pi \left(\frac{d_1}{\varepsilon_1} + \frac{d_2}{\varepsilon_2} + \frac{d_3}{\varepsilon_3} \right)}$$

Indices 1, 2, and 3 indicate layers of different permittivity and thickness, d , for a configuration with three materials.

2. Alternately connected multiplate configuration:

$$C = \frac{2\epsilon_r A}{36 \cdot 10^9 \pi}$$

This is the expression of capacitance for a transducer of n alternately connected plates. This transducer has $n - 1$ times the capacitance of one pair of plates.

SUMMARY

Capacitance Transducer

Principle:

Capacitance is a function of effective area of the conductors, the separation between the conductors, and the dielectric strength of the material. The governing equation is

$$C = \frac{\epsilon A}{d}$$

The constant of proportionality ϵ , known as the permittivity, is a function of the type of material separating the plates.

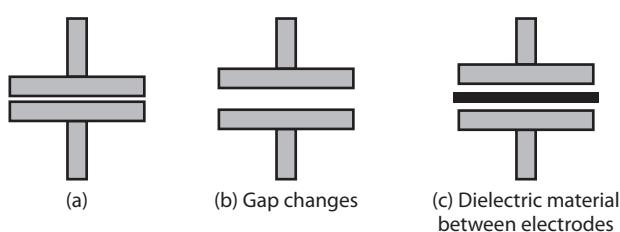
The variation in capacitance between two separated electrodes is used for the measurement of many physical phenomenon. A change in capacitance can be brought about by varying the following parameters.

- Changing the distance between the two parallel electrodes.
- Changing the dielectric constant, permittivity, of dielectric medium ϵ .
- Changing the area of the electrodes, A .

Description

Figure 3-24 shows the variable capacitance principle for displacement measurement.

FIGURE 3-24



Features

Capacitance transducers can be used in high humidity, high temperature, or nuclear radiated zones.

They are very sensitive and have high resolution. They can be expensive and need significant signal conditioners.

Applications

Capacitance transducers are generally only suitable for measuring small displacements. Examples of these are surface profile sensing, wear measurement, or crack growth.

3.4 Digital Sensors for Motion Measurement

Digital transducers are ideal devices for motion measurement. They produce a digital output which can be interfaced to the computer. They have become increasingly attractive because of the following properties.

- Signal conditioning simplicity
- Minor susceptibility to electro-magnetic interference

While they are used to measure linear or angular displacement, digital transducers also are used to measure force, pressure, and liquid level with the appropriate mechanical or electromechanical translators.

3.4.1 Digital Encoders

Encoders are widely used for applications involving measurement of linear or angular position, velocity, and direction of movement. They are used not only as a part of computerized machines but also in many precision-measurement devices, motion control applications, and quality assurance of equipment at various stages of production. Encoders are used in tensile-test instruments to precisely measure the ball screw position used to apply tension or compression to the test specimen. They are used in automated test stands used when angular positions of windshield wiper drives and switch positions are tested.

The most popular encoders are linear- or rotary-type optical encoders. Other configurations, such as contact-type encoders, have serious limitations due to contact wear and low resolution.

3.4.2 Encoder Principle

An encoder is a circular device in the form of a disk on which a digital pattern is etched. The inscribed pattern is sensed by means of a sensing head. The rotary disk is normally coupled to a shaft. As the shaft rotates, a different pattern is generated for each resolvable position. The sensing mechanism can be a photoelectric device with slots acting as transparent optical windows.

An optical encoder generally is used to precisely measure rotational movement. Its main advantages are simplicity, accuracy, and suitability for sensitive applications. Optical encoders are considered one of the most reliable and least expensive motion-feedback devices available and are used widely in a broad range of modern applications. Information obtainable from an optical encoder includes direction, distance, velocity, and position.

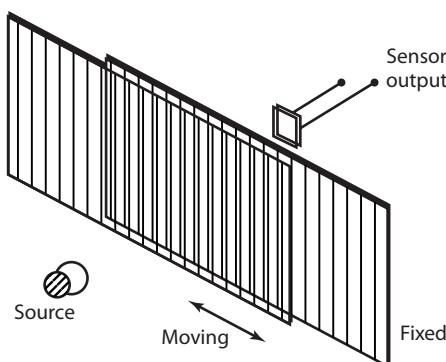
There are two types of encoders; *incremental* and *absolute*. An incremental encoder provides a simple pulse each time the object to be measured has moved a given distance. An absolute encoder provides a unique binary word coded to represent a given position of the object.

3.4.3 Incremental Encoders

Incremental encoders for angular measurement consist of a sensing shaft attached to a disk which is divided into an equal number of sectors on the circumference. In the linear type of encoders, there are equal segments along the length of travel. The readings are sensed by direct electrical contact with a brush or wiper or optically using optical slits or gratings. Since it counts the lines on a disk, the more lines, the higher the resolution. This specification is expressed as pulses per revolution, which is an important factor in encoder selection.

Incremental rotary encoders are very useful for measuring shaft rotation and primarily consist of three components: a light source, a coded wheel, and a photoelectric sensor. Figure 3-25 shows an encoder measuring system which uses transmission gratings. As the movable grating translates with respect to a fixed grating, the pulses are counted to provide position information.

FIGURE 3-25 GRATING TRANSDUCER PRINCIPLE



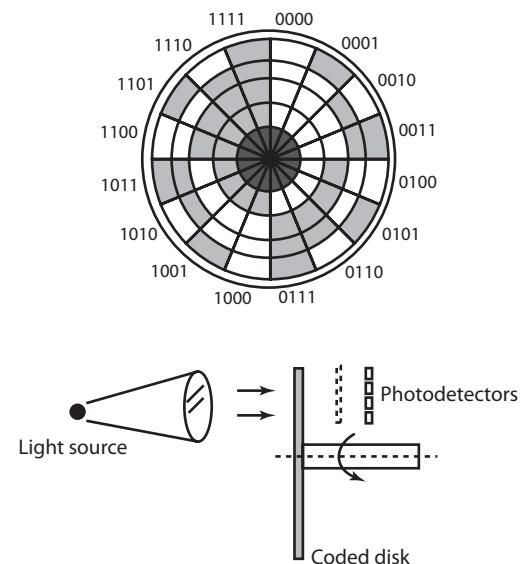
3.4.4 Absolute Encoders

The absolute encoder normally has a light source which emits a beam of light onto a photoelectric sensor called a photo detector. This converts the receiving light into an electrical signal, as shown in Figure 3-26. An optical encoded wheel (circular absolute grating) is mounted between the light source and photo detector. The encoded wheel has several concentric circular tracks that are divided into sectors. Manufactured into the surface of the coded wheel are alternating opaque and transparent sections. When the opaque section of the wheel passes in front of the light, the detector is *turned off*, and no signal is generated. When the transparent section of the wheel passes in front, the detector is *turned on*, and a signal is generated. The result is a series of signals corresponding to the rotation of the coded wheel. By using a counter to count these signals, it is possible to find out how far the wheel has rotated. Velocity information also can be obtained by differencing the pulses.

Incremental encoders are more commonly used than absolute encoders because of their simplicity and lower cost. Incremental encoders are used for both velocity and position measurement and are one of the most reliable and inexpensive devices available for this task.

3.4.5 Linear Encoder (Reflection Type)

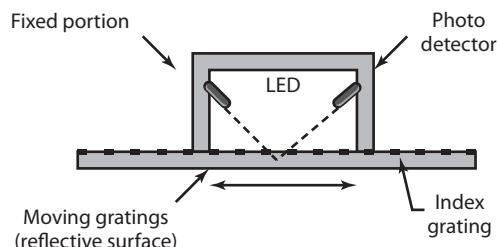
Optical gratings are used both in linear and radial forms, with the latter being rotated directly by the lead screw or a rack-and-pinion arrangement. Recent years have seen increasing use of steel or

FIGURE 3-26 OPTICAL ENCODING

steel-backed reflection scale grating, which for many engineering purposes is preferred to transmission gratings because of the increased durability and rigidity of steel gratings in comparison to optical gratings. In linear reflection-type encoders, the light must pass to the scale grating through the index grating and be reflected back through the index grating to the photoelectric sensor. Figure 3-27 shows a linear measuring system using reflection gratings. The fixed portion of the transducer box consists of a source of light, associated optics, and the detection system. The output of the detector is shown in the form of a digital read out. These types of transducers are popular in the machine-tool industry.

3.4.6 Moiré Fringe Transducers

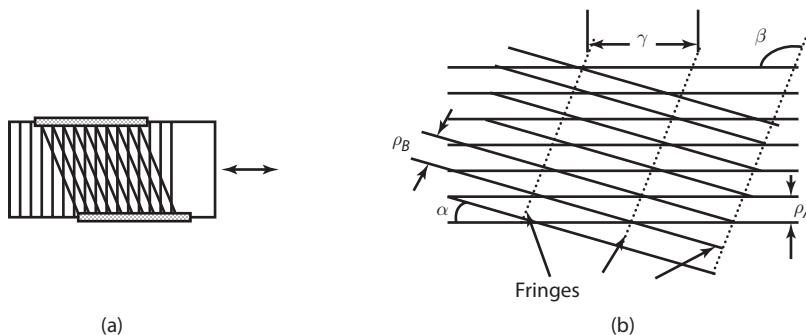
The moiré fringe principle is used in some types of digital transducers. These transducers also are used to measure length, angle, straightness, and circularity of motion. The transducer can supply information about the variable required and is relatively unaffected by external effects. An essential

FIGURE 3-27 LINEAR ENCODER (REFLECTION TYPE)

element of a transducer is an optical grating. An optical grating consists of regular succession of opaque lines separated by clear spaces of equal width. The lines are at right angles to the length of the grating. When two sections of such a grating are superimposed with the lines at slight angle to each other, a moiré fringe pattern with approximately a sinusoidal distribution of intensity results from the integrated interference effects of the interaction of the lines on each grating.

When one grating is moved with respect to the other at right angles to its lines, the moiré fringe pattern travels at right angles to the direction of movement. The sense of movement depends on the sense of relative travel of the gratings. This principle is shown in Figure 3-28.

FIGURE 3-28 (A) MOIRE FRINGES (B) FRINGE SEPARATION



Analysis of geometric relationships between the moiré fringes and the grating pair producing them leads to a finer comprehension of the potentialities of the moiré fringe measuring techniques.

$$v = \frac{\rho_A \rho_B}{[\rho_A^2 \sin^2 \alpha + (\rho_A \cos \alpha - \rho_B)^2]^{1/2}} \quad (3-39)$$

where

ρ_A, ρ_B = pitches of the gratings *A* and *B*, respectively

γ = fringe separation

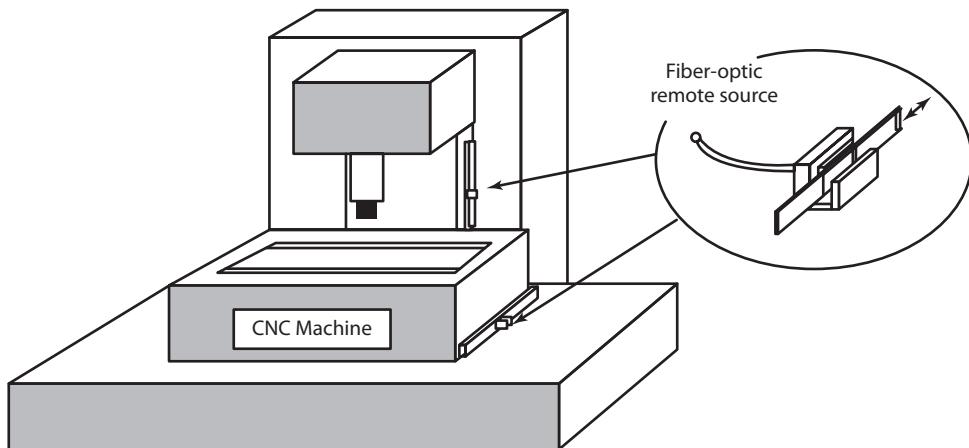
α = acute angle formed by the intersecting gratings

β = acute or obtuse angle between the lines of the first gratings and the fringe

3.4.7 Applications

Whenever encoders are used, they have to be calibrated for that specific situation. This is important because of the differing sizes, resolution requirements, and the specific nature of the movement. For example, Figure 3-29 shows encoders that are mounted to measure the displacements in two axial directions of a high-precision machine tool.

The distance to be traveled and the direction of travel are transmitted to the processor as reference values. This data gives reference values to the controller and the drive motor. If these values do not agree, the motor continues the rotation. Once they agree, the processor sends a stop signal to the controller, indicating the final slide position. If a new reference value is provided, the process is continued.

FIGURE 3-29 DIGITAL TRANSDUCERS FOR MACHINE TOOL MEASUREMENT

Absolute encoders are used in applications where the location of an object or identifying its position is of special interest. Unlike the incremental encoder, which determines position by counting pulses from the datum, the absolute encoder reads the system of coded tracks to establish the position. These encoders do not lose position when power is off. Each position is uniquely identified by a nonvolatile position verification device. Absolute encoders are chosen for situations, where establishing position status is desired as well as the possibility of avoiding equipment damage. This feature is useful in satellite tracking antennas, where occasional position verification is necessary, or in situations where an object is inactive for long periods of time or moves at very slow rates. Whenever the power is turned on, true position can be verified. Absolute encoders are not affected by stray signals from electrical noise and also can be used for serial data output for long-distance transmission. The absolute encoder is either a linear or an angular type. They may be single or multi-turn devices—the latter having higher accuracy and resolution.

Application in the Manufacturing Industry

- Machine slide position in numerically controlled machine tools
- Vertical and horizontal boring machines and precision lathes
- Gauging applications, such as in measuring calipers or digital height gauges
- As extensometers and measuring scales in structural research

The savings in indirect operator time using a digital measurement system often justifies the capital cost of transducer and display devices. Other advantages include further savings resulting from reduced scrap, operator fatigue, improved floor-to-floor time, and easier fitting.

Encoders in various configurations are possible with scaling in units of millimeters or inches, while the use of dual inch–metric capability is popular in the machine-tool industry. Angular encoders are calibrated to read degrees, minutes, seconds of arc, or (alternatively) decimal fractions of the degree. It is common to attach optical shaft encoders to the lead screw of the machine tool to digitize the screw position. The use of linear encoders eliminates the error caused by backlash in the lead screw and other mechanical transmission systems.

For applications requiring high resolution, the size of the transparent and opaque sections must be made very small, and the light source must be properly aligned in order for the photo detector to sense a change in light. Multiplication techniques can be used to increase the resolution. Four times magnification is commonly achieved by externally counting the rising and falling edges of each channel. For example, a 5,000 ppr quadrature encoder can generate 20,000 ppr using this technique.

SUMMARY

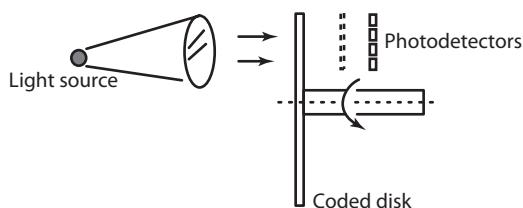
Rotary Encoder

Encoders have both linear and rotary configurations. Rotary encoders are available in two forms.

1. Incremental encoders produce digital pulses as shaft rotates, allowing relative displacement of shaft to be measured.
2. Absolute encoders have a unique digital word corresponding to each rotational position of the shaft.

Incremental encoders (Figure 3-30) are useful for measuring shaft rotation and consist of primarily three components: a light source, a coded wheel, and a photoelectric sensor. An incremental encoder provides a simple pulse each time the object to be measured has moved a given distance.

FIGURE 3-30



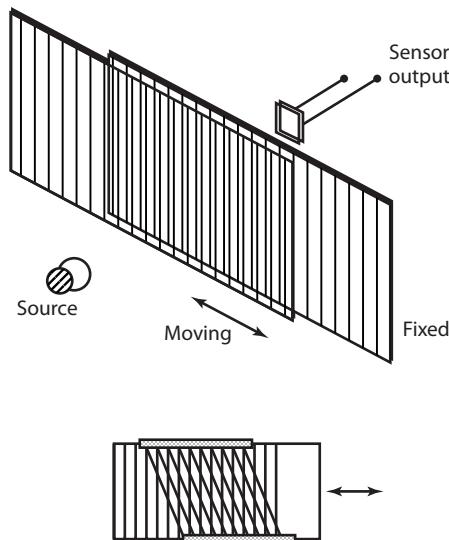
Typical absolute encoders have a coded wheel mounted between the light source and photo detector. Manufactured into the surface of the coded wheel are alternating opaque and transparent sections in a digital pattern. This results in a series of signals corresponding to the rotation of the coded wheel. By using a counter to count these signals, it is possible to find the wheel rotation. Velocity information also can be obtained by differencing the pulses.

Moiré Fringe Transducer

When two sections of optical gratings are superimposed with the lines at slight angle to each other, a moiré fringe pattern is generated (Figure 3-31). The interference effect of the lines provides a sinusoidal distribution of intensity. When one grating is moved with respect to the other at right angles to its lines, the moiré fringe pattern travels at right angles to the direction of movement; the sense of movement depends on the sense of relative travel of the gratings.

Applications

- Encoders are used for measurement of linear or angular position, velocity, and direction of movement.
- Used in computerized manufacturing machines, motion-control applications, and quality assurance of equipment.
- Used in tensile-test instruments to precisely measure the ball screw position.

FIGURE 3-31

- Used in automated test stands used when angular positions of windshield wiper drives and switch positions are tested.
- Incremental encoders commonly are used for counting applications.
- The moiré fringe transducers also are used to measure length, angle, straightness, and circularity of motion.

3.5 Force, Torque, and Tactile Sensors

Mechatronic systems in automated manufacturing environments require extensive environmental information to make intelligent decisions. Such information relates to the tasks of material handling, machining, inspection, assembly, painting, etc. Assembly tasks and automated handling tasks require controlled operations like grasping, turning, inserting, aligning, orienting, and screwing. Every situation has somewhat different sensing requirements.

This section discusses some of the techniques used for force and torque sensing. A precise measurement of strain is an important consideration in measurement. Strain measurement is used as a secondary step in the measurement of many process variables, including flow, pressure, weight, and acceleration. Electrical-resistance strain gauges are widely used to measure strains due to force or torque. When a force is applied to a structure, it undergoes deformation. The gauge, which is bonded to the structure, is deformed by strain, and its electrical-resistance changes in a nearly linear fashion.

If a piece of metal wire is stretched, not only does it get longer and thinner, but its resistance increases. The greater the strain experienced by the wire, the greater is the change in resistance.

There are a number of ways in which resistance can be changed by a physical phenomenon. The resistance, R , of a metal depends on its area, length, and electrical resistivity. It is possible to express the resistance of a conductor at a constant temperature, T , as

$$R_0 = \frac{\rho l}{A_0}$$

where

ρ = resistivity, $\Omega\text{-m}$

R_o = sample resistance, Ω

l = length, m

A_o = cross-sectional area, m^2

3.5.1 Sensitivity of Resistive Transducers

If a specimen is subjected to tension, causing an increase in length, its longitudinal dimension will increase, and its lateral dimension will decrease. If a resistance gauge made of this conducting material is subjected to a positive strain, its length increases while its cross-sectional area decreases. Since the resistance of the conductor is dependent on its length, cross-sectional area, and specific resistivity, the change in strain is due to the change in dimension or specific resistivity.

For a circular wire of length, L ; cross-sectional area, A ; and diameter, D , the resistance of the wire before straining is

$$R = \frac{\rho L}{A} \quad (3-40)$$

Let us subject the wire to tension which causes the strain. Tension increases length and reduces the diameter, which in turn reduces the area of cross section. Let the stress applied to the strain gauge be s in N/m^2 . Additional definitions are

ΔL = change in length of wire

ΔA = change in area of cross-section

ΔD = change in diameter

ρ = resistivity

ν = Poisson's ratio

$$\text{Strain } \varepsilon = \frac{\Delta L}{L}$$

In order to find how ΔR depends on the material physical quantities, Equation 3-40 is differentiated with respect to applied stress s .

$$\frac{dR}{dS} = \frac{\rho}{A} \frac{\partial L}{\partial S} - \frac{\rho L}{A^2} \frac{\partial A}{\partial S} + \frac{L}{A} \frac{\partial \rho}{\partial S} \quad (3-41)$$

Dividing Equation 3-41 throughout by Equation 3-40 yields

$$\frac{1}{R} \frac{dR}{dS} = \frac{1}{L} \frac{\partial L}{\partial S} - \frac{1}{A} \frac{\partial A}{\partial S} + \frac{1}{\rho} \frac{\partial \rho}{\partial S} \quad (3-42)$$

The change in resistance is due to two items:

1. Unit change in length $\Delta L/L$
2. Unit change in area $\Delta A/A$

Since the area $A = \frac{\pi D^2}{4}$, we can write

$$\frac{\partial A}{\partial S} = 2\frac{\pi}{4}D\frac{\partial D}{\partial S} \quad (3-43)$$

and

$$\frac{1}{A}\frac{dA}{dS} = \frac{\frac{2\pi}{4}D}{\frac{\pi}{4}D^2}\frac{\partial D}{\partial S} = \frac{2}{D}\frac{\partial D}{\partial S} \quad (3-44)$$

Equation 3-44 can be written as

$$\frac{1}{R}\frac{dR}{dS} = \frac{1}{L}\frac{\partial L}{\partial S} - \frac{2}{D}\frac{\partial D}{\partial S} + \frac{1}{\rho}\frac{\partial \rho}{\partial S} \quad (3-45)$$

Poisson's ratio is defined as

$$\nu = -\frac{\text{lateral strain}}{\text{longitudinal strain}} = -\frac{\frac{\partial D}{D}}{\frac{\partial L}{L}} \quad (3-46)$$

$$\frac{\partial D}{D} = -\nu \frac{\partial L}{L}$$

$$\frac{1}{R}\frac{\partial R}{\partial S} = \frac{1}{L}\frac{\partial L}{\partial S} + \nu \frac{2}{L}\frac{\partial L}{\partial S} + \frac{1}{\rho}\frac{\partial \rho}{\partial S} \quad (3-47)$$

For small variations, the relationship in these equations can be written as

$$\frac{\Delta R}{R} = \frac{\Delta L}{L} + 2\nu \frac{\Delta L}{L} + \frac{\Delta \rho}{\rho} \quad (3-48)$$

Sensitivity or gauge factor, G_f , is defined as the ratio of unit change in resistance to unit change in length:

$$G_f = \frac{\frac{\Delta R}{R}}{\frac{\Delta L}{L}}$$

and

$$\frac{\Delta R}{R} = G_f \frac{\Delta L}{L} = G_f \varepsilon \quad (3-49)$$

Gauge factor also can be expressed as

$$G_f = \frac{\frac{\Delta R}{R}}{\frac{\Delta L}{L}} = 1 + 2\nu + \frac{\frac{\Delta \rho}{\rho}}{\frac{\Delta L}{L}} = 1 + 2\nu + \frac{\frac{\Delta \rho}{\rho}}{\frac{\varepsilon}{\varepsilon}} \quad (3-50)$$

The change in resistivity occurs because of the piezoresistive effect, which is explained as an electrical resistance change which occurs when the material is mechanically deformed. In some cases, the effect is a source of error. If the change in resistivity or piezoresistive effect of the material is neglected, the gauge factor becomes

$$G_f = 1 + 2\nu \quad (3-51)$$

The gauge factor gives an idea of the strain sensitivity of the gauge in terms of the change in resistance per unit strain. Although strain is a unitless quantity, it is a common practice to express strain as a ratio of two units as m/m. Poisson's ratio for all metals is between 0 and 0.5. The gauge factor for metal can vary from 2 to 6. For semiconductors, it can vary between 40 to 200. Some common materials and their gauge factors are listed in Table 3-2.

TABLE 3-2

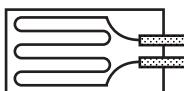
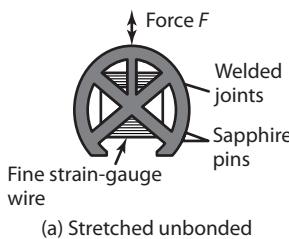
Material	Gauge Factor
Nickel	-12.6
Manganese	+0.07
Nicrome	+2.0
Constantan	+2.1
Soft Iron	+4.2
Carbon	+20
Platinum	+4.8

The gauge factor is normally supplied by the manufacturer from a calibration made of a number of gauges from a sample batch. The gauge factor for various metals ranges from -12 for nickel to +4 for soft iron. This indicates that changes in resistivity of a material could be quite significant while measurements are made.

3.5.2 Strain Gauges

A resistance strain gauge consists of a grid of fine resistance wire of about $20 \mu\text{m}$ in diameter. The elements are formed on a backing film of electrically insulating material. Current strain gauges are manufactured from constantan foil, a copper-nickel alloy, or single-crystal semiconductor materials. The gauges are formed either mechanically or by photochemical etching. Strain-gauge transducers are of two types: unbonded and bonded.

Unbonded Strain Gauges In an unbonded strain gauge (Figure 3-32(a)), the resistance wire is stressed between the two frames. The first frame is called the fixed frame, and the second is called

FIGURE 3-32 STRAIN GAUGES

a moving frame. The wires in the unbonded gauges are connected such that the input motion of one frame stretches one set of wires and compresses another set of wires.

As an example, a $20\ \mu\text{m}$ diameter wire is wound between insulated pins with one attached to a stationary frame and the other to a movable frame. For a particular stress input, the winding experiences either an increase or decrease in stress, resulting in a change in resistance. The output is connected to a Wheatstone bridge for measurement. With this type of strain gauge, measurement of small motions as small as a few microns can be made.

Bonded Strain Gauge Bonded strain-gauge transducers are widely used for measuring strain, force, torque, pressure, and vibration. The gauges have a backing material. Bonded strain gauges (Figure 3-32(b)) are made of metallic or semiconductor materials in the form of a wire gauge or thin metal foil. When the gauges are bonded to the surface, they undergo the same strain as that of the member surface. The coefficient of thermal expansion of the backing material should be matched to that of the wire.

Strain gauges are sensitive devices and are used with an electronic measuring unit. The strain gauge is normally made part of a Wheatstone bridge, so the change in its resistance due to strain either can be measured or used to produce an output, which can be displayed. Strains as low as a fraction of a micron can be measured using strain gauges. Table 3-3 presents characteristics of bonded strain gauges.

For precise measurement, the strain gauges should have the following properties.

- A high gauge factor increases the sensitivity and causes a larger change in resistance for a particular strain.
- The gauge characteristics are chosen so that the variation in resistance is a linear function of strain. If the gauges are used for dynamic measurements, the linearity should be maintained over the desired frequency range. High resistance of the strain gauge minimizes the effect of resistance variation in the signal-processing circuitry.
- Strain gauges have a low temperature coefficient and absence of the hysteresis effect.

TABLE 3-3 BONDED STRAIN GAUGES

Material	Gauge Factor	Resistance Ω	Resistance–Temperature Coefficient $\Omega/\Omega/^\circ\text{C}$	Comments
Nichrome, Ni:80%,Cr:20%	2.5	—	0.1×10^{-3}	For use under 1200°C
Constantan, Ni:45%,Cu:55%	2.1	100	$\pm 0.02 \times 10^{-3}$	400°C
Platinum	4.8	50	4.0×10^{-3}	For high temperature use
Silicon	−100 to +150	200	—	—
Nickel	−12	—	4.8×10^{-3}	—

EXAMPLE 3.7

A compressive force is applied to a structure causing the strain, $\varepsilon = -5(10)^{-6}$. Two separate strain gauges are attached to the structures, where one is a nickel wire strain gauge of gauge factor of −12.1 and another is a nichrome wire strain gauge of gauge factor of 2. Calculate the value of resistance of the gauges after they are strained. The resistance of strain gauge is 120 Ω.

Solution

Let us consider tensile strain as positive and compressive strain as negative.

$$\text{Strain, } \varepsilon = -5(10)^{-6}$$

$$\frac{\Delta R}{R} = G_f \cdot \varepsilon; \Delta R = RG_f \cdot \varepsilon$$

Change in resistance for Nickel strain gauge,

$$\Delta R = (120)(-12.1)(-5)(10)^{-6} = 7.26(10)^{-3}\Omega$$

Change in resistance for Nichrome strain gauge,

$$\Delta R = (120)(2)(-5)(10)^{-6} = -1.2(10)^{-3}\Omega$$

The value of resistance of nickel strain gauge increases, whereas, the value of resistance of nichrome strain gauge decreases.

EXAMPLE 3.8

A resistance wire strain gauge with a gauge factor of 2 is bonded to a steel structure member subjected to a stress of 100 MN/m². The modulus of elasticity of steel is 200 GN/m². Calculate the percentage change in value of the gauge resistance due to the applied stress.

Solution

$$\text{Strain} = \varepsilon = \frac{S}{E} = \frac{100(10)^6}{200(10)^9} = 0.5(10)^{-3} \text{ m/m}$$

$$\text{Gauge factor} = \frac{\frac{\Delta R}{R}}{\varepsilon}$$

Thus,

$$\frac{\Delta R}{R} = G_f \cdot \varepsilon = (2)(0.5)(10)^{-3} = 0.001$$

$$\text{Percentage change in } \frac{\Delta R}{R} = 0.1\%$$

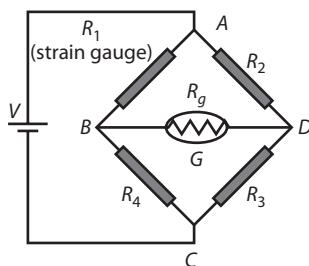
Bridge Circuit Arrangement The Wheatstone bridge circuit is used to measure the small changes in resistance that result in most strain-gauge applications. The change in resistance either can be measured or provided as an output that is processed by the computer. Figure 3-33 shows an arrangement of a bridge circuit. In the balanced bridge arrangement, strain-gauge resistance, R_1 , forms one arm of the Wheatstone bridge, while the remaining arms have resistances R_2 , R_3 , and R_4 . Between the points A and C of the bridge, there is a power supply; between points B and D, there is a precision galvanometer. The galvanometer gives an indication of the presence of current through that leg. For zero current to flow through the galvanometer, the points B and D must be at the same potential. The bridge is excited by the direct current source with voltage, V and R_g is the resistance in the galvanometer. The bridge is said to be balanced when there is no current flowing through the galvanometer.

The condition of balance is

$$\frac{R_1}{R_4} = \frac{R_2}{R_3} \quad (3-52)$$

If R_1 changes due to strain, the bridge (which is initially in the balanced condition) becomes unbalanced. This may be balanced by changing R_4 or R_2 . The change can be measured and used to indicate the change in R_1 . This procedure is useful for measuring static strains.

FIGURE 3-33 BRIDGE CIRCUIT WITH STRAIN GAUGE



In the unbalanced bridge arrangement, the current through the galvanometer or the voltage drop across it is used to indicate the strain. This is useful for measuring dynamic as well as static strains.

3.5.3 Offset Voltage

As shown in Figure 3-33, G is a null deflector that is used to compare potentials of point- B and D . The potential difference between points B and D is $\Delta V = V_D - V_B$. If all the resistance values (R_1, R_2, R_3, R_4) chosen in the bridge circuit are same, then the voltage at points B and D are the same, ΔV will be zero, and the bridge is balanced.

Let us consider R_1 as the strain gauge. If R_1 is strained, its resistance value changes, and the bridge becomes unbalanced, causing a nonzero ΔV . If any other resistance value is adjusted, the bridge can be brought back to a balanced condition. The adjusted value of any resistor needed to force ΔV to zero is equal to the strained value of the strain gauge. The current flowing through the bridge arms is computed as

$$\text{Current through } ABC: I_1 = \frac{V}{R_1 + R_4}$$

$$\text{Current through } ADC: I_2 = \frac{V}{R_2 + R_3}$$

The voltage drop across $R_3 = (I_2)R_3$, and the voltage drop across $R_4 = (I_1)R_4$. The voltage offset is given by

$$\Delta V = V_D - V_B = \frac{R_3 V}{R_2 + R_3} - \frac{R_4 V}{R_1 + R_4}$$

$$\Delta V = V \cdot \frac{R_3 R_1 - R_4 R_2}{(R_2 + R_3)(R_1 + R_4)}$$

In data acquisition systems where the ratio $\frac{\Delta R}{R}$ is small, the following method is suitable.

Constant supply voltage to the bridge is V , and ΔV is the output voltage.

$$\Delta V = \frac{R_3 R_1 - R_4 R_2}{(R_2 + R_3)(R_1 + R_4)} \cdot V$$

Use $R_1 = R + \Delta R$ and R_2, R_3, R_4 equal to R . Therefore,

$$\Delta V = \left(\frac{R(R + \Delta R) - R^2}{(R + R)(R + \Delta R + R)} \right) V$$

$$\Delta V = \frac{\Delta R}{4R + 2\Delta R} V$$

If

$$\frac{\Delta R}{R} = \delta; \quad \Delta V = \frac{\delta}{4 + 2\delta} V$$

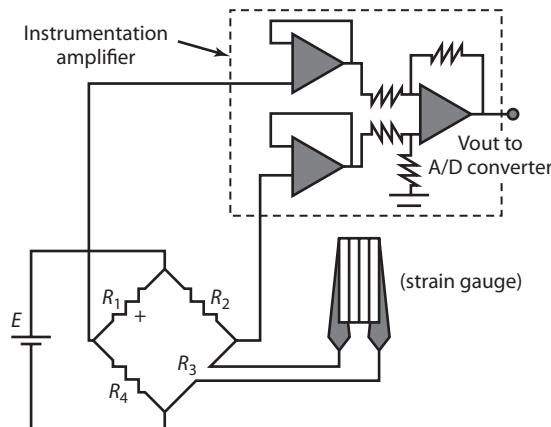
In systems, where the value of δ is small,

$$\Delta V = \frac{\delta}{4} V \text{ or } \delta = \frac{4\Delta V}{V} \text{ or } \Delta R = \frac{4R\Delta V}{V}$$

Signal Enhancement Strain-gauge devices with signal-conditioning equipment are designed to balance the bridge automatically and provide the strain value in terms of microstrains. Data acquisition systems for force and strain measurement are programmed to provide the unbalanced offset voltage, which is proportional to the gauge resistance.

Figure 3-34 shows an arrangement of an instrumentation amplifier to be connected to the input channels of the data acquisition system.

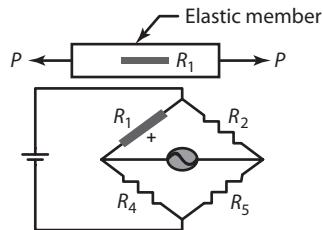
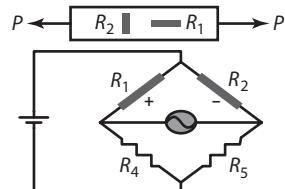
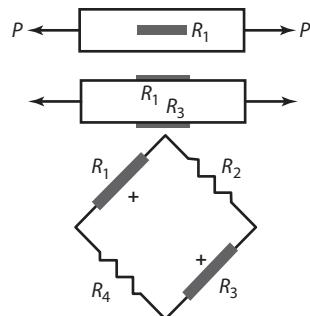
FIGURE 3-34 BRIDGE CIRCUIT WITH INSTRUMENTATION AMPLIFIER



Possible Strain-Gauge Arrangement When more than one arm of the bridge circuit contains strain transducers and their resistances change, the bridge output is due to the combined effect of these changes. More than one strain gauge, if suitably arranged, can lead to a higher signal-enhancement factor and a larger change in output voltage for a given strain.

For example, in Figure 3-33, R_3 is the original strain gauge, and if we use R_1 as another strain gauge placed in a location such that it has same strain as R_3 the bridge output will be double the value obtained for a single gauge. In many experimental situations, there are areas of tension and compression in the same object with similar strain but of opposite sign. In such situations, care must be taken in arranging strain gauges in such a way that the adjacent arms of the bridge have strains of opposite nature.

In Figure 3-35, R_1 measures changes due to *axial tensile strain*. In Figure 3-36, strain gauge R_1 is bonded to the elastic member to measure axial tensile strain. R_1 changes due to axial tensile strain. R_2 measures changes due to *transverse compressive strain*. In the arrangement shown in Figure 3-37, both R_1 and R_3 are subjected to axial tensile strain of the same amount, and R_1 and R_3 form opposite arms of the bridge. This causes a signal enhancement factor of 2.

FIGURE 3-35 POSSIBLE ARRANGEMENT STRAIN GAUGES TO MEASURE P **FIGURE 3-36 POSSIBLE ARRANGEMENT OF GAUGES TO MEASURE P** **FIGURE 3-37 POSSIBLE ARRANGEMENT OF GAUGES TO MEASURE TENSION**

In the example shown in Figure 3-38, R_1 has tensile strain, and R_2 has compressive strain. R_3 also has tensile strain, and R_4 has compressive strain. Strain gauges R_1 , R_2 , R_3 and R_4 are bonded at the root of the cantilevers, where the bending stresses are maximum. In the arrangement shown in

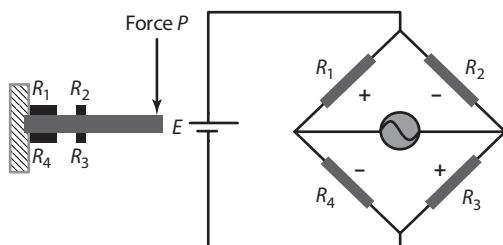
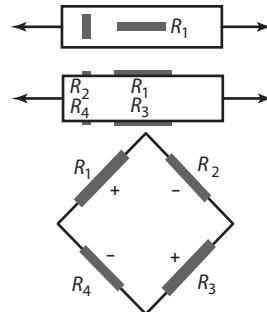
FIGURE 3-38 CANTILEVER DEFLECTION MEASUREMENT

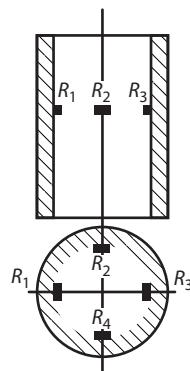
Figure 3-39, four active gauges are used with R_2 and R_4 arranged at right angles to R_1 and R_3 to produce a signal enhancement factor of $2(1 + \nu)$, where ν denotes Poisson's ratio.

FIGURE 3-39 ALTERNATE ARRANGEMENTS



In the arrangement shown in Figure 3-40, the strain gauges are arranged in such a way that R_1 and R_3 measure axial strains, while R_2 and R_4 measure the *circumferential strains*, which have strain of the opposite nature.

FIGURE 3-40 HOLLOW CYLINDER WITH AXIAL LOADING



Temperature Effects in Strain Gauges The strain-gauge measuring environment is often influenced by temperature changes. The electrical resistivity of most alloys changes with temperature, increasing as temperature rises and decreasing as it falls. As shown in Table 3-3, metals used in strain gauges have a temperature coefficient (α_0) of the order of $0.004/\text{ }^\circ\text{C}$. The resistance at temperature T is given as

$$R_T = R_{T_0}(1 + \alpha_0 \Delta T) \quad (3-53)$$

Resistance change due to change in temperature ΔT is

$$\Delta R_T = R_{T_0} \alpha_0 \Delta T \quad (3-54)$$

For example, if the temperature changes by one degree, the change in resistance is calculated as $\Delta T = 1^\circ$, $\alpha_0 = 0.004/\text{ }^\circ\text{C}$, $R_{T_0} = 120 \Omega$, and $\Delta R_T = 0.48 \Omega$.

When a strain gauge is bonded to the member being tested, its resistance will be affected by a change in temperature. This effect is independent of any strain applied to the gauge. The recording instrument cannot differentiate between the changes in the resistance due to temperature and strain. In addition, unless the coefficient of the linear expansion of the gauge is the same as that of the material to which it is bonded, the temperature change during measurement also will be a source of false strain due to differential expansion.

Temperature Compensation Temperature compensation is achieved in two manners:

1. Using a dummy gauge.
2. Using more than one active gauge with proper arrangement of gauges.

If active and dummy gauges are mounted on the adjacent arms of a bridge, variation in temperature will not affect the bridge. The active gauge is subjected to strain as well as temperature change, while the dummy gauge is subjected to temperature change only. Since active and dummy gauges form adjacent arms of the bridge, the output due to temperature change is zero, as both active and dummy gauges change identically due to temperature. Furthermore, it is desirable to choose a gauge material with a coefficient of thermal expansion very close to that of the material under test.

Since it is inconvenient to calculate and apply temperature correction after the measurement is made, the temperature compensation can be made in the experimental setup itself. The gauges are suitably arranged so that adjacent arms have strains of opposite nature. This procedure ensures signal enhancement as well as temperature compensation.

Acceleration Sensing Using Strain Gauges Strain gauges are used in a variety of electrical transducer devices. Their advantages include ease of instrumentation, high accuracy, and excellent reliability. One of the most common configurations used in pressure, force, displacement, and acceleration transducers is the cantilever configuration with strain gauges mounted at the base, shown in Figure 3-41. A point mass of weight W is used as the acceleration-sensing element, and the cantilever (mounted with gauges) converts the inertial force into a strain.

FIGURE 3-41 ACCELERATION SENSING

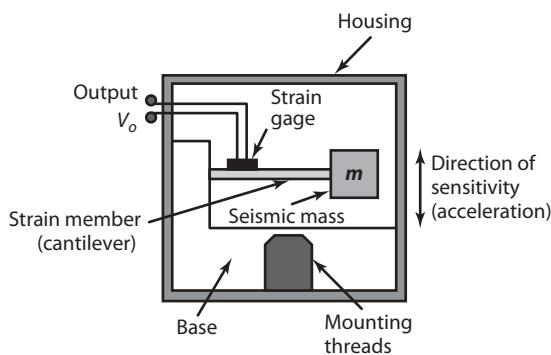
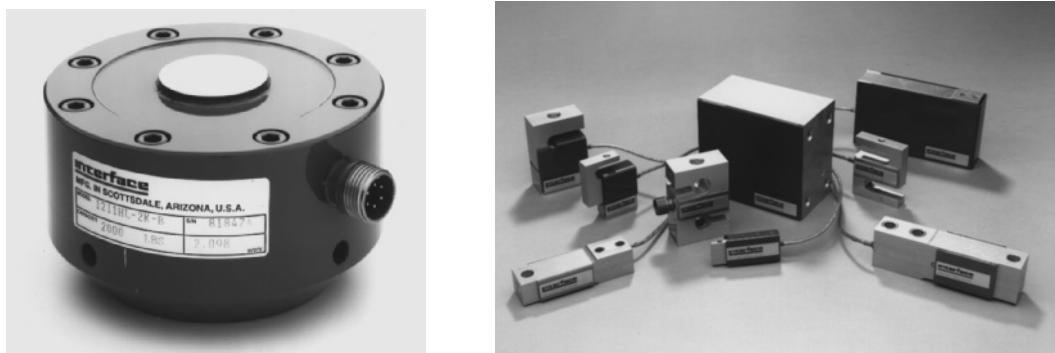


Figure 3-42 presents a photograph of a load cell being used in a force measurement application.

FIGURE 3-42 LOAD CELL



Courtesy of Interface, Scottsdale, AZ.

Semiconductor Strain Gauges Semiconductor strain gauges are very useful in low strain applications. Use of semiconductor silicon has notably increased during the last few years. In a semiconductor gauge, the resistivity changes with strain as well as with physical dimensions. Changes in electron and hole mobility with changes in crystal structure as strain is applied results in larger gauge factors than possible with the metal gauges. Gauge factors of semiconductor gauges are between 50 and 200.

Semiconductor strain gauges physically appear as a band or strip of material with an electrical connection. The gauge is either bonded directly to the test element, or if encapsulated, it is attached by the encapsulation material. Signal conditioning is essentially a bridge circuit with temperature compensation.

There is also a need to linearize the output, because the basic characteristic of resistance versus strain is nonlinear. For good linearity of the output voltage with respect to strain, it is desirable to maintain a constant gauge current. This is accomplished by maintaining constant voltage excitation or by suitable modification, which produces constant current in the bridge arm in addition to constant voltage.

The benefits of semiconductor strain gauges are low power consumption and low heat generation. In addition, the mechanical hysteresis is negligible.

$$\frac{\delta R}{R} = G_f \epsilon + G_f \epsilon^2 \quad (3-55)$$

The resistance of semiconductor strain gauges varies from 1000 to 5000 Ω . They are usually made from *p*- or *n*-type silicon material.

3.5.4 Tactile Sensors

Tactile sensors are used in many applications ranging from fruit picking to monitoring human prosthetic implants; however, the major area of application is in the biomedical field. Tactile sensors are used for the following.

- Study the forces developed by the human foot during motion.
- Study the forces developed during various types of hand functions.
- Monitor the artificial knee and sense the forces developed.

Other areas of application include the field of robotics, where tactile sensors can be placed on the gripper of the manipulator to provide feedback information from the workpiece. Besides being used as a touch sensor, gripping force sensors detect the force with which the object is gripped, pressure sensors detect the pressure applied to the object, and slip sensors can detect if the object is slipping. In addition, other industrial applications of the tactile sensors include the study of forces developed by fastening devices.

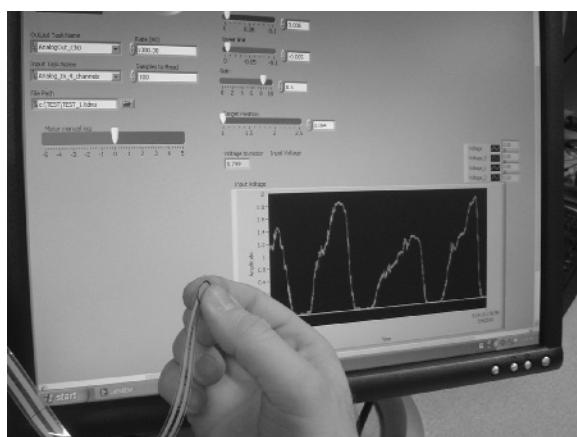
A tactile sensing system has the ability to detect the following.

1. Presence of a part
2. Part shape, location, and orientation
3. Contact pressure distribution
4. Force magnitude and direction

The major components of tactile sensors include:

- Touch surface
- Transducer
- Structure and control interface

FIGURE 3-43 PHOTOGRAH OF THE TACTILE SENSOR



Shetty, University of Hartford.

Some tactile sensors are designed using piezoelectric films. Piezoelectric (Piezo) film consists of polyvinylidene fluoride (PVDF) that has undergone special processing to enhance its piezoelectric properties. Piezo film develops an electrical charge proportional to induced mechanical stress or strain. As a result, it produces a response proportional to the rate of stress rather than to the stress magnitude. This sensor is passive—that is, its output signal is generated by the piezoelectric film without the need for an excitation signal. The piezoelectric tactile sensor can be fabricated with the PVDF film strips imbedded into a rubber skin. To measure surface vibration, the film is bonded to the surface. As the surface vibrates, it stretches the surface in a cyclical manner, generating a voltage. Piezo-film voltage output is relatively high.

A resistive tactile sensor known as a force sensing resistor (FSR) can be fabricated using material whose electrical conductivity changes with strain. FSR consists of a material whose resistance changes

with applied pressure. Such materials are known as conductive elastomers fabricated of silicone rubber, polyurethane, and other compounds. The basic operating principle of elastomeric tactile sensors is based either on varying the contact area when the elastomer is squeezed between two conductive plates or on changing the thickness. When the external force varies the contact area at the interface of the elastomer, changes result in a reduction of electrical resistance. Compared with a strain gauge, the FSR has a much wider dynamic range. Miniature tactile sensors are used extensively in robotic applications where good spatial resolution, high sensitivity, and wide dynamic range are required.

SUMMARY

Strain Gauges

The resistance, R , of a resistance wire depends on its area, length, and electrical resistivity.

$$R_0 = \frac{\rho l}{A_0}$$

where

ρ = resistivity, $\Omega\text{-m}$

R_0 = sample resistance, Ω

l = length, m

A_0 = cross-sectional area, m^2

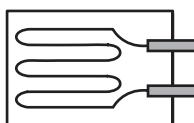
Sensitivity or gauge factor, G_f , is defined as the ratio of unit change in resistance to unit change in length.

$$G_f = \frac{\frac{\Delta R}{R}}{\frac{\Delta L}{L}}$$

Bonded Strain Gauges

Bonded strain gauges (Figure 3-44) are made of metallic or semiconductor materials in the form of a wire gauge or thin metal foil. When the gauges are bonded to the surface they undergo the same strain as that of the member surface.

FIGURE 3-44 BONDED WIRE STRAIN GAUGE

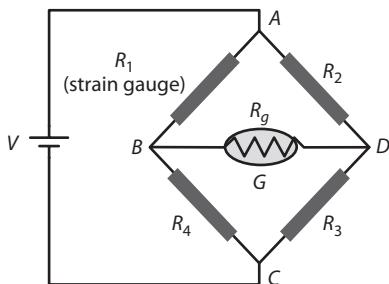


Strain gauges are very sensitive devices and are used with an electronic measuring unit. The resistance strain gauge is normally made part of a Wheatstone bridge (Figure 3-45) so that the change in its resistance due to strain can either be measured or used to produce an output which can be displayed or recorded.

Features

Strain gauges should have the following features:

- A high gauge factor increases its sensitivity and causes a larger change in resistance for a particular strain.
- High resistance of the strain gauge minimizes the effect of resistance variation in the signal processing circuitry. Choose gauge characteristics such that resistance is a linear function of strain.

FIGURE 3-45 BRIDGE CIRCUIT ARRANGEMENT

- For dynamic measurements, the linearity should be maintained over the desired frequency range.
- Low temperature coefficient and absence of the hysteresis effect add to the precision.

Applications

- Strain-gauge transducers are used for measuring strain, force, torque, pressure, and vibration.
- In some applications, strain gauges are used as a primary or secondary sensor in combination with other sensors.

Tactile Sensors

- Tactile sensors are used in applications ranging from fruit picking to monitoring human prosthetic implants.
- Biomedical applications include the study of forces during human foot motion, during various types of hand functions, and monitoring and sensing the forces developed in knee implants
- In robotics, the tactile sensors are placed on the gripper of the manipulator to provide feedback; pressure sensors detect the pressure applied to the object, and slip sensors can detect slip

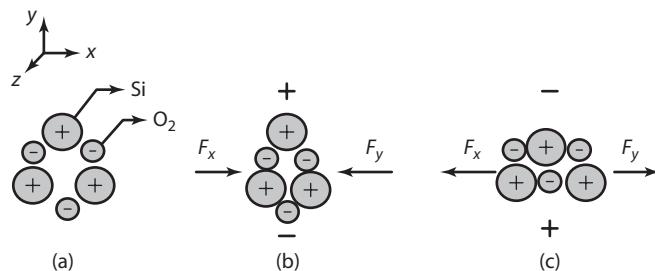
3.6 Vibration—Acceleration Sensors

3.6.1 Piezoelectric Transducers

Piezoelectric transducers depend upon the characteristics of certain materials that are capable of generating electric voltage when they deform. Piezoelectric materials, when subjected to mechanical force or stress along specific planes, generate electric charge. The property of generating an electric charge when deformed makes piezoelectric materials useful as primary sensors in instrumentation.

The best-known natural material is quartz crystal (SiO_2). Rochelle salt is also considered a natural piezoelectric material. Artificial materials using ceramics and polymers, such as PZT (lead zirconium titanate), PVDF (polyvinylidene fluoride), BaTiO_3 (barium titanate), and LS (Lithium Sulfate) also exhibit the piezoelectric phenomenon.

Piezoelectric Effect A piezoelectric material such as a quartz crystal can be cut along its axes in x , y , and z directions. Figure 3-46 shows a view along the z -axis. In a single-crystal cell, there are three atoms of silicon and six atoms of oxygen. Oxygen atoms are lumped in pairs. Each silicon

FIGURE 3-46 PIEZOELECTRIC EFFECT IN A CRYSTAL

atom carries four positive charges, and oxygen atoms carry two negative charges. A pair of oxygen atoms carries four negative charges. When there is no external force applied on the quartz crystal, the quartz cell is electrically neutral.

When compressive forces are applied along the x -axis, as shown in Figure 3-46(b), the hexagonal lattices become deformed. The forces shift the atoms in the crystal in such a manner that positive charges are accumulated at the silicon atom side and negative charges at the oxygen pair side. The crystal tends to exhibit electric charges along the y -axis. On the other hand, if the crystal is subjected to tension along the x -axis, as in Figure 3-46(c), a charge of opposite polarity is produced along the y -axis. To transmit the charge which has been developed, conductive electrodes are applied to the crystal at the opposite side of the cut.

The piezoelectric material acts as a capacitor with the piezoelectric crystal acting as the dielectric medium. The charge is stored because of the inherent capacitance of the piezoelectric material itself.

The piezoelectric effect is reversible. If a varying potential is applied to the proper axis of the crystal, it changes the dimension of the crystal, thereby deforming it. A piezoelectric element used for converting mechanical motion to electrical signals is thought of as both a charge generator and a capacitor. This charge appears as a voltage across the electrodes. The magnitude and polarity of the induced surface charges are proportional to the magnitude and direction of the applied force. For the arrangements shown in Figure 3-47 and 3-48, the charge generated, Q , is defined as:

$$Q = dF \text{ (Longitudinal effect)} \quad (3-56)$$

$$Q = dF \frac{a}{b} \text{ (Transverse effect)}$$

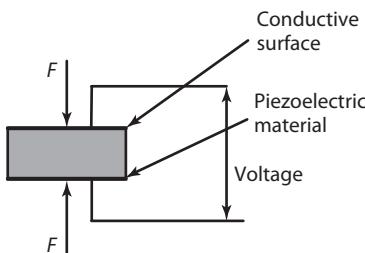
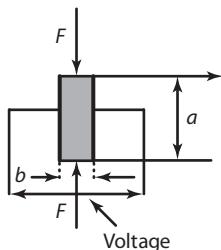
FIGURE 3-47 LONGITUDINAL EFFECT

FIGURE 3-48 TRANSVERSE EFFECT

Here d is the *piezoelectric coefficient* of the material. It is also known as *the charge sensitivity factor* of the crystal. For a typical quartz crystal, $d = 2.3 \times 10^{-12} \text{ F/N}$ or 2.3 pF/N where F is the applied force, in newtons.

If the ratio of $\frac{a}{b}$ is greater than one, the transverse effect produces more charge than the longitudinal effect. The force, F , results in a change in thickness of the crystal. If the original thickness of the crystal is t and Δt is the change in thickness due to the applied force, Young's modulus, E , can be expressed as the ratio of stress and strain:

$$\text{Young's Modulus: } E = \frac{\text{Stress}}{\text{Strain}} = \frac{\frac{F}{A}}{\frac{\Delta t}{t}} = \frac{Ft}{A\Delta t}$$

Rewriting the expression, we have

$$F = \frac{AE}{t} \Delta t \quad (3-57)$$

where

A = area of the crystal, m^2

t = thickness of the crystal, m

Piezoelectric Output From Equations 3-56 and 3-57, we have

$$\text{Charge } (Q) = \frac{dAE\Delta t}{t} C \quad (3-58)$$

The charge at the electrodes produces the voltage

$$V = \frac{Q}{C} \quad (3-59)$$

The capacitance of the piezoelectric material between the two electrodes is

$$C = \epsilon \frac{A}{t} = \epsilon_0 \epsilon_r \frac{A}{t} \quad (3-60)$$

Here ϵ_r is the dielectric constant (permittivity) of the material, and ϵ_o is that for free space. Thus,

$$V = \frac{Q}{C} = \frac{dF}{\epsilon_r \epsilon_o \frac{A}{t}} = \frac{dtF}{\epsilon_r \epsilon_o A} \quad (3-61)$$

Expressing g as the crystal voltage sensitivity factor, the voltage can be written as

$$g = \frac{d}{\epsilon_r \epsilon_o} \text{ Vm/N} \quad (3-62)$$

$$V = \frac{gtf}{A} = g \cdot t \cdot P$$

Also, $g = \frac{V}{t \cdot P} = \frac{V/t}{P}$ where V/t is the *electrical field strength* and P is the pressure or stress.

Table 3-4 presents the basic properties and characteristics of typical piezoelectric materials.

TABLE 3-4 BASIC CHARACTERISTICS OF PIEZOELECTRIC MATERIALS

Material	Density ($\times 10^3 \text{ kg/m}^3$)	Permitivity ϵ_r	Young's Modulus $E (10^{10} \text{ N/m}^2)$	Piezoelectric Charge Sensitivity $d (\text{pF/N})$
Quartz(SiO ₂)	2.65	4.5	7.7	2.3
Barium Titanate BaTiO ₃	5.7	1700	11	78
PZT	7.5	1200	8.3	110
PVDF	1.78	12	0.3	20 to 30 (based on crystal axes)

Typical values for g , which is the crystal voltage sensitivity factor, are

$$\text{BaTiO}_3 = 12 \times 10^{-3} \text{ Vm/N} \quad \text{Quartz} = 50 \times 10^{-3} \text{ Vm/N}$$

Typical values of permitivity, ($\epsilon_r \epsilon_o$), are

$$\begin{aligned} \text{BaTiO}_3 &= 12.5 \times 10^{-9} \text{ F/m} \\ &= (1700)(8.85)(10)^{-12} \\ \text{Quartz} &= 40 \times 10^{-12} \text{ F/m} \\ &= (4.5)(8.85)(10)^{-12} \end{aligned}$$

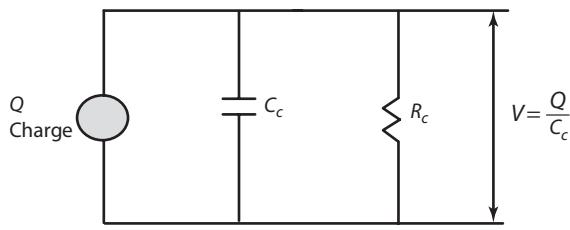
Piezoelectric materials are used in a variety of applications where force, pressure, acceleration, and vibration measurements are taken. The major application of the piezoelectric sensor is in situations where the charge does not have much time to leak off. It is also used as the sensor in ceramic- or crystal-type pick ups, where the needle causes distortion of the crystal and the voltages generated are amplified by charge amplifiers, which have the additional capacity of reducing loading effects on piezoelectric transducers.

Sensitivity, natural frequency, nonlinearity, hysteresis, and temperature effects are the primary considerations when selecting piezoelectric transducers. Piezoelectric pressure sensors are used for the measurement of rapidly varying pressures as well as shock pressures. Sensors made of quartz materials generally exhibit stable frequency response from 1 Hz to 20 kHz—the natural frequency being of the order of 50 kHz. Quartz crystals can be used over a temperature range of -185 to $+288$ °C compared to ceramic devices, which are limited to -185 to $+100$ °C.

Equivalent Circuit of a Piezoelectric Transducer The dynamic properties of a piezoelectric transducer can be represented by an equivalent circuit derived from the electrical and mechanical parameters of the transducer. The basic equivalent circuit is shown in Figure 3-49. The charge generated, Q , is across the capacitance C_c , and its leakage resistance is R_c . The charge source can be replaced by a voltage source, as per Equation 3-63 and drawn in series with a capacitance C_c and resistance R_c .

$$V = \frac{Q}{C} = \frac{dF}{C_c} \quad (3-63)$$

FIGURE 3-49 EQUIVALENT CIRCUIT

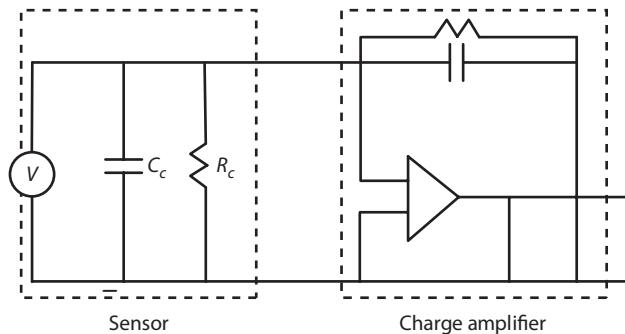


When the piezoelectric crystal is coupled with leads and cables as well as a readout device, the voltage depends not only on the element but also on the capacitance of cables, charge amplifier, and display. The total capacitance is expressed as

$$C_T = C_c + C_{\text{cable}} + C_{\text{display}}$$

A typical arrangement is shown in Figure 3-50, where the sensing element and charge amplifiers are presented.

FIGURE 3-50 CHARGE AMPLIFIER FOR PIEZOELECTRIC TRANSDUCER



The feedback resistance of the charge amplifier is kept high so that this circuit draws very low current and produces a voltage output that is proportional to the charge. Figure 3-51 shows the piezoelectric crystal interface, and the combined equivalent circuit is shown in Figure 3-52.

FIGURE 3-51 PIEZOELECTRIC CRYSTAL INTERFACE

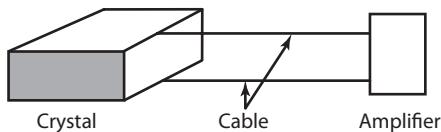


FIGURE 3-52 COMBINED EQUIVALENT CIRCUIT

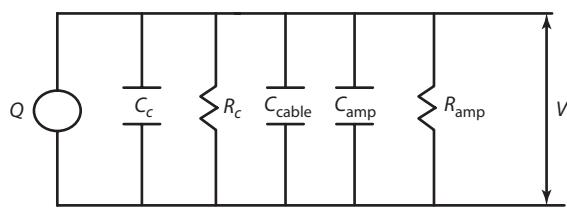


Figure 3-53 presents a photograph of a piezoelectric pressure transducer manufactured by the Kistler Instruments Corp.

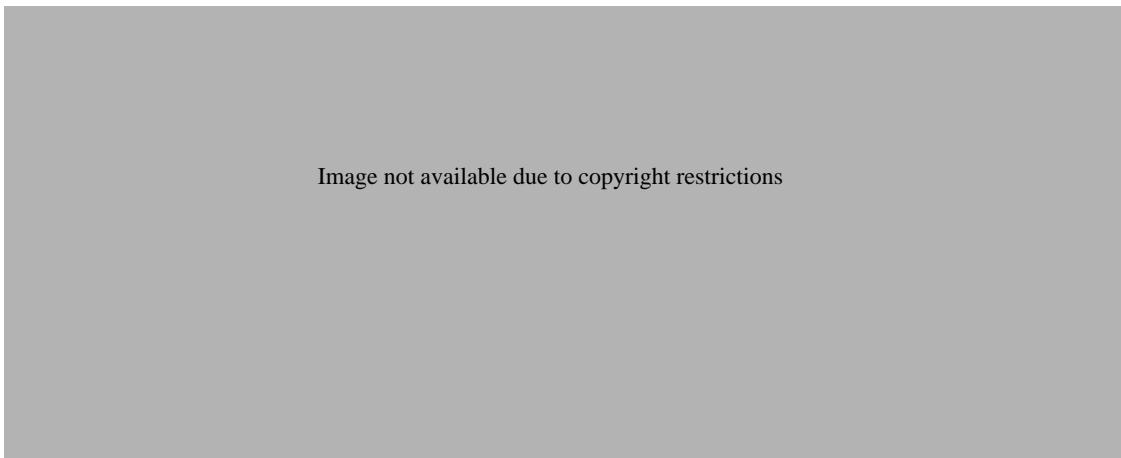


Figure 3-54 presents a photograph of a piezoelectric translator used for high-precision motion measurement.

Figure 3-55 presents a photograph of a rotating cutting force dynamometer for machine-tool applications.

Analogy Equations Using mathematical models, solutions of equations describing one physical form can be applied to analogous systems in other fields. The analogy approach is discussed in detail

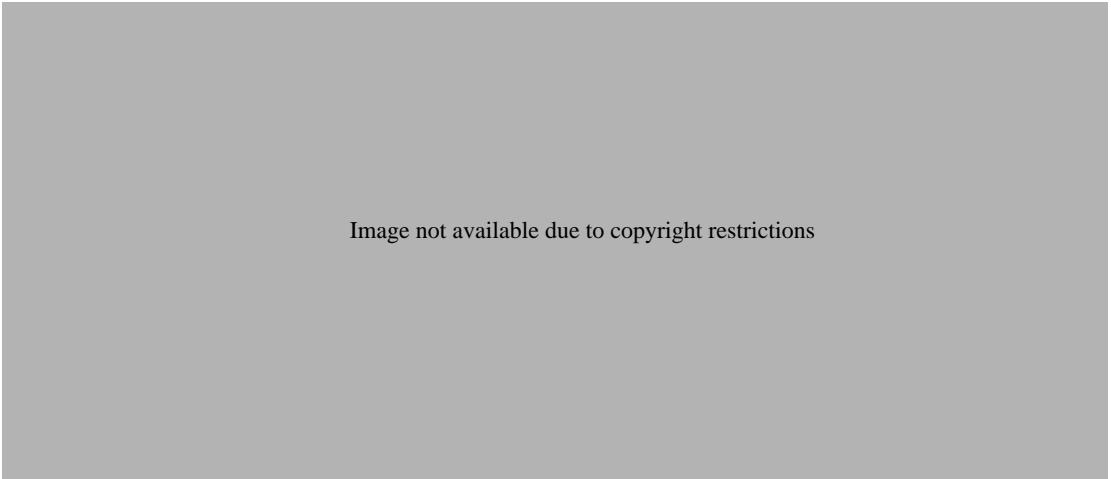


Image not available due to copyright restrictions

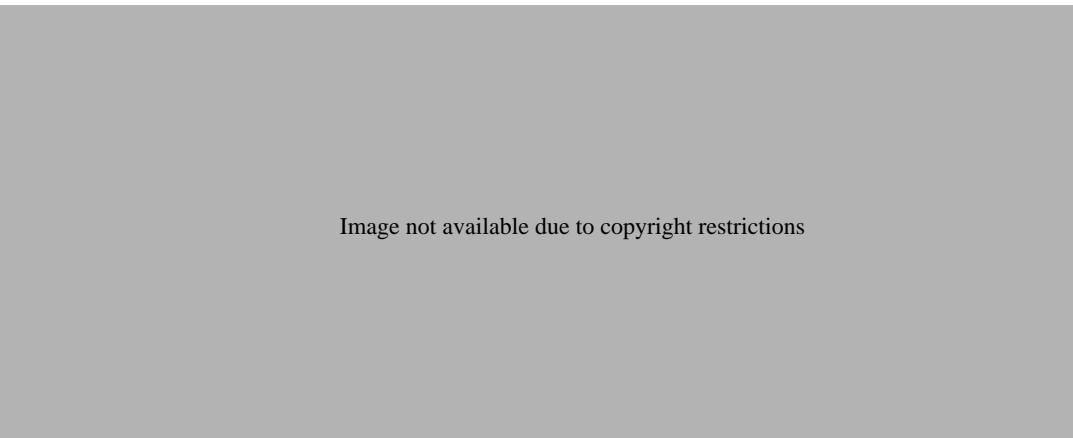


Image not available due to copyright restrictions

in earlier chapters. These analogies also can be applied to a piezoelectric transducer element. Using mechanical elements (such as inertial elements, spring, and damper), a mechanical system can be analyzed. C , L , and R represent mechanical parameters of compliance, mass, and viscous resistance of the element, respectively. The mechanical analogy in terms of displacement can be expressed as

$$F = m \frac{d^2x}{dt^2} + c \frac{dx}{dt} + \frac{x}{k}$$

$$v = \frac{dx}{dt}$$

$$F = m \frac{dv}{dt} + cv + \frac{1}{k} \int v dt$$

Differential equations in terms of current and velocity are also developed. R , L and C represent the viscous resistance of an element, mass, and parameters of mechanical parameters of compliance. In a $R-L-C$ series electrical network, the applied voltage equals the drop across the resistor, plus the drop across the inductor, plus the drop across the capacitor.

$$\text{Electrical; } = Ri + L \frac{di}{dt} + i + \frac{1}{c} \int idt$$

The configuration of the piezoelectric element is an important consideration in the industrial use of these elements. The shape of the element could be a disc, plate, or in tubular form. It may be operated under normal, transverse, or shear modes. For example, a small piezoelectric transducer 4 mm in diameter and 10 mm long weighs around 2 grams, operates at 177 °C, and has voltage sensitivity of 0.1 mV/N.

Acceleration Measurement by Piezoelectric Transducer The piezoelectric accelerometer is constructed as follows. It consists of a housing and contains a mass attached to the mechanical axis of the crystal. The piezoelectric element in the form of a cylinder is first bonded to a central pillar. Then a cylinder mass is bonded to the outside of the PZT element. Acceleration in the direction of the cylinder axis causes a shear force on the element, which provides its own spring force. The acceleration of the piezoelectric material generates electric potential when subjected to mechanical strain along a predetermined axis. The initial calibrating force is predetermined between the mass and spring using a preloaded spring.

As the housing of the accelerometer is subjected to vibrations, the force exerted on the piezoelectric element by the mass is altered. The charge generated on the crystal is sensed using a charge amplifier. A force F applied to the crystal develops a charge, $Q = dF$. When a varying acceleration is applied to the mass crystal assembly, the crystal experiences a varying force.

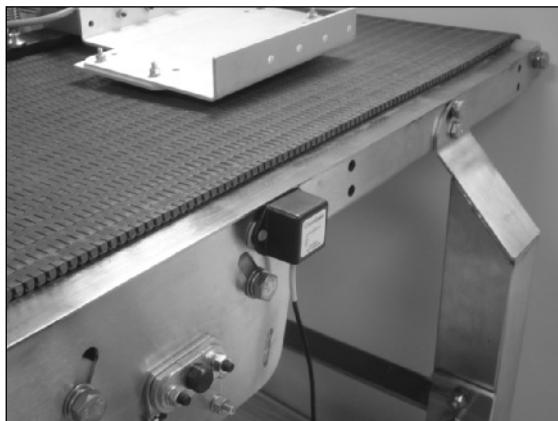
$$\begin{aligned} F &= Ma \\ Q &= dF = dMa \\ V &= \frac{dF}{C} = \frac{dMa}{C} \end{aligned} \tag{3-64}$$

Here a is the acceleration and V is the voltage produced. Thus, the output is a measure of the acceleration. Figure 3-56 presents a photograph of a typical accelerometer.

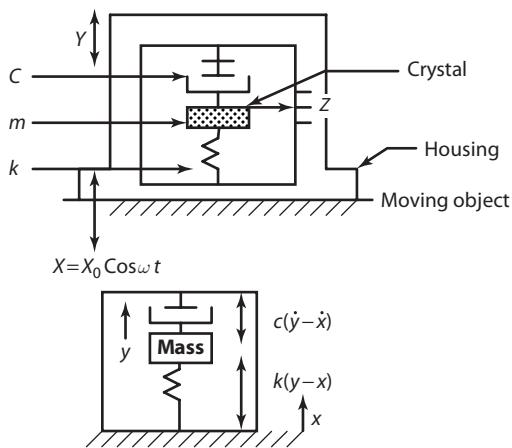
Because of the high stiffness of the piezoelectric material, the natural frequency of such devices can be as high as 125 kHz, which provides an ability to measure at high frequencies. The accelerometer (Figure 3-57) is of small size and has a small weight (0.25 kg). The crystal is a source with high output impedance, and the electrical matching of the impedance between the transducer and the circuitry is usually a critical matter in the design of the display system. Piezoelectric materials used as sensing elements for acceleration have been employed in seismic instrumentation.

The base of the device is attached to the object whose motion is to be measured. Inside the piezoelectric acceleration transducer, mass m is supported on spring of stiffness k and a viscous damper with damping coefficient c . The motion of the object results in the motion of the mass relative to the frame. A transducer equation is obtained by considering the inertial forces of the mass and the restoring force of the spring and the damper.

$$m \frac{d^2y}{dt^2} + c \frac{dy - x}{dt} + k(y - x) = 0$$

FIGURE 3-56 ACCELEROMETER

Shetty, University of Hartford.

FIGURE 3-57 PIEZOELECTRIC ACCELEROMETER

where y = absolute motion of the mass.

The relative motion, $z = y - x$, is expressed as

$$\begin{aligned} m \frac{d^2(z + x)}{dt^2} + c \frac{dz}{dt} + kz &= 0 \\ (mD^2 + cD + k)z &= -mD^2x \end{aligned}$$

where $D = \frac{d}{dt}$. The equation is of second order and relates the input and output of the transducer.

Velocity Measurement by Piezoelectric Transducer It is possible to measure velocity by first converting the velocity into a force using a viscous damping element and then measuring the

resulting force with a PZT sensor. If acceleration data is available through an accelerometer, the velocity is obtained by integrating this device. Velocity transducers are constructed using piezoelectric accelerometers and integrating amplifiers. Double integration provides displacement information. The principle of piezoelectric velocity transducer is illustrated in Figure 3-58.

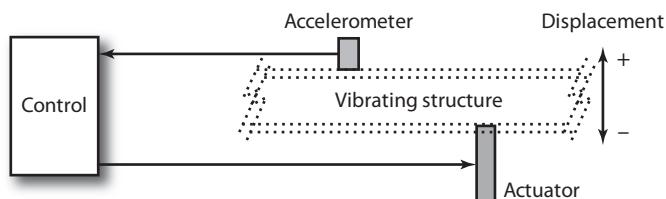
FIGURE 3-58 VELOCITY TRANSDUCER



3.6.2 Active Vibration Control

Active vibration control can be defined as a technique in which the vibration of a structure is reduced by applying a counterforce to the structure that is appropriately out of phase but equal in force and amplitude with the original vibration. As a result, two opposing forces cancel with each other, and the structure essentially stops vibrating. A schematic of a representative active vibration control system is shown in Figure 3-59.

FIGURE 3-59 SCHEMATIC OF ACTIVE VIBRATION CONTROL SYSTEM



The vibration control system consists of a high-speed microprocessor-based system, a vibrating structure, and an actuator. The structural vibrations are monitored by a motion sensor, such as an accelerometer. The resulting output voltage from the motion sensor is fed into a high-speed digital-signal processing device. The processing device calculates the appropriate phase inversion and the counterforce amplitude needed to reduce the original vibration characteristics. The output voltage from the computer is amplified and drives the actuator. The expansion and contraction of the actuator produces a force which counteracts the original vibration amplitude and reduces the vibration of the structure. It should be noted that this vibration control must theoretically take place in real time with the original vibration. It also should be noted at this point that, in practice, the vibration of a structure can not be stopped—it can only be reduced. This is essential due to the response-time limitations of the control system, the response-time limitations of the actuator itself, and the high rate of change of the structural vibration's spectral characteristics.

There are several areas where active vibration control can be applied. One such area is in isolating a mass from another vibrating mass rather than using traditional passive devices, such as springs and dampers. This is especially useful in the isolation of microelectronics and signal-processing units that are extremely sensitive to even the slightest vibrations. Another use of

active vibration control is in the precision manufacturing area. Vibrations and resultant acoustic emissions have the ability to damage the instrumentation and can be harmful for human health. Chatter and vibrations, if present in a machine-tool structure, also can make a severe impact on machining accuracies and can reduce surface quality. Elimination of unwanted vibrations created by a process can improve process accuracy. By controlling the vibrations of the cutting tools, closer tolerances can be achieved, and tool wear can be reduced. The advantage of active vibration control over other passive methods (i.e., springs and dampers) is that the structural vibrations can be reduced at a much faster rate.

3.6.3 Magnetostrictive Transducers for Vibration Control

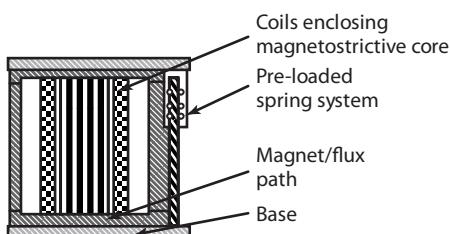
The piezoelectric type of actuator has been popular in active vibration control because of its fast response times. However, very high voltages are required to produce only micro-cm strains. Magnetostrictive materials, on the other hand, produce fairly substantial strains in the presence of relatively low magnetic fields. Magnetostrictive materials are also able to produce much higher counter forces. Magnetostrictive materials, however, do have high-frequency limitations, whereas the piezoelectric materials can oscillate well into the megahertz range. The actuator with the best promise for real-time vibration control is the magnetostrictive transducer.

Magnetostrictive Transducer Principle Magnetostriction is a property of certain materials, namely iron, nickel, cobalt, and respective alloys, whereby the material strains in the presence of a magnetic field. There are fifteen such rare earth elements that are part of the periodic table. The magnetic field is imposed by feeding a current through a coil surrounding the magnetostrictive material. The rare earth materials, especially magnetostrictive materials, are capable of producing strains of the order of 2,000 ppm. Certain alloys of iron and rare earth elements are capable of producing strains in excess of 2,000 ppm under certain circumstances. One such material is an alloy of terbium, iron, and dysprosium. Known commercially as Terfenol-D, it exhibits good magnetostrictive properties and is the most commonly used actuator element.

The basic elements of the actuator is shown in Figure 3-60. It consists of the coil which encloses the magnetostrictive rod, magnetic poles that conduct the flux to the rod, the DC flux from the permanent magnet to the rod, the air gap that allows the rod to expand and contract freely, the head and tail mass or the base, and spring systems that are used to provide the proper preload to the rod.

When a magnetostrictive material is surrounded by a coil and an AC current is fed to the coil, both the positive and negative portions of the cycle produce positive strains in the magnetostrictive material. However, this presents a problem when one wants to produce both positive and negative strains. This phenomenon is of importance while using the materials to actively control vibrations.

FIGURE 3-60 BASIC ELEMENTS OF THE TERFENOL-D ACTUATOR



In other words, the oscillating structure is pulled down (counterforce is applied in the negative direction) when the vibration amplitude is positive and pushes the oscillating structure up (i.e., counterforce is applied in the positive direction) when the vibration amplitude is negative. In both cases, the goal is to push or pull the vibration amplitude toward its neutral position so that the structural vibrations are significantly reduced.

The magnetostrictive strain, S , can be defined as the ratio of the expansion length, Δl to the original length, l , due to the applied magnetic field intensity, H . The magnetic field intensity, H , provided by the coil to the magnetostrictive material is defined as

$$S = \frac{\Delta l}{l} \quad (3-65)$$

$$H = \frac{NI}{l_c}$$

where I is the current through the coil, N is the number of coil turns, and l_c is the axial length of coil turns. The magnetostrictive actuator, if used in the linear region, converts electrical energy into mechanical energy. It also can be used to convert mechanical energy into electrical energy. It can be seen that the device may be used as both a transducer and an actuator.

Applications In the design of magnetostrictive transducer for real-time applications, the problem of the material straining in only one direction in the presence of both positive and negative currents is addressed by introducing a biasing field. The bias is usually accomplished either by placing a permanent magnet around the material or by introducing a DC bias field into the circuit. Due to the magnetic field from the permanent magnet, the material experiences an initial expansion or strain. The design size of the permanent magnet is suitably chosen so that the initial expansion is about one half the total expansion limit of the magnetostrictive material used. When the positive cycle of the AC current is presented, the field from the magnet and the field from the coil gets added, resulting in positive expansion of the material. When the negative cycle of the current is presented, the two fields cancel each other, and the material shrinks. Through the use of biasing, the actuator can be used to control the oscillating structures. If the use of the magnetostrictive actuator is limited to positive strain, a bias is not required for the application.

Magnetostrictive materials can operate from cryogenic temperatures up to 200°C. The transducer is highly reliable because of the minimal number of moving parts. Some of the current applications of magnetostrictive transducers include robotics, valve control, micro-positioning, and active vibration control. Other areas of applications include fast-acting relays, high-pressure pumps, and as high-energy, low-frequency sonic sources.

SUMMARY

Piezoelectric Transducer

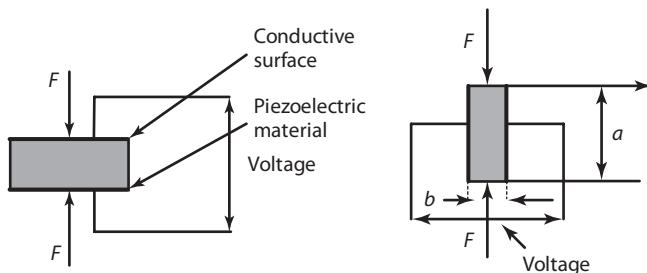
Piezoelectric materials, when subjected to mechanical force or stress along specific planes, generate electric charge. The best-known natural material is quartz crystal (SiO_2). Rochelle salt is also considered a natural piezoelectric material.

For the arrangement shown in Figure 3-61, the charge generated, Q , is defined as

$$Q = dF \text{ (Longitudinal effect)}$$

$$Q = dF \frac{a}{b} \text{ (Transverse effect)}$$

Here d is the *piezoelectric coefficient* of the material.

FIGURE 3-61

Magnetostrictive Transducer Theory

Magnetostriction is a property of certain materials, namely iron, nickel, cobalt, and respective alloys, whereby the material strains in the presence of a magnetic field. The most commonly used actuator element is commercially known as “Terfenol-D.”

When a magnetostrictive material is surrounded by a coil and an AC current is fed to the coil, both the positive and negative portions of the cycle produce positive strains in the magnetostrictive material. This phenomenon is of importance while using the materials to actively control vibrations.

Applications

- Piezoelectric materials are used in a variety of applications where force, pressure, acceleration, and vibration measurements are taken.
- Used as the sensor in ceramic- or crystal-type pick ups where the needle causes distortion of the crystal and the voltages generated are processed.

Features

- Sensitivity, natural frequency, nonlinearity, hysteresis, and temperature effects are the primary selection considerations.
- Sensors made of quartz materials generally exhibit stable frequency response from 1 Hz to 20 kHz, with the natural frequency being of the order of 50 kHz.
- Quartz crystals can be used over a temperature range of -185 to $+288$ °C compared to ceramic devices, which are limited to -185 to $+100$ °C.

Applications of Magnetostrictive Transducers

Current applications include micro-positioning, stress measurement. Other engineering applications include inspection of steel pipes and tubes, condition monitoring of machinery such as combustion engines, and onboard sensing of crash events for vehicle safety system operations.

3.7 Sensors for Flow Measurement

Flow sensing for measurement and control is one of the most critical areas in the modern industrial process industry. Regardless of the state of the fluid, gas, or liquid, accurate flow measurements are critical. In some situations, optimum performance of a machine is dependent on the

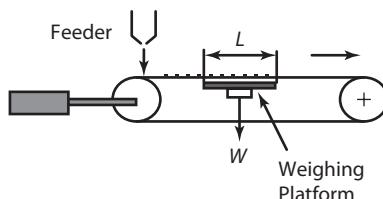
correct mix of definite proportions of liquids. The continuous manufacturing process relies on accurate monitoring and inspections involving raw materials, products, and waste throughout the process.

3.7.1 Solid Flow

While monitoring the bulk of solid materials in transit, it is necessary to weigh the quantity of material for some fixed length of the conveyor system. A flow transducer in a solid measurement is actually the assembly of a conveyor, hopper opening, and weighing platform. Small crushed particles of a solid material are carried by conveyor belt or through pipes in a slurry which is pumped through the pipes.

As can be observed from Figure 3-62, the flow is measured as the necessary weight of the quantity of material on a fixed length of the conveyor system.

FIGURE 3-62 SOLID FLOW MEASUREMENT



In this situation, the flow measurement becomes weight measurement. The material on the platform displaces a transducer, usually a load cell, which is calibrated to provide an electrical output proportional to the weight of the solid flow. Weight is usually measured by a load cell, which is calibrated to give an indication of the solid flow.

$$\text{Flow rate } Q = \frac{WR}{L} \quad (3-66)$$

where

Q = flow (kg/min)

W = weight of material on section of length L

R = conveyor speed (m/min)

L = length of weighing platform (m)

3.7.2 Liquid Flow

The basic continuity equation in flow calculations is the continuity equation which states that if the overall flow rate in the system is not changing with time then the flow rate past any section is constant. The continuity equation in the simplest form can be expressed as

$$V = Q/A$$

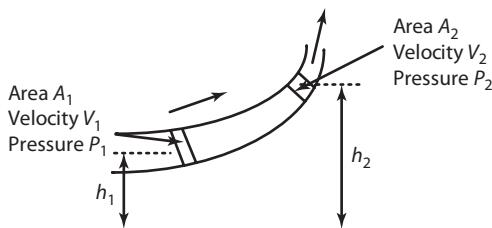
where

V = flow velocity

Q = volume flow rate

Volume flow rate is expressed as a volume delivered per unit time. The common units are cubic meters per hour and litres per hour. Mass flow rate or mass of flow per unit time is expressed in kg/hr. Figure 3-63 illustrates the fluid flow phenomenon through varying cross-sectional areas.

FIGURE 3-63 LIQUID FLOW THROUGH VARYING CROSS-SECTIONAL AREA



Incompressible fluid flow through a pipe under equilibrium conditions can be expressed by Bernoulli's theorem, which states that the sum of the pressure head, velocity head, and elevation at one point is equal to another point.

Equation 3-67 represents conservation of energy with no energy loss between points A and B . The first term represents energy stored as pressure; the second term represents kinetic energy; and the third term represents energy due to position.

$$\frac{P_2}{\rho} + \frac{V_2^2}{2g} + h_2 = \frac{P_1}{\rho} + \frac{V_1^2}{2g} + h_1 \quad (3-67)$$

$$Q = EA_2 \sqrt{\frac{2g(P_1 - P_2)}{\rho}} \text{ where } E = \frac{1}{\sqrt{1 - \left(\frac{A_2}{A_1}\right)^2}} \quad (3-68)$$

where

V_1, V_2 = mean fluid velocity at points 1 and 2 (m/s)

ρ = fluid density (N/m^3)

P_1 and P_2 = pressures at two different points

g = acceleration of gravity

h_1 and h_2 = elevation above a given datum level

The most common flow-measurement technique is to measure a pressure differential along a flow line. Sensors based on differential pressure measurement, rotameters, ultrasonic flow transducers, turbine flow transducers, electromagnetic flow transducers and laser anemometers are used for this measurement.

3.7.3 Sensors Based On Differential Pressure

Flow sensors of this type use an obstruction along the flow line, such as a nozzle, orifice plate, Venturi tube, or pitot tube. Using Bernoulli's equation with some modification, the basic relationship between the pressure differential and flow rate is expressed as

$$Q = \frac{C_d a}{\sqrt{1 - \left(\frac{a}{A}\right)^2}} \sqrt{\frac{2 \Delta p}{\rho}} \quad (3-69)$$

where

p = density of fluid

a = area of cross section pipe at constriction

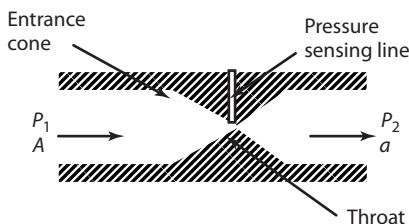
A = area of cross section pipe prior to constriction

Δp = pressure differential between two tapping points

C_d = discharge coefficient

The discharge coefficient indicates the amount of disturbance to the flow stream at the area of restriction, called the throat (Figure 3-64). This illustrates the flow sensing principle using an obstruction.

FIGURE 3-64 FLOW SENSING

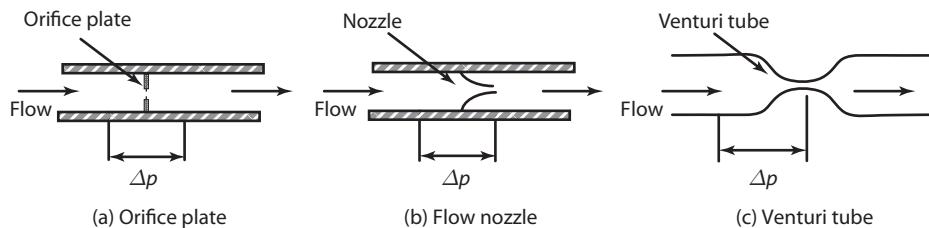


The conventional devices for flow sensing employ one of the following three arrangements,

1. Orifice plate
2. Nozzle
3. Venturi tube

As shown in Figure 3-65, these all use a calibrated restriction in the flow line and thereby measure the pressure drop across the obstruction. The velocity of flow is considerably higher on the downstream side of the obstruction. According to Bernoulli's theorem, there is a pressure drop, and the magnitude of this drop is proportional to the velocity of flow through the obstruction. The relationship between the pressure drop and the flow velocity is nonlinear. In addition, the obstruction must be designed for a specific range of flows and velocities. Flows with lower velocities may not register any substantial pressure drop.

The orifice plate flow transducer is the least expensive device but has a limited measurement span. It can be used for both liquid and gas flow with reasonable accuracy. In orifice plate meters, circular

FIGURE 3-65 RATE OF FLOW SENSORS

holes are cut in thin plates and bolted between flanges along the length of the pipe. The pressure tapping for flow rate measurements can be obtained by a variety of methods. For pipes of 5 cm and larger, the pressure tappings are made at distances of D and $D/2$ in the upstream and downstream directions, respectively, where D is the diameter of the pipe. These instruments are inexpensive and generally have a long, maintenance-free life.

The nozzle and Venturi tubes are more sophisticated and expensive transducers compared to the orifice plate flow transducer. They are more accurate, operate over a wide range of flow, and are less susceptible to flow losses. Venturi tubes offer the best accuracy compared to nozzle flow and orifice plate transducers. Their design consists of three sections: the converging section at the upstream, the throat, and the diverging conical section at the downstream. The cylindrical throat section experiences a decrease in pressure and an increase in velocity. At this point, the flow rate is steady. The Venturi tube is expensive to construct and must be calibrated. Because of this, Venturi tubes are not suitable for fluids that collect on the tiny wall pockets as they flow.

The nozzle flow meter is similar to the Venturi meter but occupies considerably less space. The design of the nozzle combines the simplicity of the orifice plate with the low losses of the Venturi tube. The fluid passes through the minimum flow area and expands suddenly to the pipe area. The absence of a downstream cone brings the pressure loss to the same level of the orifice meter. Nozzle flow meters can be used for both liquids and gases in situations where the volumetric flow rate has to be measured with reasonable accuracy. They are less expensive than Venturi tubes, have a longer life, and do not require recalibration.

Pitot Tube The pitot tube is the oldest flow rate-sensing instrument. It transforms the kinetic energy of the fluid into potential energy in the form of a static head. The difference between the impact (or the dynamic pressure) and the static pressure can be related to the flow rate. The velocity head is converted into impact pressure, and the difference between the static pressure and the impact pressure becomes a measure of the flow rate.

The pitot tube is widely used for air speed measurements onboard aircraft. It consists of a cylindrical probe installed in a pipe line. As the fluid approaches the probe, the velocity decreases until it reaches zero at the point of impact on the probe. The deceleration increases the pressure. P_1 and V_1 are the upstream pressure and velocity, and P_2 and V_2 are the pressure and velocity in the neighborhood of the object. At the point of impact, V_2 is zero. From Bernoulli's theorem, the velocity of fluid flow is computed, as

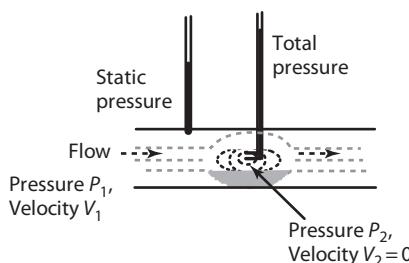
$$\frac{P_2}{\rho} = \frac{V_1^2}{2g} + \frac{P_1}{\rho} \quad (3-70)$$

Solving for velocity and introducing the correction factor, C_v , to account for nonuniform velocity in the pipe yields

$$V = C_v \sqrt{\frac{2}{\rho} g(P_2 - P_1)} \quad (3-71)$$

The Pitot tube in Figure 3-66 has two concentric tubes. The inner tube connects the impact hole to one side of a differential pressure gauge, and the outer tube has a series of holes bored into it to sense the static pressure. Velocity at a point is determined by the pressure differential generated by this pitot tube. Total pressure in the inner tube is equal to the sum of the static pressure and the pressure due to impact of the fluid stream.

FIGURE 3-66 STANDARD PITOT TUBE USED FOR FLOW MEASUREMENT



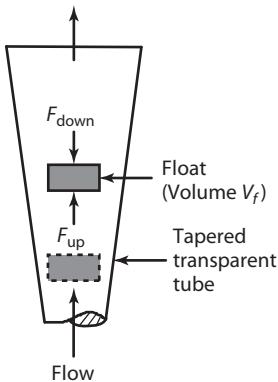
Rotameter The rotameter is another device widely used in the process-control industry for flow measurement. It consists of a tapered glass tube and a float. The float rises until the annular passage is larger enough to pass all material through pipe. The float is constructed with a diameter that completely blocks the inlet. When the flow starts in the pipeline and the fluid or gas reaches the float, the buoyant effect of fluid or gas makes the float lighter. The float passage remains closed until the pressure of the flowing material plus the fluid buoyancy effect exceeds the downward pressure due to the weight of the float. The float then rises and floats within the medium in proportion to the flow at a given pressure. The float then comes to dynamic equilibrium.

An increase in flow rate causes the float to rise, and a decrease in flow rate causes the float to drop. The forces acting on the float in the vertical column of the liquid are shown in Figure 3-67. The downward forces include the effective weight of the float, F_w , as well as the forces acting on the upper surface of the float, F_d . They are shown in Equation 6-68. The upward forces include the forces acting upward on the lower surface of the float, F_{up} , and the drag force, F_{drag} , which tends to pull the float in the upward direction. The value of this force depends upon the float design, the flow conditions, and the absolute viscosity of the fluid.

$$F_{down} = F_w + F_d = V_f(\rho_2 - \rho_1) + (p_2)A_f \quad (3-72)$$

V_f is the float volume, A_f is the surface area of the float, ρ_2 and ρ_1 are the densities of float material and liquid, respectively, and p_2 is the pressure per unit area on the upper surface of the float.

$$F_{up} = F_{up} + F_{drag} = (p_1)A_f + F_{drag} \quad (3-73)$$

FIGURE 3-67 SCHEMATIC OF ROTAMETER

Under equilibrium conditions and neglecting viscous drag forces, Equation 3-73 becomes

$$(p_1)A_f = V_f(\rho_2 - \rho_1) + (p_2)A_f \quad (3-74)$$

$$(p_2 - p_1) = \frac{V_f}{A_f}(\rho_2 - \rho_1)$$

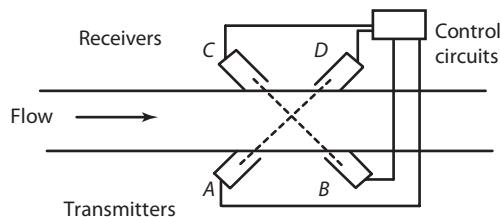
Substituting and accounting for the discharge coefficient produces the desired flow equation, we have

$$Q = C_d EA_2 \sqrt{2g \frac{V_f}{A_f} \left(\frac{\rho_2 - \rho_1}{\rho_1} \right)} \quad (3-75)$$

If the rotameter is connected to a variable inductance transducer, an electrical output can be generated in proportion to the flow. This principle is used in the induction variable area flowmeter. The rotameter acts as the primary sensor of the flow. An inductive transducer is the secondary transducer which provides a signal as an armature connected to it changes position as the float position changes. Two coils are connected to the arms of an AC bridge circuit. When the armature is symmetrically located with respect to the two coils, their impedances are equal, and the bridge is balanced, producing no output. If there is fluid flow, the float changes position resulting in the movement of the soft iron armature. This causes a change in the impedance of the coils. The bridge becomes unbalanced. Since the output voltage is a function of the flow rate, the output voltage is amplified and used to operate a servo motor.

3.7.4 Ultrasonic Flow Transducers for Flow Measurement

Ultrasonic flow meters measure fluid velocity by passing high-frequency sound waves through the fluid. Sometimes called transit time flowmeters, they operate by measuring the transmission time difference of an ultrasonic beam passed through a homogeneous fluid contained in a pipe at both upstream and downstream locations. Figure 3-68 illustrates the principles of ultrasonic flow sensing.

FIGURE 3-68 ULTRASONIC FLOW SENSING

The transducer consists of transmitter and receiver pairs. One pair, *A* and *B*, act as transmitters, and the other pair, *C* and *D*, act as receivers. If a sound pulse is transmitted from transmitter *B* to receiver *C*, the transit time is calculated as

$$t_{BC} = \frac{d}{\sin \alpha(C - V \cos \alpha)} \quad (3-76)$$

If the pulse is transmitted from transmitter *A* to receiver *D*, the transit time is

$$t_{AD} = \frac{d}{\sin \alpha(C + V \cos \alpha)} \quad (3-77)$$

where

d = diameter of the tube (m)

V = velocity of fluid flow (m/s)

α = the angle between the path of sound and the pipe wall

C = sound velocity in the fluid (m/s)—assume $V \ll C$

The transit time difference, Δt , is the difference between Equation 3-76 and Equation 3-77. It is proportional to flow velocity and fluid flow and can be used as an input to the computer. By measuring the transit times at both upstream and downstream locations, the fluid velocity can be expressed independently of the sound velocity in the fluid. Since the measurement is independent of the velocity of sound through the fluid, the effects of pressure and temperature are avoided.

$$\frac{t_{BC} - t_{AD}}{(t_{BC})(t_{AD})} = \frac{2V \sin \alpha \cos \alpha}{d} \quad (3-78)$$

Figure 3-69 presents a photograph of an ultrasonic level sensor with a digital read out.

Ultrasonic Doppler Flow Meter The Doppler effect is a useful technique used to measure the velocity of a fluid and hence its flow. In Doppler flow meters, continuous ultrasonic waves are beamed into the fluid. The transducer is normally bonded to the wall of the pipe so as to transmit a beam into the flow. The particles in the fluid scatter the beam and cause a frequency shift which is

FIGURE 3-69 ULTRASONIC LEVEL SENSOR

Courtesy of Gems Sensors, Inc. Plainville, CT.

proportional to the particle velocity. If f_r and f_t are the respective receiving and transmitting frequencies, then the Doppler shift, f_d , can be represented as

$$f_d = f_r - f_t$$

Ultrasonic flow meters are used to measure liquid velocities with minimal pressure loss. The flow measurement is insensitive to pressure, temperature, and viscosity variations. The method has advantages, including bi-directional sensing, high accuracy, wide ranges, and a rapid response. Although it is an expensive technique, it can be employed for measurement in tubes and pipes of varying sizes.

3.7.5 Drag-Force Flow Meter

In this type of flow meter, a suitable obstruction is inserted into the flow path. As a result, the fluid applies a drag force on the object which is sensed and used as a measure of the flow. The drag force, F_d , acting on the object immersed in the fluid is represented by Equation 3-79:

$$F_d = \frac{1}{2} C_d \rho g V^2 A \text{ (N)} \quad (3-79)$$

where

C_d is the coefficient of the drag

A is the area of cross section

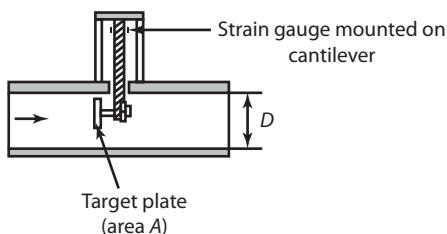
ρ is the fluid density $\left(\frac{\text{Kg}}{\text{m}^3} \right)$

V is the velocity (m/s)

The drag force of the body can be measured by attaching the drag body to a suitable force monitoring device.

Figure 3-70 shows a cantilever beam arrangement with bonded strain gauges. The drag force is transmitted as a strain in the cantilever beam. The strain is suitably calibrated and measured. The main advantage of this type of flow meter is its high dynamic response. The accuracy of the instrument is $\pm 0.5\%$ and repeatability $\pm 0.1\%$. Drag force flow meters are useful for highly viscous flows, such as hot asphalt, tar, or slurries at high pressures.

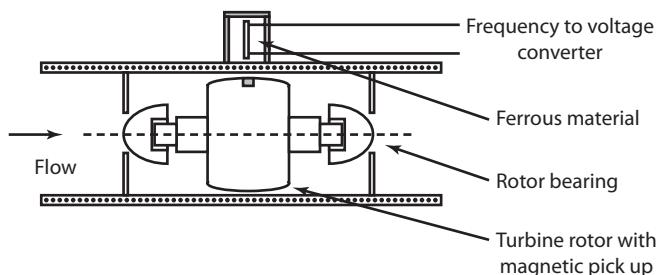
FIGURE 3-70 DRAG FORCE TYPE FLOW SENSOR



3.7.6 Turbine Flow Meter

The turbine flow meter is a popular method for flow measurement. As shown in Figure 3-71, a permanent magnet is enclosed in a rotary body. Each time the rotating magnet passes the pole of the pick up coil, the change in the permeability of the magnetic circuit produces a voltage signal at the output terminal. The output signal is a frequency that is proportional to the flow rate. The voltage pulse is counted by means of a digital counter to give the total flow.

FIGURE 3-71 FLOW SENSING BY TURBINE FLOW METER



The main advantage of the turbine flow meter is the linear relationship between the volume flow rate and the angular velocity of the rotor which is

$$Q = kn \quad (3-80)$$

where

Q is the volume flow rate

k is a constant depending on the fluid property

n is the rotor angular velocity (rad/s)

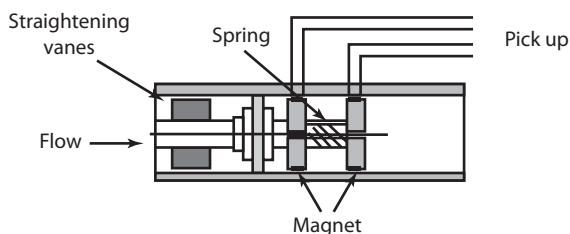
Turbine flow meters are not suited for fluids that contain abrasive particles. Any damage to the turbine blades must be followed by an immediate recalibration of the meter. The paddle wheel flow meter is a variation of the turbine flow meter. In such flow meters, the fluid drives a small paddle wheel that is located on the side of the pipe.

3.7.7 Rotor Torque Mass Flow Meter

In some applications, it is necessary to measure the mass flow rate rather than the volume flow rate. Such applications exist in process-control industries as well as aerospace industries where mass flow rate information is needed. The measurement concept is based on Newton's second law of motion, wherein the force required to alter the velocity of the fluid stream is used as a measurement.

Figure 3-72 describes the basic rotor torque mass flow meter. The fluid is given a constant rotational velocity in a direction normal to the direction of flow. The fluid is first passed through straightening vanes to remove any angular swirls and then allowed to flow through an assembly which consists of a set of vanes rotating at constant speed about the axis of the flow meter.

FIGURE 3-72 ROTOR TORQUE MASS FLOW METER



The torque needed to drive the rotating vanes is proportional to the magnitude of the angular momentum applied to the fluid, which in turn is proportional to the mass of the fluid through the assembly.

The torque, T , transmitted to the impeller is expressed by

$$T \propto \frac{d}{dt} (I\omega) \quad (3-81)$$

$$I = mk^2$$

$$T \propto \frac{d}{dt} (mk^2\omega)$$

$$\therefore$$

$$T \propto mk^2\omega$$

where

T = torque transmitted

I = mass moment of inertia

ω = angular velocity

k = radius of gyration

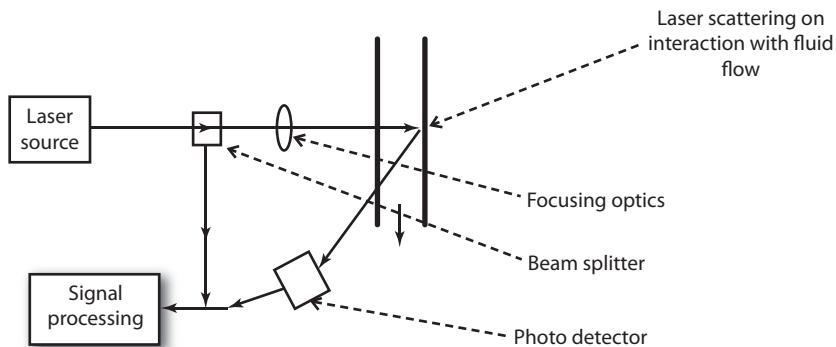
3.7.8 Fluid Measurement Using Laser Doppler Effect

Laser Doppler anemometers utilize a non-invasive procedure to measure the instantaneous flow velocities of liquids or gases flowing in a transparent channel. The technique can be employed only in situations where

- Adequate transmission of laser light through the fluid is possible.
- The fluid contains sufficient particles of contamination so that the laser beam can use the effect of scattering.

As shown in Figure 3-73, the principle is based on the Doppler shift phenomenon in which the frequency of the scattered light from the moving object differs from that of the incident beam by an amount proportional to the fluid velocity.

FIGURE 3-73 LASER DOPPLER ANEMOMETER



A laser beam is focused at a point in the fluid where the velocity is to be measured. The laser beam is scattered by the small particles flowing in the liquid. Due to viscous effects, the small particles move at the same velocity as the fluid, so the measurement of the particle velocity is the same as the fluid velocity. Signal processing of the photodetector output produces the magnitude of the Doppler frequency shift, which is directly proportional to the instantaneous velocity of flow.

$$\text{Frequency shift: } \Delta f = \frac{2V \cos\theta}{c} f_0 \quad (3-82)$$

Here, V is the particle velocity, f_0 is the frequency of the laser beam, θ is the angle between the laser beam and the particle in the fluid, and c is the speed of light. The output voltage of the instrument is directly proportional to the instantaneous velocities of the fluids.

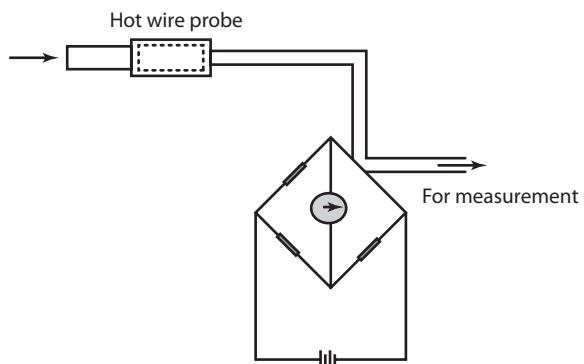
Related developments in the area of laser anemometry include the dual-beam laser velocimeter, which looks at the interference pattern of two laser beams interacting on the fluid at a plane. The interaction results in a fringe pattern, and the fringe separation is a measure of the fluid velocity. The laser Doppler velocimeter is used for a wide range of velocities of fluid and gas flows. High accuracy's in the range of $\pm 0.2\%$ are possible. These instruments have been used in the aerospace industry to measure vortex flow near the wing tips of aircraft, flow between the gas turbine compressor blades, investigation of boundary layers, combustion phenomenon in jet propulsion systems, and in biological areas for *in vivo* blood-flow measurement.

3.7.9 Hot Wire Anemometers

Hot wire anemometry is an important method of fluid velocity measurement and is primarily used for mean and fluctuating velocity measurements. The method is used in aerodynamic applications to measure liquids and gases at high speeds and to measure non-conductive liquids at low speeds.

Its operation is based on the principle that the convective heat transfer from a small $5 \mu\text{m}$ diameter platinum-tungsten wire is a function of the fluid velocity. The wire is heated by the passage of current through it (Figure 3-74). When it is exposed to the fluid flow, heat is dissipated from the wire by convection, and there is a decrease in the wire resistance. The rate of heat loss depends on the shape and characteristics of the wire, properties of the fluid, and the fluid velocity. By maintaining the first two factors at constant values, the instrument response becomes a function of the fluid velocity only.

FIGURE 3-74 SCHEMATIC OF HOT WIRE OPERATION



Basic heat-transfer equations can be explained using King's law for convective heat transfer from the heated wire: by

$$\frac{hD}{K} = 0.3 + 0.5 \left(\frac{\rho v D}{\mu} \right)^{0.5} \quad (3-83)$$

where

h = convective coefficient of heat transfer

K = thermal conductivity of hot wire

ρ = density of fluid

v = velocity of fluid stream

D = diameter of hot wire

μ = coefficient of viscosity of the fluid

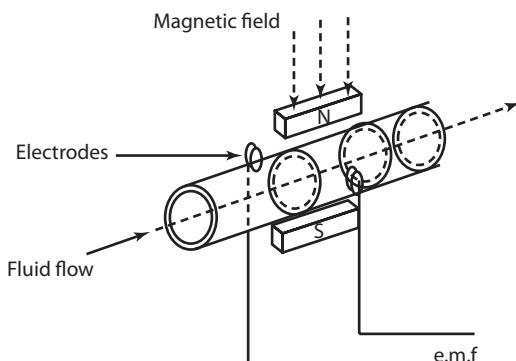
The output of the bridge circuit with a calibrated computer interface provides a measure of the fluid flow velocity. Hot wire anemometers are suited for measurement in clean fluids. One important application is the measurement of fluid turbulence achieved by using proper compensation circuitry and calibration.

3.7.10 Electromagnetic Flow Meter

The operating principle of the electromagnetic flow meter is based on the voltage which is generated in an electrically conducting fluid as it moves through a magnetic field. This method is useful for measuring flows of conducting liquids that may have abrasive materials and are not suited for other measurement methods. It cannot be used for electrically non-conducting fluids (like gases) and produces satisfactory results for low conductivity fluids (like water).

Figure 3-75 illustrates the operating principle of the electromagnetic flowmeter. In electromagnetic flow sensing, a pair of electrodes are inserted on the opposite sides of a non-conducting and nonmagnetic pipe which carries the liquid. The pipe is surrounded by an electromagnet, which produces the magnetic field. The voltage is induced across the electrodes. The magnitude of the emf is proportional to the rate at which the field lines are cut. Assuming a constant magnetic field, the magnitude of the voltage appearing across the electrodes will be proportional to the velocity.

FIGURE 3-75 ELECTROMAGNETIC FLOW METER



According to Faraday's law, the induced voltage, e , is given by

$$e = Blv \times 10^{-8} V \quad (3-84)$$

where

B = magnetic flux density

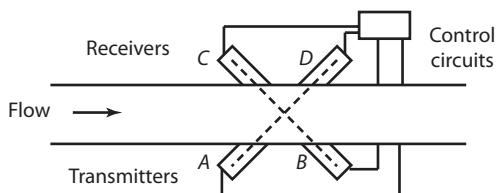
l = length of the conductor (pipe diameter)

V = velocity of the conductor (cm/s)

Electromagnetic flow sensing can be used in pipes of any size. The use of electro-magnetic sensors will not cause any obstruction in the fluid flow and will not cause any specific pressure drop. The output voltage has a large linear range and a good transient response. The output is not affected by variations in viscosity, pressure, or temperature. In summary, electromagnetic flow meters are useful for monitoring corrosive fluids, solid contaminated liquids, paper pulp, detergents, cement slurries, and greasy liquids.

SUMMARY**Flow Sensors****Flow Sensors For Flow Measurement:**

Ultrasonic flow meters measure fluid velocity by passing high frequency sound waves through the fluid. They operate by measuring the transmission time difference of an ultrasonic beam passed through a homogeneous fluid contained in a pipe at both an upstream and downstream location.

FIGURE 3-76 ULTRASONIC FLOW SENSING**Measurement Using Laser Doppler Effect:**

This principle is based on the Doppler shift phenomenon in which the frequency of the scattered light from the moving object differs from that of the incident beam by an amount proportional to the fluid velocity. The beam is focused at a point in the fluid where the velocity is to be measured. Signal processing of the photodetector output produces the magnitude of the Doppler frequency shift which is directly proportional to the instantaneous velocity of flow.

$$\text{Frequency shift; } \Delta f = \frac{2V \cos\theta}{c} f_0$$

where V is the particle velocity, f_0 is the frequency of the beam, θ is the angle between the laser beam, and the particle c is the speed of light. The output voltage of is proportional to the instantaneous velocities of the fluids.

Applications

These techniques have been used in the aerospace industry to measure vortex flow near the wing tips of aircraft, flow between the gas turbine compressor blades, investigation of boundary layers, combustion phenomenon in jet propulsion systems, and in biological areas for *in vivo* blood-flow measurement.

Features

High accuracy in the range of $\pm 0.2\%$ is possible.

Electromagnetic Flow Meter Theory**Principle:**

The electromagnetic flow meter is based on the voltage which is generated in an electrically conducting fluid as it moves through a magnetic field. A pair of electrodes are inserted on the opposite sides of a non-conducting and nonmagnetic pipe which carries the liquid. The pipe is surrounded by an electromagnet, which produces the magnetic field. The voltage is induced across the electrodes. The magnitude of the emf is proportional to the rate at which the field lines are cut. Assuming a constant magnetic field, the magnitude of the voltage appearing across the electrodes will be proportional to the velocity.

Applications

Electromagnetic flow meters are useful for monitoring corrosive fluids, solid contaminated liquids, paper pulp, detergents, and cement slurries.

Features

- Can be used in pipes of any size.
 - Use of electro-magnetic sensors will not cause any obstruction in the fluid flow
 - The output has a large linear range and a good transient response. The output is not affected by variations in viscosity, pressure and temperature.
-

3.8 Temperature Sensing Devices

Temperature is one of the most familiar engineering variables. Its measurement and control is one of the earliest known metrological achievements. Temperature measurement is based on one of the following principles.

1. Material expansion based on change in length, volume, or pressure.
2. Based on the change in electrical resistance.
3. Based on contact voltage between two dissimilar metals.
4. Based on changes in radiated energy.

An RTD is a length of wire whose resistance is a function of temperature. The design consists of a wire that is wound in the shape of a coil to achieve small size and improve thermal conductivity. In many cases the coil is protected from the environment by a protecting tube which inevitably increases response time, however, this enclosure is essential when RTDs are used in hostile environments.

Resistance relationships of most metals over a wide range of temperature are given by quadratic equations. A quadratic approximation to the $R-T$ curve is a more accurate representation of the resistance variation over a span of temperatures. It includes both a linear term and a term that varies as the square of the temperature. An analytical approximation is represented as,

$$R = R_o(1 + \alpha(T - T_0) + \beta(T - T_0)^2 + \dots) \quad (3-85)$$

Here R_o is the resistance at absolute temperature T and α and β are material constants which dependent on the purity of material used.

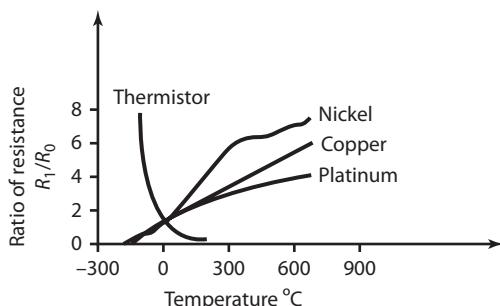
An examination of the resistance versus temperature curves of Figure 3-77 shows that the curves are quite linear in short ranges. This observation is employed to develop approximate analytical equations for resistance versus temperature of a particular metal.

Over a small temperature range of 0°C to 100°C , the linear relationship is written as,

$$R_t = R_0(1 + \alpha(T - T_0)) \quad (3-86)$$

Here α is the temperature coefficient of resistivity. Typical values of α for three materials are $\text{Cu} = 0.0043 /^\circ\text{C}$; $\text{Pt} = 0.0039 /^\circ\text{C}$; $\text{Ni} = 0.0068 /^\circ\text{C}$.

An estimation of RTD sensitivity can be calculated from typical values of the linear fractional change in resistance with temperature, as shown in Figure 3-76. The sensitivity for platinum is

FIGURE 3-77 RESISTANCE TRANSDUCER CHARACTERISTICS (OF PURE METALS)

0.004/°C and for nickel is 0.005/°C. Usually, a specification provides calibration information, either as a graph of resistance versus temperature or as a table of values from which the sensitivity can be determined.

An RTD has a response time of 0.5 to 5 seconds or more. The speed of the response is governed by its thermal conductivity which governs the time required to bring the device into thermal equilibrium with its environment. The operating range of an RTD depends on the type of wire used as the active element, for example, a typical platinum RTD has an operating range between -100 to 650°C, and an RTD constructed from nickel has a range in the vicinity of -180°C to 300°C.

Variation of the resistance in a sensing element is measured using some form of electrical bridge circuit. Such a circuit may employ either the deflection mode of operation or the null mode. Resistance variations in a typical RTD tend to be quite small—in the vicinity of 0.4%. Because of these small fractional resistance changes with temperature, process-control applications require the use of a bridge circuit in which the null condition is accurately detected.

3.8.1 Thermistors

A thermistor is a temperature transducer whose operation relies on the principle of change in semiconductor resistance with change in temperature. The particular semiconductor materials used in a thermistor vary widely to accommodate temperature ranges, sensitivity, resistance ranges, and other factors. The characteristics depend on the peculiar behavior of semiconductor resistance versus temperature. When the temperature of the material is increased, the molecules begin to vibrate. Further increases in temperature cause the vibrations to increase, which in turn increase the volume occupied by the atoms in the metal lattice. Electron flow through the lattice becomes increasingly difficult, which causes electrons in the semiconductor to detach resulting in increased conductance. In summary, an increase in temperature decreases electrical resistance by improving conductance. The semiconductor becomes a better conductor of current as its temperature is increased. This behavior is just the opposite of a metal. An important distinction, however, is that the change in semiconductor resistance with respect to temperature is highly nonlinear.

Individual thermistor curves are approximated by the following nonlinear equation,

$$\frac{1}{T} = A + B \ln R + C(\ln R)^3 \quad (3-87)$$

where

T = temperature in kelvins

R = resistance of thermistor

A, B, C = curve fitting constants

The temperature range measured with a typical thermistor is between -250°C and 650°C . The high sensitivity of the thermistor is one of its significant advantages. Changes in resistance of 10% per degree Celsius are not uncommon.

Because a thermistor exhibits such a large change in resistance with respect to temperature, there are many possible circuits which can be used for their measurement. A bridge circuit with null detection is most frequently used because the nonlinear behavior of the thermistor makes it difficult to use as a primary measurement device. Thermistors using null-detecting bridge circuits and proper signal conditioning provide extremely sensitive temperature measurements.

Since the thermistor is a bulk semiconductor, it can be fabricated in many forms including discs, beads, and rods varying in size from a bead of one millimeter in diameter to a disc several centimeters in diameter and several centimeters thick. By varying the manufacturing process and using different semiconducting materials, a manufacturer can provide a wide range of resistance values at any particular temperature.

The response time of a thermistor depends primarily on the quality and quantity of material present as well as the environment. When encapsulated for protection against a hostile environment, the time response is increased due to the protection from the environment.

3.8.2 Thermocouples

When two conductors of dissimilar material are joined to form a circuit the following effect is observed.

When the two junctions are at different temperatures, θ_1 , and θ_2 , small emf, e_1 and e_2 , are produced at the junctions and the algebraic sum of these causes a current.

This effect is known as the Seebeck effect. The Peltier effect is the inverse of the Seebeck effect and described as follows.

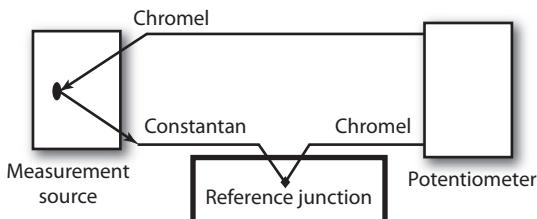
When the two dissimilar conductors which are joined together have a current passed through them, the junction changes its temperature as heat is absorbed or generated.

Another effect, called the Thomson effect, predicts that, in addition to the Peltier emf, another emf occurs in each material of a thermocouple which is due to the longitudinal temperature gradient between its ends when it forms part of a conductor.

When a thermocouple is used to measure an unknown temperature, the temperature of the thermo-junction, called the reference junction, must be known by some independent means and maintained at constant temperature.

Figure 3-78 shows a typical thermocouple circuit using a chromel constantan thermocouple, reference junction, and a potentiometric circuit to monitor the output voltage. Calibration of the thermocouple is performed by knowing the relationship between the output emf and the temperature of the measuring junction.

The standards for the production of thermocouples are provided by The National Institute of Standards and Technology (NIST). Table 3-5 presents standard thermocouple characteristics.

FIGURE 3-78 SCHEMATIC OF THERMOCOUPLE CIRCUIT**TABLE 3-5 STANDARD THERMOCOUPLE CHARACTERISTICS**

Type	Material	Operating Range	Accuracy
K	Chromel/Alumel	–200 to 1350	+/- 3°C
J	Iron/Constantan	–200 to 800	+/- 3°C
E	Chromel/Constantan	–200 to 1000	+/- 1.5°C
R	Platinum/Platinum Rhodium (10%)	–50 to 1600	+/- 2°C
S	Platinum/Platinum Rhodium (13%)	–50 to 1600	+/- 2°C
T	Copper/Constantan	–200 to 400	+/- 2°C

Chromel is an alloy of nickel and chromium, alumel is an alloy of nickel, aluminium is an alloy of nickel, and constantan is an alloy of copper. Thermocouple materials are divided into two categories: base metal types and rare metal types using platinum, rhodium, and iridium.

The general requirements for industrial thermal transducers are

- High output electromotive force.
- Resistance to the chemical changes when it comes in the contact with the fluids.
- Stability of voltage developed.
- Mechanical strength in their temperature range.
- Linearity characteristics.

The resultant emf of a particular transducer may be increased by multiplying the number of hot and reference junctions. If there are three measuring junctions, the emf is enhanced appropriately. If the thermocouples in this arrangement are at different temperatures, the resultant emf is a measure of the mean value.

Susceptibility to interference is an important consideration in any measurement application. Temperatures measured in hostile environments; in the presence of strong electrical, magnetic, or electromagnetic fields; or near high voltages are susceptible to interference. Susceptibility can be reduced by using non-contact methods of temperature detection.

3.8.3 Radiative Temperature Sensing

Bodies at any temperature emit radiation and absorb radiation from other bodies. A body at a temperature greater than 0°K radiates electromagnetic energy in an amount that depends on its temperature and physical properties. A sensor for thermal radiation need not be in contact with the surface

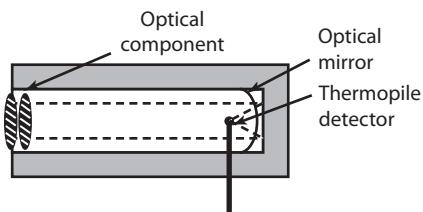
to be measured. Since the radiation emitted by an object is proportional to the fourth power of its temperature, the following relationship exists.

$$W = \sigma T^4 \quad (3-88)$$

Here W is the flux of energy radiated from an ideal surface and σ is the Stefan-Boltzmann constant.

Commercial radiation thermometers or radiometers vary in their complexity and accuracy. A schematic of a basic radiometer is shown in Figure 3-79 schematic of thermocouple circuit.

FIGURE 3-79 SCHEMATIC OF RADIATION THERMOMETER



The thermopile detector is subjected to radiation from a heat source whose temperature is to be detected. The resulting rise in temperature is recorded by measuring the thermoelectric power produced by a thermopile detector. A pyrometer is a device that measures the temperature of an object by measuring its radiated energy using an optical system. The radiation emitted by the object passes through the lens system and impacts the thermal sensor. The increase in temperature of the thermopile is a direct indication of the temperature of the radiation source.

An optical pyrometer identifies the temperature of a surface by the color of the radiation emitted by the surface. Other methods of temperature detection include optical fiber thermometers, acoustic temperature sensors, interferometric sensors, and thermochromic solution sensors.

3.8.4 Temperature Sensing Using Fiber Optics

Several concepts of temperature monitoring using fiber optics have been investigated. Operating principles based on intensity modulation in the optical fibers while under the influence of temperature has been discussed in the fiber-optic section of this chapter. In one type of reflective sensor, the displacement of a bimetallic element under the influence of temperature is measured providing an indication of temperature variation. In another type of sensor, an active sensing material (such as a liquid crystal) is used which produces fluorescence. The spectral response of the material as it is placed in the path of temperature is calibrated to produce a temperature output. The concept of micro bending is also used for temperature measurement. Using the thermal expansion of component structure, the sensor can measure the temperature by altering the fiber bend radius with temperature.

3.8.5 Temperature Sensing Using Interferometrics

Interferometric sensing is another method used for temperature measurement. It is based on the light intensity of interfering light beams. One is a reference beam, and the other, which travels through a temperature sensitive medium, is delayed. The length of the delay is a function of the temperature. The resulting phase shift between the two beams excites the interference signal.

Under extreme conditions temperature measurement may become a difficult task. Examples of such conditions include:

- Cryogenic temperature ranges such as high radiation levels inside nuclear reactors.
- Temperature measurement inside a sealed enclosure with a known medium, in which no contact sensors can be inserted and the enclosure is not transmissive for the infrared radiation.

In such unusual conditions, acoustic temperature sensors may be useful. The operating principle of this sensor is based on the relationship between temperature of the medium and the speed of sound.

SUMMARY

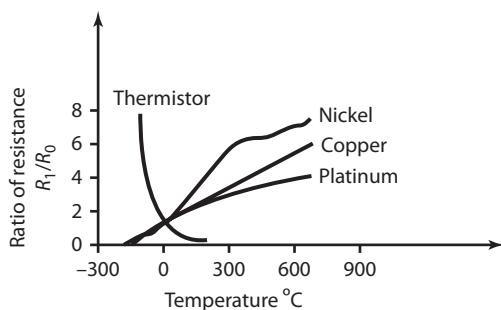
Temperature Sensors

RTD is a length of wire whose resistance is a function of temperature. It consists of a wire that is wound in the shape of a coil to achieve small size and improve thermal conductivity.

Thermistors

A thermistor is a transducer whose operation relies on a change in semiconductor resistance with change in temperature. Increase in temperature decreases electrical resistance by improving conductance. A Semiconductor becomes a better conductor of current as its temperature is increased. Individual thermistor curves (Figure 3-80) are approximated by the nonlinear equation,

FIGURE 3-80



$$\frac{1}{T} = A + B \ln R + C (\ln R)^3$$

where

T = temperature in kelvins

R = resistance of thermistor

A, B, C = curve fitting constants

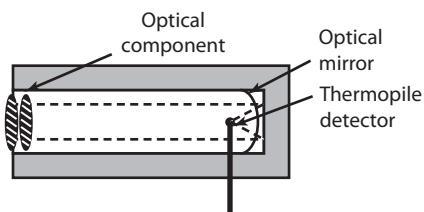
Radiative Temperature Sensing:

The radiation emitted by an object is proportional to the fourth power of its temperature,

$$W = \sigma T^4$$

where W is the flux of energy radiated from an ideal surface, and σ is the Stefan-Boltzman constant. The thermopile detector (Figure 3-81) is subjected to radiation from a heat source whose temperature is to be detected.

FIGURE 3-81 SCHEMATIC OF RADIATIVE THERMOMETERS



Pyrometers measure the temperature of an object by measuring its radiated energy using an optical system. The radiation emitted by the object passes through the lens system and impacts the thermal sensor. Increase in temperature of the thermopile is a direct indication of the temperature of the radiation source. An optical pyrometer identifies the temperature of a surface by the color of the radiation emitted by the surface.

Features

Since the thermistor is a bulk semiconductor, it can be fabricated in many forms, including discs, beads, and rods varying in size from a bead of one millimeter in diameter to a disc several centimeters in diameter and thickness.

Other methods include optical-fiber thermometers, acoustic sensors, interferometric sensors, and thermo-chromic solution sensors.

Applications

- The operating range of an RTD depends on the type of wire used as the active element.
- Platinum RTD has an operating range between -100 to 650°C ,
- Nickel RTD constructed from nickel has a range in the vicinity of -180°C to 300°C .
- Temperature range measured with a typical thermistor is between -250°C and 650°C .

3.9 Sensor Applications

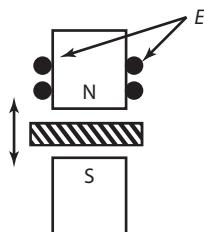
3.9.1 Eddy Current Transducer

Eddy current transducers are used to detect the presence of nonmagnetic but conductive materials. They are also used in nondestructive testing applications, including flaw inspections and location of defects.

Defects may include changes in composition, structure, and hardness, as well as cracks and voids. In addition to detecting the presence or absence of an object, eddy current transducers can be used to determine material thickness and non-conductive coating thickness. Depending on the application, eddy current transducers can vary in diameter from 2 to 30 mm. Direct contact with the specimen is not required, which makes it ideal for unattended continuous process monitoring.

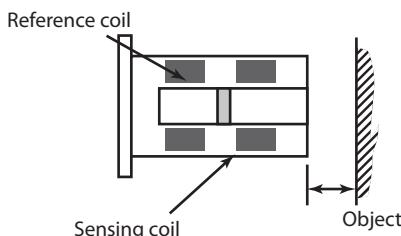
When a conducting material is placed in a changing magnetic field, an electromotive force (EMF) is induced in it. This EMF causes localized currents to flow, which are known as eddy currents. Eddy currents can be induced in any conductor but are most noticeable in solid conductors.

For example, when the magnetic core of a transformer or rotating machine is subjected to a change in magnetization, eddy currents are produced. Figure 3-82 shows the principle behind the eddy current transducer.

FIGURE 3-82 EDDY CURRENT PRINCIPLE

A nonferrous plate moves in a direction perpendicular to the lines of flux of a magnet. Eddy currents generated in the plate are proportional to the velocity of the plate. The eddy currents set up a magnetic field in a direction that opposes the magnetic field that creates them. The output voltage is proportional to the rate of change of eddy currents in the plate.

The eddy current sensor, shown in Figure 3-83, has two identical coils, one coil is used as a reference, and the second coil is used to sense the magnetic current in the conductive object.

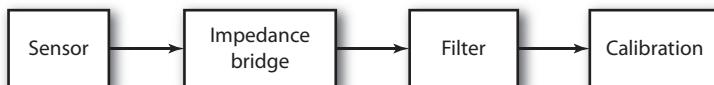
FIGURE 3-83 SENSING AND REFERENCE COILS IN AN EDDY CURRENT TRANSDUCER

Eddy currents produce a magnetic field which opposes that of the sensing coil, resulting in a reduction of flux. When the plate is nearer to the coil, the eddy currents as well as the change in magnetic impedance are both larger. The coils form two arms of an impedance bridge. The bridge has a supply frequency usually 1 MHz or higher. In the absence of a target object, the output of the impedance bridge is zero. As a target moves closer to the sensor, eddy currents are generated in the conducting medium because of radio frequency (RF) magnetic flux from the active coil. Inductance of the active coil increases, causing a voltage output in the bridge circuit.

Eddy current transducers are designed with shielded and unshielded configurations. The shielded transducer has a metal guard around the ferrite core and the coil assembly. This shielding focuses the electromagnetic field to the front of the transducer and allows the transducer to be installed in a metal structure without influencing the range of detection. The unshielded transducer can sense from its sides as well as its front.

The block representation of the signal processing in an eddy current transducer is shown in Figure 3-84. Using sensitive eddy current transducers, differential motions of .001 mm are easily detected. Eddy current transducers are attractive because of their low cost, small size, high reliability, and their effectiveness while operating at elevated temperatures.

FIGURE 3-84 SIGNAL PROCESSING IN EDDY CURRENT TRANSDUCERS

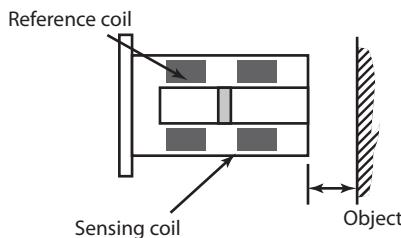


SUMMARY

Eddy Current Transducers

When a conducting material is placed in a changing magnetic field, an electromotive force (EMF) is induced in it. This EMF causes localized currents to flow and are called eddy currents. A nonferrous plate moves in a direction perpendicular to the lines of flux of a magnet. Eddy currents are generated in the plate that are proportional to the velocity of the plate. The output voltage is proportional to the rate of change of eddy currents in the plate.

FIGURE 3-85



Applications

- Eddy current transducers are used as proximity sensors.
- Used in non-destructive testing applications, including flaw inspections and defect location.
- Used to determine material thickness and non-conductive coating thickness.

Features

Direct contact with the specimen is not required which makes it ideal for unattended continuous process monitoring.

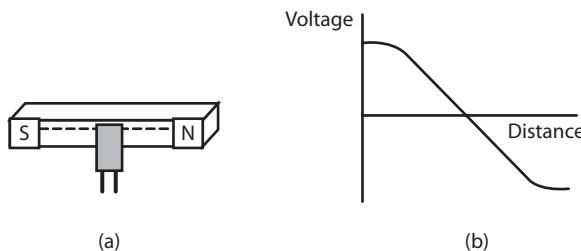
Hall Effect

The Hall effect is the generation of a transverse voltage in a conductor or semiconductor carrying current in a magnetic field. See Section 3.9.2 “Hall Effect” on the next page for a complete discussion of the Hall effect. The Hall effect results in the production of an electric field perpendicular to the directions of both the magnetic field and the current with a magnitude proportional to the product of the magnetic field strength, the current, and various properties of the conductor.

Position Sensing

As the magnet moves back and forth at that fixed gap (Figure 3-86), the magnetic field induced by the element becomes negative as it approaches the north pole and positive as it approaches the south pole.

FIGURE 3-86



Applications

- Hall effect sensors are used for proximity, level, and flow sensing applications.
- Devices based on the Hall effect include Hall-effect vane switches, Hall-effect current sensors, and Hall-effect magnetic-field strength sensors.

Features

- Hall effect sensors provide liquid-level measurement without any electrical connections inside the tank.
- Tend to be more expensive than inductance proximity sensors, but have better signal-to-noise ratios and are suitable for low speed operation.

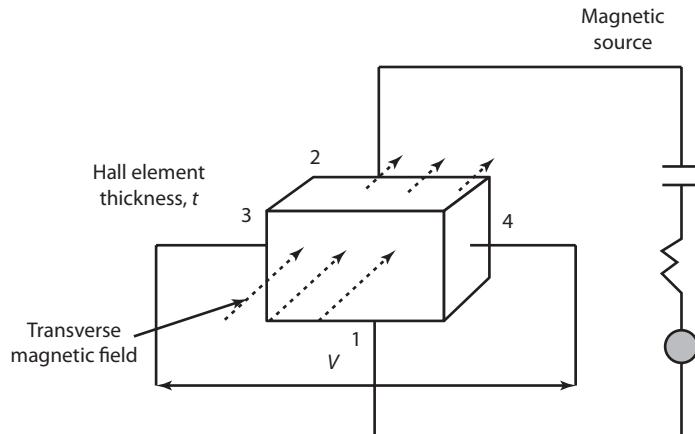
3.9.2 Hall Effect

Hall effect transducers are used to measure position, displacement, level, and flow. They can be used as an analog motion sensing device as well as a digital device. The Hall effect occurs when a strip of conducting material carries current in the presence of a transverse magnetic field, as shown in Figure 3-87. The Hall effect results in the production of an electric field perpendicular to the directions of both the magnetic field and the current with a magnitude proportional to the product of the magnetic field strength, the current, and various properties of the conductor. An electron of charge, e , traveling in a magnetic field, B , with a velocity v , experiences a Lorenz force F , and it is represented by

$$F = e(v \times B) \quad (3-89)$$

An electric field, known as Hall's field, counterbalances Lorenz's force and is represented by an electric potential. The voltage produced may be used to produce field strength or a current.

Figure 3-87 shows the Hall effect principle. Current is passed through leads 1 and 2 of the element. The output leads are connected to the element faces 3 and 4. These output ends are at the same potential when there is no transverse magnetic field passing through the element. When there is a magnetic flux passing through the element, a voltage V appears between output leads. This voltage

FIGURE 3-87 HALL EFFECT PRINCIPLE

is proportional to the current and the field strength. The output voltage is represented in terms of element thickness, the flux density of the field, the current through the element, and the Hall coefficient as

$$V = H \frac{IB}{t} \quad (3-90)$$

where

H = Hall coefficient, which can be defined as transverse electric potential gradient per unit magnetic field per unit current density. The units are V-m per A-Wb/m²

I = current through the element (A)

B = flux density of the field (Wb/m²)

t = thickness of the element (m)

The overall sensitivity of the transducer depends on the Hall coefficient. The Hall effect may be either negative or positive, depending on the material crystalline structure, and is present in metals and semiconductors in varying amounts based on the characteristics of the materials.

EXAMPLE 3.9 Flux Density Measurement Using a Hall Element

A Hall element with dimensions $4 \times 4 \times 2$ mm is used to measure flux density. The Hall coefficient (H) is -0.8 V-m per A-Wb/m². Find the voltage developed if the field strength is 0.012 Wb/m² and the current density is 0.003 A/mm².

Solution

$$\begin{aligned} \text{Current} &= \text{Current density} \times \text{area} \\ &= 0.003 \times 4 \times 4 = 0.0048 \text{ A} \end{aligned}$$

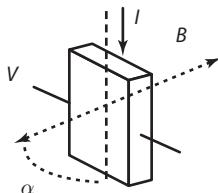
The voltage generated is

$$V = \frac{\text{HIB}}{t} = \frac{-0.8 \times 0.048 \times 0.012}{0.002}$$

$$V = 0.23 \text{ V}$$

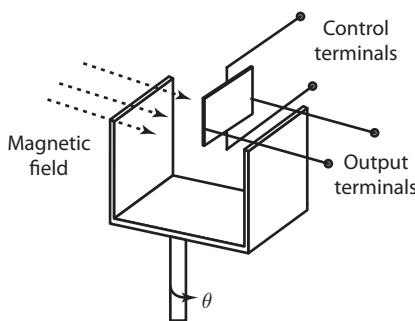
Rotational Measurement The basic operating principle of the Hall effect, which produces an output voltage proportional to a small rotary displacement, is shown in the Figure 3-88.

FIGURE 3-88 HALL ELEMENT FOR ANGULAR MEASUREMENT



The Hall sensor is suspended between the poles of a permanent magnet connected to the shaft, as shown in Figure 3-89. The probe is stationary, and the permanent magnet connected to the shaft rotates. With a constant control current applied to the electrical contacts at the end of the probe, the Hall voltage generated across the probe is directly proportional to the sine of the angular displacement of the shaft. Small rotations up to six degrees can be measured precisely with such probes. The main advantage of such devices is that they have nocontact, small size, and good resolution.

FIGURE 3-89 ROTATIONAL TRANSDUCER



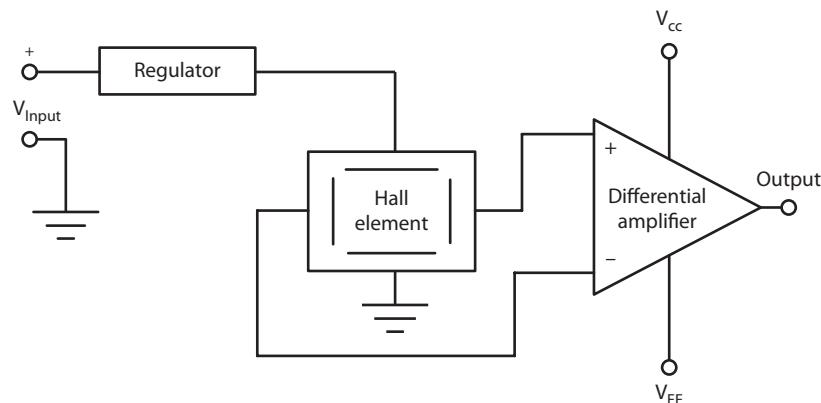
Output voltage generated for a rotation of α degrees is summarized as

$$V = \text{HIB} \frac{\sin \alpha}{t} \quad (3-91)$$

Here α is the angle between the magnetic field and the Hall plate.

Constructional Details of a Hall Effect Sensor The Hall element requires signal conditioning to make the output usable for most applications. The signal conditioning electronics needed are amplifier stage and temperature compensation. Voltage regulation is needed when operating from an unregulated supply. Figure 3-90 illustrates a basic Hall effect sensor. If the Hall voltage is measured when no magnetic field is present, the output is zero (Figure 3-87). However, if voltage at each output terminal is measured with respect to ground, a non-zero voltage will appear. This is the *common mode voltage* (CMV) and is the same at each output terminal. It is the potential difference that is zero. The amplifier shown in Figure 3-90 must be a differential amplifier in order to amplify only the potential difference (i.e., the Hall voltage).

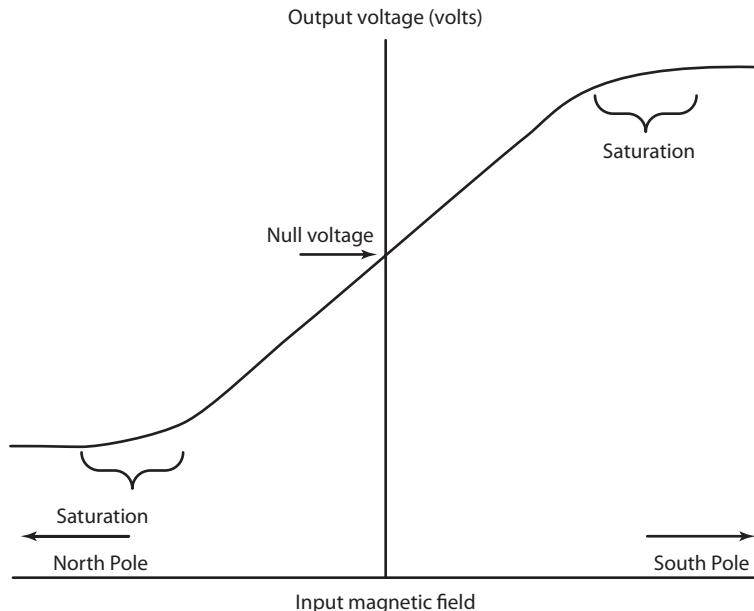
FIGURE 3-90 BASIC ANALOG OUTPUT HALL EFFECT SENSOR



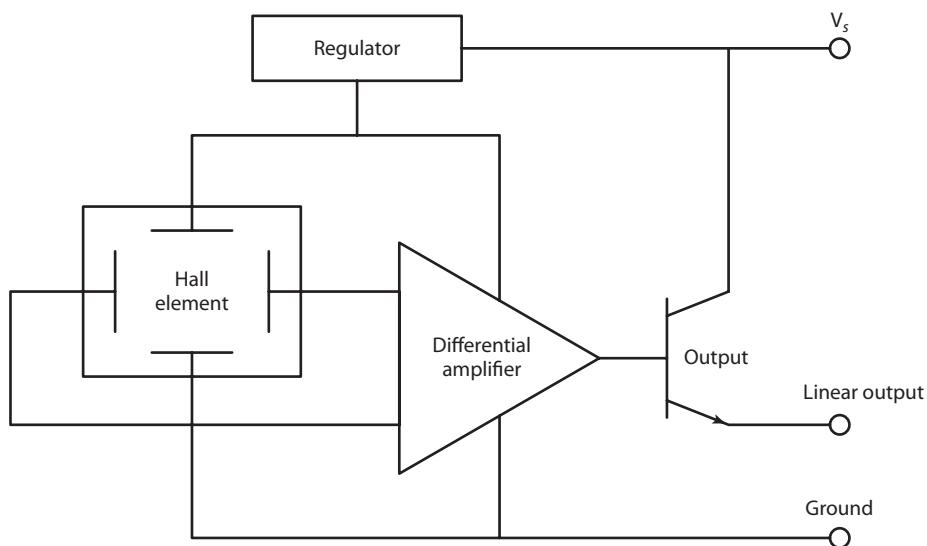
The Hall voltage is a low-level signal on the order of $30 \mu\text{V}$ in the presence of a one gauss magnetic field. This low-level output requires an amplifier with low noise, high input impedance, and moderate gain. A differential amplifier with these characteristics can be readily integrated with the Hall element using standard bipolar transistor technology. Temperature compensation is also easily integrated. As was shown by Equation 3-91, the Hall voltage is a function of the input current. The purpose of the regulator in Figure 3-90 is to hold this current constant so that the output of the sensor only reflects the intensity of the magnetic field. As many systems have a regulated supply available, some Hall effect sensors may not include an internal regulator.

Analog Output Sensors The sensor described in Figure 3-90 is a basic analog output device. Analog sensors provide an output voltage that is proportional to the magnetic field to which it is exposed. The sensed magnetic field can be either positive or negative. As a result, the output of the amplifier will be driven either positive or negative. Hence, a fixed offset or bias is introduced into the differential amplifier which appears on the output when no magnetic field is present and is referred to as a null voltage. When a positive magnetic field is sensed, the output increases above the null voltage. Conversely, when a negative magnetic field is sensed, the output decreases below the null voltage, but remains positive. This concept is illustrated in Figure 3-91.

Also, the output of the amplifier cannot exceed the limits imposed by the power supply. In fact, the amplifier will begin to saturate before the limits of the power supply are reached. This saturation is illustrated in Figure 3-91. It is important to note that this saturation takes place in the amplifier and not in the Hall element. Thus, large magnetic fields will not damage the Hall effect sensors, but rather drive

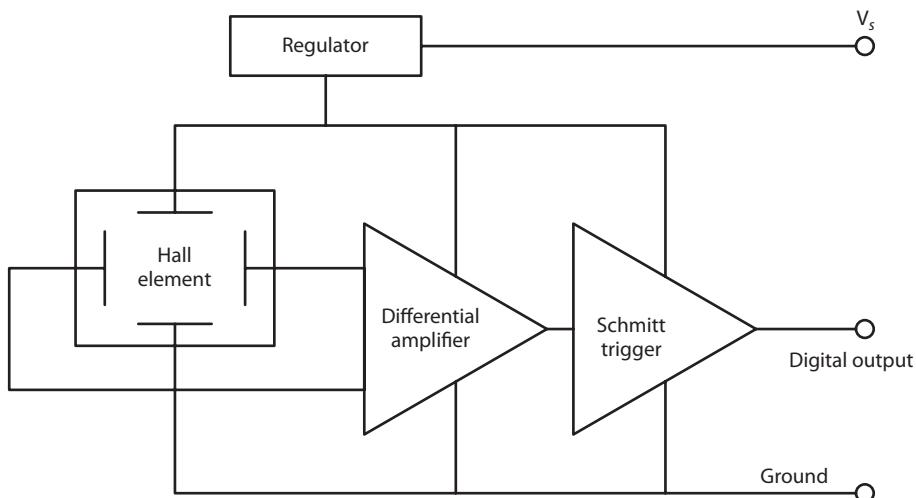
FIGURE 3-91 HALL EFFECT SENSOR'S CHARACTERISTIC CURVE

them into saturation. To further increase the interface flexibility of the device, an open emitter, open collector, or push-pull transistor is added to the output of the differential amplifier. Figure 3-92 shows a complete analog output Hall effect sensor incorporating all of the previously discussed circuit functions.

FIGURE 3-92 ANALOG OUTPUT HALL EFFECT SENSOR

Digital Output Sensors The digital Hall effect sensor has an output that is just one of two states: ON or OFF. The basic analog output device illustrated in figure 3-90 can be converted into a digital output sensor with the addition of a Schmitt trigger circuit. Figure 3-93 illustrates a typical internally regulated digital output Hall effect sensor. The Schmitt trigger compares the output of the differential amplifier with a preset reference. When the amplifier output exceeds the reference, the Schmitt trigger turns on. Conversely, when the output of the amplifier falls below the reference point, the output of the Schmitt trigger turns off.

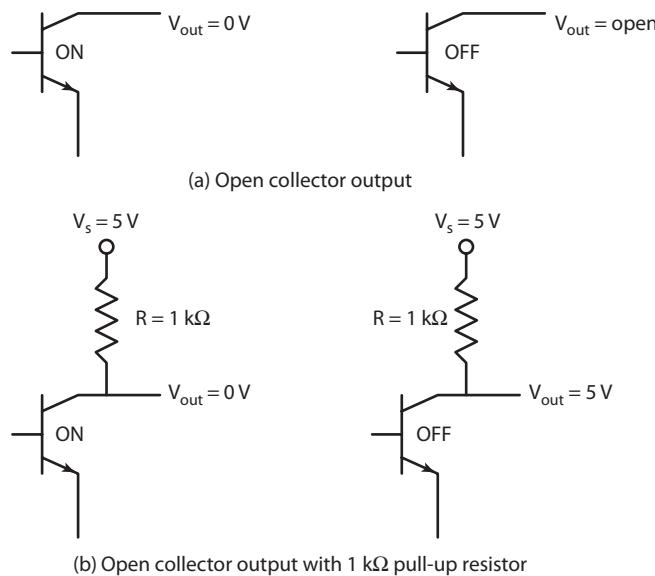
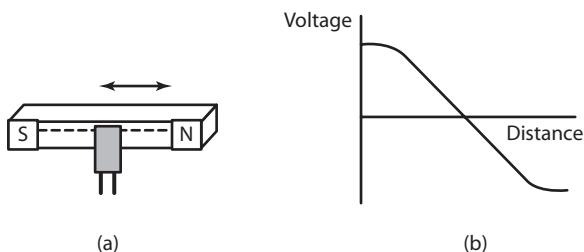
FIGURE 3-93 DIGITAL OUTPUT HALL EFFECT SENSOR



Open-Collector Output and Pull-Up Resistor (Ref. 2) A Hall effect encoder with open-collector output either drives the output LOW or lets it float. Hence, to drive logic HIGH with an open-collector output, we should add an external resistor, called a pull-up resistor, as shown in Figure 3-94(b).

Applications of Hall Effect Transducers Hall effect transducers are widely used as proximity sensors, limit switches, liquid level measurement, and flow measurement. They are also used for sensing deflections in biomedical implants. Hall effect transducers are constructed in various configurations depending on the application. Hall effect principle is used to make devices such as, Hall-effect vane switches, Hall-effect current sensors, and Hall-effect magnetic field strength sensors. Hall effect sensors tend to be more expensive than inductance proximity sensors but have better signal-to-noise ratios and are suitable for low-speed operation.

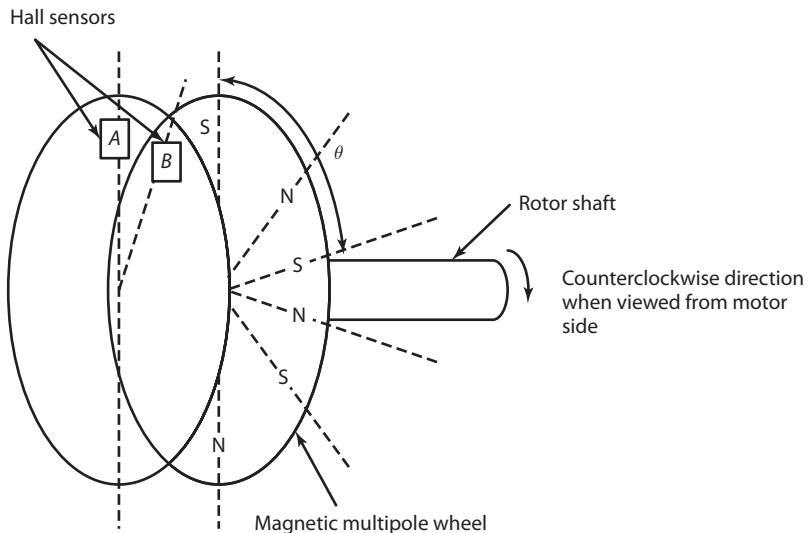
Position Sensing Figure 3-95(a) shows a schematic of a Hall effect sensor used for sensing sliding motion. A tightly controlled gap is maintained between the magnet and the hall element. As the magnet moves back and forth at that fixed gap, the magnetic field induced by the element becomes negative as it approaches the North Pole and positive as it approaches the South Pole. This type of position sensor features mechanical simplicity, and when used with a large magnet, it can detect position over a long magnet travel.

FIGURE 3-94 OPEN COLLECTOR OUTPUT WITH AND WITHOUT PULL-UP RESISTOR**FIGURE 3-95 (A) SLIDING SENSOR (B) OUTPUT CHARACTERISTICS**

The output characteristic of the sensor has a fairly large linear range, as shown in Figure 3-95(b). It is necessary to maintain rigidity in linear motion and prevent any orthogonal movements of the magnet when the sensor is used for measuring sliding motion.

Method for Measuring the Angular Position of a Motor Shaft Figure 3-96 shows the setup of using Hall effect sensor along with a permanent magnet multi-pole wheel for measuring the position, and Figure 3-97 shows the constructional details of a motor with one such inbuilt Hall effect encoder (sensor). As seen in Figure 3-96, there are two Hall sensors, A and B, which are required to measure the position and the direction of rotation of the rotor shaft.

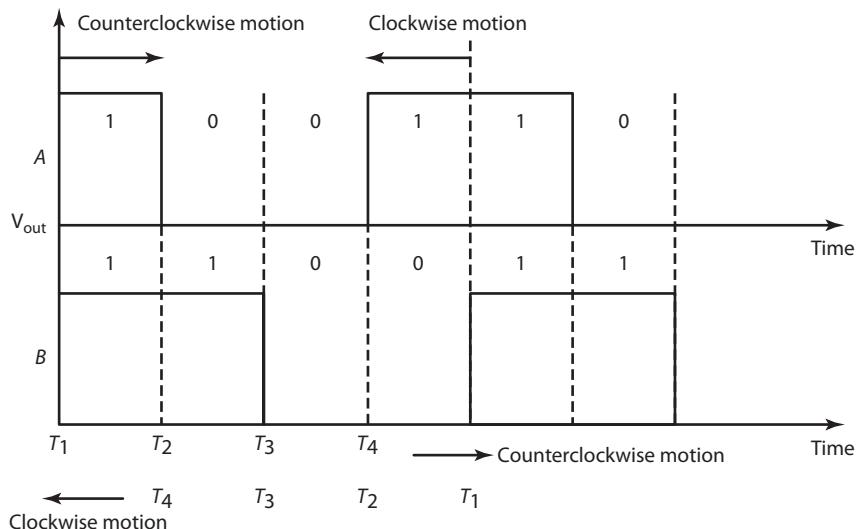
We know that, when the South Pole comes in front of the Hall element, a positive voltage is developed and the trigger is turned ON. With the North Pole, a negative voltage (or zero voltage with the bias in the differential amplifier) is developed and the trigger is turned OFF. With the current position of the poles on the wheel and the sensors, as shown in the Figure 3-96, if the rotor rotates by an angle θ in counterclockwise direction when viewed from the motor side,

FIGURE 3-96 HALL SENSORS AND MAGNETIC WHEEL SETUP**FIGURE 3-97 CONSTRUCTIONAL DETAILS OF A MOTOR WITH INBUILT HALL SENSOR (REF. 3)**

www.walab.ctw.utwente.nl/Lectures/110325/DataSheets/MaxonEncoderinfo.pdf.

the output signals from the digital output Hall sensors A and B will be of the form represented in Figure 3-98.

As seen from Figure 3-98 there is a 90° phase difference between the output signals; hence, these sensors are also known as quadrature encoders. The ON (1) and OFF (0) states of the output signals from A and B are used to create the logic for measuring the position as well as the direction of the motor. Figure 3-99 shows the tabular representation of these states for 1 pulse (i.e., for the

FIGURE 3-98 HALL SENSORS OUTPUT SIGNAL**FIGURE 3-99 HALL SENSORS OUTPUT STATES CHART**

Time	T_1	State of A is changing	T_2		T_3	State of A is changing	T_4
A	1		0		0		1
Time	T_2		T_3		T_4		
A	0		0		1		
Time	T_1		T_2	State of B is changing	T_3		T_4
B	1		1		0		0
Time	T_2		T_3		T_4		
B	1		0		0		

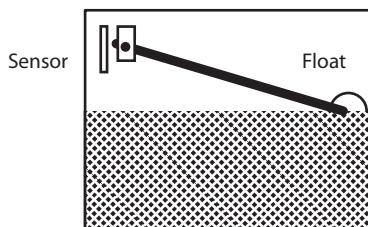
rotation of the rotor shaft in counterclockwise direction by an angle θ) for the setup shown in Figure 3-96. Also, it would be important to know here that if we have n -pole wheel, we get $n/2$ pulses for every revolution of the rotor shaft. With a quadrature encoder, we get 4 counts for every pulse. From Figure 3-99, if we compare the state of A with the previous state of A and the state of B with the previous state of B, we find that if the state of A or state of B is changing, we have to increment the count by 1 if it is moving in the same direction or decrement it by 1 if it is moving in the opposite direction. The decision for incrementing or decrementing can be made if we compare the state of A with previous state of B, as shown in Figure 3-100 for both counterclockwise and clockwise movement of the rotor shaft.

Considering counterclockwise direction of the motor to be positive, we would need to increment the count by 1 if the state of A is different from the previous state of B and decrement the count by 1 if the state of A is same as the previous state of B. Based on the discussion, a logic was developed to count the rotation of the motor shaft which is discussed further in Chapter 7.

FIGURE 3-100 COMPARISON CHART OF SENSOR A STATE WITH PREVIOUS STATE OF SENSOR B

Counter-clockwise direction	Time	T_1	The two states are different	T_2	The two states are different	T_3	The two states are different	T_4
	B	1		1		0		1
	Time	T_2		T_3		T_4		
	A	0		0		1		
Clockwise direction	Time	T_1	The two states are same	T_2	The two states are same	T_3	The two states are same	T_4
	B	0		0		1		1
	Time	T_2		T_3		T_4		
	A	0		0		1		

Liquid Level Measurement Determining the height of a float is one method of measuring the level of liquid in a tank. Figure 3-101 illustrates an arrangement of a Hall element and a float in a tank made of non-ferrous material (e.g., aluminum).

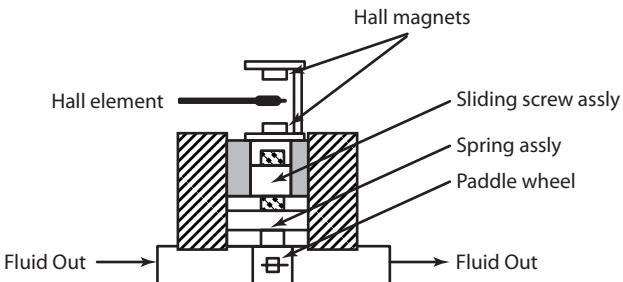
FIGURE 3-101 LIQUID LEVEL BY HALL EFFECT

As the liquid level goes down, the magnet moves closer to the sensor, causing an increase in output voltage. This system provides liquid level measurement without any electrical connections inside the tank.

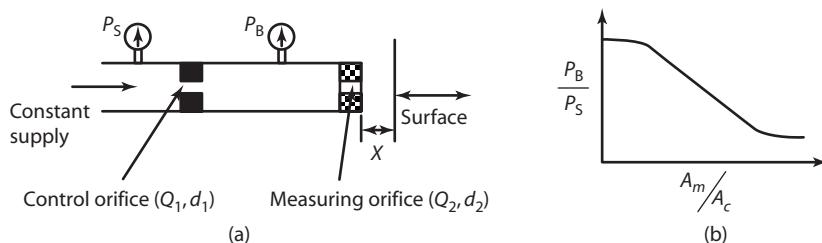
Flow Measurement Figure 3-102 shows how a Hall element is used for flow measurement. The chamber has fluid-in and fluid-out provisions. As the flow rate through the chamber increases, a spring-loaded paddle turns a threaded shaft. As the shaft turns, it raises a magnetic assembly that energizes the transducer. When the flow rate decreases, the coil spring causes the assembly to go down which reduces the transducer output. The design of the magnetic assembly and sliding screw-nut assembly is calibrated to provide a linear relationship between the measured voltage and the flow rate. Figure 3-103 presents a photograph of a typical Hall effect flow sensor.

3.9.3 Pneumatic Transducers

Pneumatic transducers are non-electrical in nature and widely used in industrial instrumentation for measurement and gauging applications. Pneumatic systems use air as a medium for transmitting signal and power. They are sensitive, simple to design, and sensitive in operation. Pneumatic transducers used for displacement convert changes in length or surface displacement into changes in pressure value. A schematic diagram of a pneumatic transducer is shown in Figure 3-104.

FIGURE 3-102 FLUID FLOW MEASUREMENT**FIGURE 3-103 HALL EFFECT FLOW SENSOR**

Courtesy of Gem Sensors, Inc. Plainville, CT.

FIGURE 3-104 PRINCIPLE OF PNEUMATIC BACK PRESSURE SENSORS

Typically, there are two chambers arranged in series and separated by an orifice. Air passes from the first to the second chamber-control orifice and to the atmosphere via the second orifice (the measuring orifice). The transducer shown has two orifices, Q_1 and Q_2 . Orifice Q_1 is called the control orifice. It has a diameter, d_1 , and effective area, A_c . The second orifice, Q_2 , is called the measuring orifice. It has a diameter, d_2 . Its effective area, A_m , is variable and depends upon the distance x , which is the displacement of the workpiece.

$$A_c = \frac{\pi}{4} d_1^2 \quad (3-92)$$

$$A_m = \frac{\pi}{4} d_2 x$$

Variation in the backpressure, P_b , can be caused by moving a resistive surface towards or away from the orifice Q_2 . Experimental results have shown that there exists a linear relationship between P_b and x over a limited range of x . Empirical results have shown that, for supply pressure between 15 kN/m² and 500 kN/m², the variation of P_b/P_s and A_m/A_e is as shown in Figure 3-104(b). The curve has a linear range P_b/P_s extending from 0.6 to 0.9. The extension to the linear part cuts the P_b/P_s axis at 1.1. The slope varies slightly, reducing with increasing supply pressure. For linear range, the relationship may be expressed as

$$\frac{P_b}{P_s} = K \frac{A_m}{A_e} + b \quad \text{for } 0.6 < \frac{P_b}{P_s} < 0.9 \quad (3-93)$$

Here $b = 1.1$ and K = slope of the curve. The backpressure P_b is measured by a pressure gauge. Overall sensitivity is given by the rate of change of output with respect to the input. If the output variable has a pressure gauge reading of ΔR , and the input variable has a surface displacement of ΔX .

The overall magnification is $\frac{\Delta R}{\Delta X}$, and the overall sensitivity is dependent on the sensitivity of the measuring head, orifice size, and the supply pressure. The measuring head sensitivity is computed as $A_m = \pi d_2 X$. Differentiating with respect to A_m , $\frac{dA_m}{dx} = \pi d_2$ reveals that the measuring head sensitivity increases with an increase in orifice size.

The overall sensitivity of the pneumatic transducer is a measure of the gauge displacement for any input change in displacement. This factor is sensitive to variations in the measuring orifice, changes in the backpressure, and also to the sensitivity of display gauges.

In addition to displacement measurement, pneumatic transducers are used in gauging applications where it is difficult to use electronic gauges because of the design limitations of high temperature, humidity, and contamination.

Figure 3-105 shows a typical plug gauge, which inspects the internal diameter within the specified limits. Figure 3-106 shows the ring gauge used for inspection of the external diameters. Figure 3-107 illustrates the principle of taper measurement and Figure 3-108 shows the principle of measurement of the straightness of precision cylindrical bores.

FIGURE 3-105 PNEUMATIC PLUG GAUGES

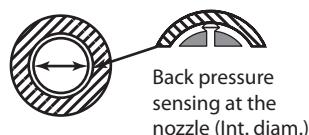


FIGURE 3-106 PNEUMATIC RING GAUGES

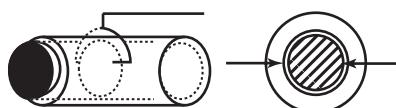
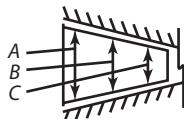


FIGURE 3-107 PNEUMATIC TAPER GAUGES**FIGURE 3-108 PNEUMATIC BORE GAUGES**

3.9.4 Ultrasonic Sensors

Ultrasonic sensors are used mainly in the areas of inspection and testing, especially for non-destructive testing. Ultrasonic waves have frequencies higher than the audible frequency of 20 kHz. The penetrative quality of ultrasonic waves makes them useful for noninvasive measurements in environments (such as radioactive, explosive, and areas which are difficult to access). They are used for distance, level, speed sensing, medical imaging devices, dimensional gauging, and robotics applications.

The ultrasonic transducer emits a pulse of an ultrasonic wave and then receives the echo from the object targeted. The ultrasonic transducer consists of a transmitter, a receiver, and a processing unit. The transducer produces ultrasonic waves normally in the frequency range of 30 to 100 kHz. Whenever an ultrasonic beam is incident on a surface, one portion of the incident beam is absorbed by the medium, another portion is reflected, and a third portion is transmitted through the medium. In proximity sensing applications, the ultrasonic beam is projected on the target, and the time it takes for the beam to echo from the surface is measured. For non-contact distance measurements, an active sensor transmits a signal and receives the reflected signal.

If there is a relative movement between the source and the reflector, the Doppler effect, discussed earlier in this chapter, is employed. Using the Doppler method, it is also possible to precisely measure the position, velocity, and fluid flow.

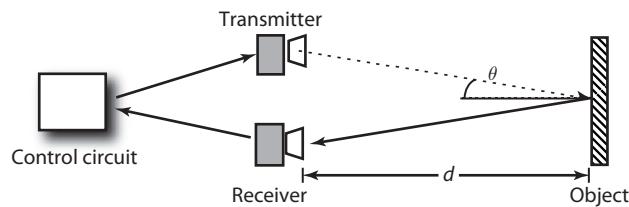
Ultrasonic automotive vehicle detection systems are based on two techniques: *pulse technique* and *Doppler shift technique*. In the pulse technique, the detector measures the time, Δt , spent between transmission and reception of an ultrasonic signal to determine the distance between transmit/receiver and the object. Using the Doppler technique, the frequency of the received ultrasonic signal changes in relation to the emitted frequency depending on the velocity, v , of the object. If the object is approaching the detector, then the frequency of the signal received increases in relation to the emitted frequency. It is reduced when the object is moving away from the detector.

Ultrasonic waves can be generated by the movement of a surface which creates compression and expansion of the medium. Transducers, such as piezoelectric transducers, are the excitation devices most commonly used for surface movement. As discussed in piezoelectric section, when an input voltage is applied to a piezoelectric element, it causes the element to flex and generate ultrasonic waves. This effect is reversible. Conversely, the element generates a voltage whenever it is subjected to vibrations such as the incoming ultrasonic waves. The typical operating frequency of the transmitting ultrasonic element is close to 32 kHz. If the ultrasonic instrument operates in the pulsed mode, then the same piezoelectric crystals are utilized for transmitting and receiving purposes.

Ultrasonic Distance Sensing The Figure 3-109 presents a range sensing system. In this figure, d is the distance to the object, v is the speed of the ultrasonic waves in the measured medium, θ is the incident angle, and t is the time for the ultrasonic waves to travel to the object and back to the receiver. Using these definitions the following equation is written,

$$\text{Distance: } d = \frac{vt \cos \theta}{2} \quad (3-94)$$

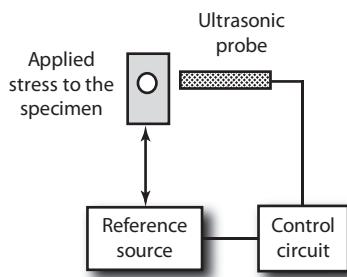
FIGURE 3-109 ULTRASONIC DISTANCE SENSING



The accuracy of the ultrasonic transducer is high and often in the order of one percent of the range measured. The sensors are used in robotics applications, where the robot manipulators need to avoid collisions and sense the distance of the object or obstruction in the vicinity of robot workspace. Some robots are provided with an ultrasonic ranging system that assists the robot in positioning the gripper relatively close to the part. This system often functions in combination with another optical proximity sensor that assists in the precise positioning.

Ultrasonic Stress Sensing Ultrasonic beams may be used for stress measurement. Figure 3-110 presents a typical stress measurement system employing ultrasonic beams.

FIGURE 3-110 ULTRASONIC STRESS SENSING



The system consists of an ultrasonic probe which is placed in contact with the specimen. The ultrasonic probe consists of an ultrasonic driver, receiver, and a control device to change the electrical signal to vibrations and vice versa. When in contact with the specimen, the ultrasonic transmitter causes waves to travel across the specimen. These waves are then received by the receiver and converted to an electrical signal.

The basic operating principle relies on changes in the propagation of sound in a specimen causing stress changes. The probe is moved around the specimen to *map* out the stress field distribution in various sections of the specimen. By rotating the probe, it is possible to determine the direction of the stress.

Ultrasonic Flow Sensing The transducer that is based on this principle has been explained in the 3.7.4 section of this chapter.

3.9.5 Range Sensors

Range sensing techniques are of special importance in manufacturing automation applications. Range sensors have been successfully employed in other areas as well, including the following.

- Automatic guidance systems for vehicles
- Robot navigation
- Collision avoidance

For example, consider an industrial scanning and recognition operation in which a sensory robot must locate objects in a container, not knowing exactly where they are. The robot has to follow the sequence of operations which could consist of the following.

1. Scanning a bin containing objects and locating the object in a three-dimensional space.
2. Determining the relative position and orientation of the object.
3. Moving the robot manipulator to the object location.
4. Positioning and orienting the robot gripper according to the objects location and layout.
5. Picking up the object and placing the object at the required location.

In a stationary robot, the gripper must be oriented to the object position. In addition, it must also have the capability of sensing the distance. In automated guided vehicle applications, the vehicle must navigate its body to the object location and then move its workholding device to grasp the object. Range sensors are typically located on the wrist of the robot manipulator. In some cases sensors also act as safety devices. Besides locating an object in a work cell, sensors are positioned to determine the human obstruction in the robot work cell.

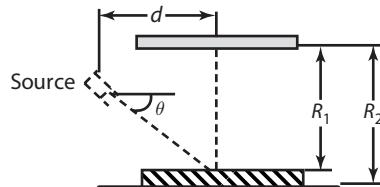
Distance sensors are also used for three dimensional shape inspection. A specimen or a machine part can be inspected on the production floor using an inspection machine such as a *coordinate measuring machine* (cmc). By finding the distance of the object from a fixed location to various points on the object, it is possible to digitize the three dimensional shape of an object into discrete points.

Distance sensors used for workpiece inspection are also known as digitizers in the machine tool industry. Digitizers are normally used in machine tools, robots, and inspection devices to locate the position of objects and to identify the geometry of the objects in a three dimensional environment. Some of these sensors also can be used as proximity devices. Proximity devices are used to give an indication of the closeness of one object to another object. A number of techniques are employed in range sensors including optical methods; acoustic, inductive, and electrical field techniques (e.g., eddy current, Hall effect, magnetic field); and others.

Range Sensing Principles The following section explains various methodologies used for range measurement. Although the focus of this section is on optical techniques, the same principle is applicable to non-optical methods.

The basic triangulation principle is the method of triangulation which applies trigonometric principles to determine the distance of an object from two previously known positions. Figure 3-111 illustrates the principle in a thickness-measuring application.

FIGURE 3-111 TRIANGULATION PRINCIPLE TO MEASURE THICKNESS ($R_2 - R_1$)



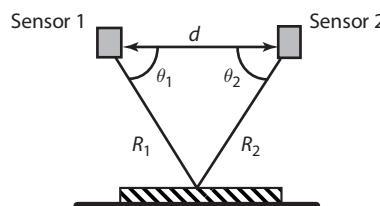
The source, typically a laser source, is focused on the surface of the object. A photodetector is used to determine the location of the spot. The distance, R_2 , and angle, θ , are known. Because the photo detector is located at a fixed distance in the work environment, the thickness of the part is calculated as

$$t = R_2 - R_1 = R_2 - d \tan \theta \quad (3-95)$$

Here d is found from the position of the light spot on the workpiece.

If two triangulation sensors are positioned a certain distance apart and both devices can align to a spot on an object, as shown in Figure 3-112, then the two devices and the object form a triangle. The distance, d , and two angles, θ_1 and θ_2 , are known. The third angle is found by subtracting the two known angles from 180°.

FIGURE 3-112 TRIANGULATION PRINCIPLE WITH TWO SENSORS



The distance from each device to the object can then be found by using the law of sines.

$$R_1 = \frac{d \sin \theta_2}{\sin [180 - (\theta_1 + \theta_2)]} \quad (3-96)$$

$$R_2 = \frac{d \sin \theta_1}{\sin [180 - (\theta_1 + \theta_2)]}$$

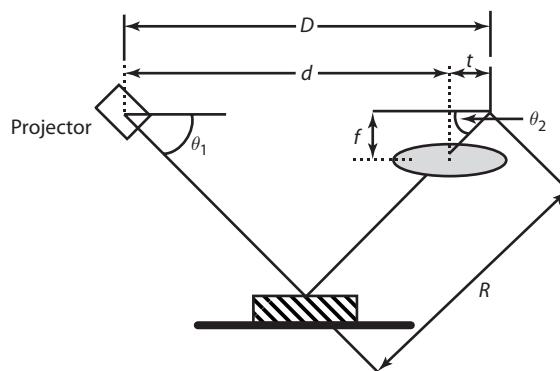
Instrumental techniques using triangulation principles include the following six methods.

1. Spot sensing method
2. Light strip sensing method
3. Camera motion method
4. Time of flight technique
5. Binocular vision technique
6. Optical ranging using position sensitive detectors

Range Sensing by Spot Projection Consider the situation in which a single imaging device is kept stationary and a projected light source scans the scene. If a single beam of light is projected onto an object, as shown in Figure 3-113, the projected beam creates a light spot on the object that is reflected into sensing device, such as a camera, which is positioned at a known distance, d , from the spot projector. This produces a triangle between the projector, object, and camera. The range, R , is calculated using the triangle, which provides the distance of the object spot from the camera. The reflected light spot produces an image point, B , in the camera image. This image point is easily detected, as a bright “spot” in the image. The distance of the image point from the center of the camera image can be determined. Furthermore, the camera focal length f , is fixed. Since the focal length, f , and the image point distance, t , form the sides of a right triangle, the angle θ_2 can be calculated as

$$\theta_2 = \tan^{-1} \frac{f}{t} \quad (3-97)$$

FIGURE 3-113 RANGE SENSING USING LASER SPOT PROJECTOR



From this, D , the distance between the projector and the image point can be calculated as

$$D = d + t$$

Where d is the distance between the projector and camera.

Depending on whether the image point is to the right (+) or left (−) of the center of the camera lens, t can be positive or negative. The angle the projector makes, θ_1 , is known and from this information the range, R , can be calculated using the law of sines as shown in Equation 3-98.

$$\frac{R}{\sin \theta_1} = \frac{D}{\sin [180^\circ - (\theta_1 + \theta_2)]} \quad (3-98)$$

$$R = \frac{D \sin \theta_1}{\sin [180^\circ - (\theta_1 + \theta_2)]}$$

Range is the distance between the image point and the object point. To calculate the range from the camera lens, subtract the distance between the lens and the image point.

Digitization of an object is performed if the light spot can scan over the entire scene and the range calculation can be computed at each point in the scan. In three-dimensional digitizers, a light spot scans the scene from right to left and top to bottom, utilizing a rotating mirror, which can traverse the beam in a three-dimensional area.

Sensing by the Use of Light Stripe The basic principle used in the light stripe method is an extension of spot sensing technique. Instead of projecting a spot of light, a stripe of light is projected on the scene. The imaging device creates a line of certain length. The image of the line is divided into individual image points, and the range is calculated for each point along the stripe. The range calculation is similar to that for spot sensing.

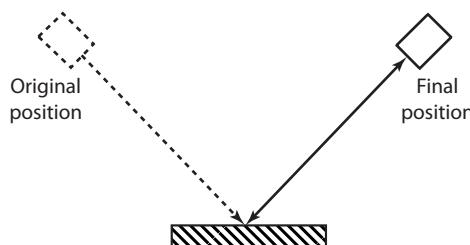
The light stripe can be formed by passing ambient or infrared light through a slit on the projector. The scene is scanned in a direction perpendicular to the stripe, resulting in a complete range mapping of the scene.

One limitation of light-stripe scanning is the poor depth resolution that is obtained for object surfaces that are parallel to the light stripe. It can be overcome by scanning the image in two directions, one perpendicular to the other.

The benefit of light striping is that it is relatively simple and fast, as opposed to spot sensing. The object boundaries and regions can be determined by connecting the end points of the light-stripe images. Thus, light striping aids in the image-segmentation process. This can be seen by examining the series of light stripe images.

Camera Motion Method Another method used in the area of active triangulation is called the camera motion method. It involves moving the camera as illustrated in Figure 3-114.

FIGURE 3-114 ACTIVE TRIANGULATION USING MOVING CAMERA



Here a single camera is moved a given distance to produce two stereo images of the scene. In an effect analogous to a stereo system, a single moving camera replaces two stationary cameras. Once the two images are obtained, the range calculations are made using the principle of disparity between the two images, as in stereovision.

Time-of-Flight Ranging Method Time-of-flight, TOF, ranging involves calculating the time required for a signal to reach and return from an object. Since distance equals the product of velocity and time, the range of an object can be written as

$$R = \frac{vt}{2} \quad (3-99)$$

Here, R is the range from the ranging device, v is the velocity of the transmitted signal, and t is the time required for the signal to reach and return from the object.

Time-of-flight ranging has been used for optics, sound, and electromagnetic sources. The determination of range, using Equation 3-99, is the same for each type of signal; however, each type of signal has its own characteristics that affect the accuracy of the range data. Two significant features of the time of flight method are (1) beam width and (2) speed of the signal.

The width of the signal beam determines the amount of detail that can be recovered during the ranging process. A wide signal beam does not produce accurate range data for small object details, because it covers a larger area than a narrower beam. Narrow beams result in higher object resolution. The faster the signal reaches and returns from an object, the more difficult it is to determine its range.

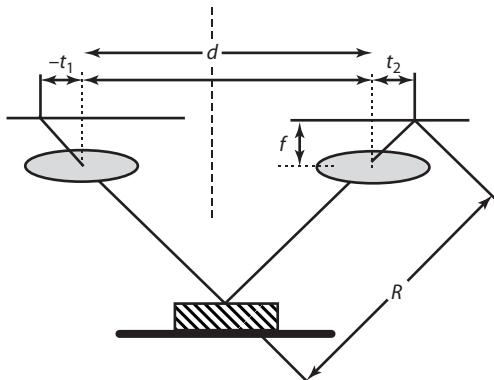
Range Sensing By Binocular Vision Binocular or stereovision, also known as passive triangulation, is analogous to human vision and sensing in terms of depth perception. Two imaging devices are placed a known distance apart. In a machine vision system, the imaging devices are usually diode-matrix or CCD cameras. Two parameters in the system are known: the distance between the cameras, d , and the focal length of the cameras. To calculate the range, R , from the cameras to a given point, P , on the object, both cameras scan the scene and generate a picture matrix. Given any point in the scene, such as point, P , there will be two pixels representing that point. One pixel is in the left camera image and the other is in the right camera image.

Each pixel is located at a given distance from the center of its image. Let t_1 be the distance that the left-camera image pixel is located from the center of its image. Let t_2 be the distance that the right-camera image pixel is located from the center of its image. If the two camera images are overlapped, the two image points, t_1 , and t_2 , will not coincide. There will be a certain distance between them.

This distance is calculated by taking the absolute value of their difference. The resulting difference is called the disparity between the two image points. The range, R , from the cameras to the object point is inversely proportional to the disparity between the values of t . As the disparity approaches zero, the range becomes infinite. Conversely, the range gets smaller as the disparity gets larger. As an example, consider the stereo system presented in Figure 3-115.

The range of any point on the object can be approximated by

$$R = \frac{d\sqrt{f^2 + t_1^2 + t_2^2}}{|t_1 - t_2|} \quad (3-100)$$

FIGURE 3-115 RANGE SENSING USING BINOCULAR VISION

where

- R = range from the left camera lens if the object point is in the right side of the scene
- = range from the right camera lens if the object point is in the left side of the scene
- = range from either camera if the object point is directly in the middle of the scene
- d = distance between camera lens centers
- f = focal length of cameras
- t_1 = distance of the image pixel from the center of the left camera lens
- t_2 = distance of the image pixel from the center of the right camera lens

The range value, R , can be the distance from the left, right, or either camera, depending on where the object point is located in the scene. If the object point is located in the right half of the scene, R is defined as the range from the left camera lens. The left and right halves of the scene are divided by an imaginary line located exactly halfway between the two cameras. Also, the individual values of t_1 and t_2 can be positive or negative, depending on the location of a given image pixel relative to the center of its respective image.

For example, if the image were between the two cameras, t_1 would be negative and t_2 positive. Note, however, that the disparity is always the absolute value of the difference between the two image points and is used in the denominator of the range equation. Hence, the position of these two points must be precisely determined.

Ideally, it would be nice to find individual pixels in one camera image that matched those of a second camera image. However, in reality, one cannot guarantee that two pixels with the same gray scale or color values were produced by the same object point. Stereo vision systems often search for similar edge or region features between two images to locate corresponding pixels. Edge-based stereo systems attempt to match stereo images by detecting intensity or color in edge mapping. Another matching technique is to take a pixel window from one image and pass it over the same general region of the second image until the best match is found. A displacement or disparity value is determined on the basis of how much the window must be displaced from the first image to match the second image. This value is then used to calculate the range.

Optical Ranging Using Position Sensitive Detectors Optical principles are widely used for precision position measurement. *Position sensitive detectors* (PSD) based on optical sources have been effectively used in photographic devices. These devices consist of a small light source and position sensitive detector. The light emitting diode and collimating lens transmit a pulse in the form of a narrow beam. After striking the object, the beam is reflected back to the detector. The received intensity is focused on the position sensitive detector. For example, let the beam be incident at a distance, t , from the center. The detector generates the output current I_1 and I_2 , which is proportional to the distance t of the light spot on its surface from the center.

The sensor consists of a silicon device and provides position signals on a light spot traveling over its surface. The photoelectric current produced at each terminal is proportional to the resistance between the electrode and the point of incidence. If I is the total current produced by the light spot and I_1 is the current at one of the output electrodes, the current produced at each terminal is proportional to the corresponding resistances and the distance between incidence and electrode. We replace the resistance's with distances as

$$I_1 = I \frac{(D - t)}{D}; I_2 = I \frac{t}{D} \quad (3-101)$$

where D is the distance between I_1 and I_2 . The ratio of currents is expressed as

$$Q = \frac{I_1}{I_2} = \frac{D}{t} - 1 \quad (3-102)$$

Solving for t yields,

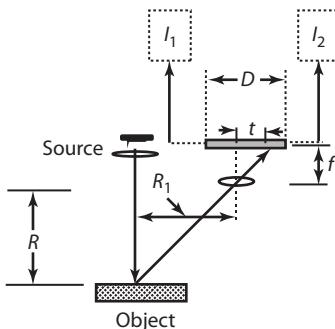
$$t = \frac{D}{Q + 1} \quad (3-103)$$

Using two triangles, the value of R is calculated as

$$\begin{aligned} R &= f \frac{R_1}{t} \\ R &= f \frac{R_1}{D} (Q + 1) \end{aligned} \quad (3-104)$$

Figure 3-116 shows the relationship between the focal length of the lens, f , the range, R , and various distances, R_1 and D . R can be calculated as shown in Equation 3-116.

FIGURE 3-116 TRIANGULATION PRINCIPLE APPLIED TO POSITION-SENSITIVE DETECTOR



Other Ranging Techniques The challenge in ultrasonic ranging is the difficulty in concentrating the sound energy into the narrow beam required to produce high object resolution for three-dimensional vision. Ultrasonic ranging is useful in robot navigation to detect the presence and range objects. Electromagnetic range sensing involves the use of radio frequency signals and is normally called radar. Radar has become useful in general, industrial, and military applications. The radio signal is transmitted into the atmosphere. The signal is reflected back from the object, and the distance or range to the object is determined using the time-of-flight relationship. Radar systems are efficient to measure the range of highly reflective metallic objects over relatively long distances but not useful for measuring relatively short distances of nonmetallic objects. Accurate depth measurement is difficult over short distances.

SUMMARY

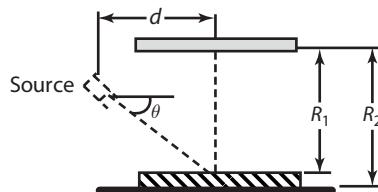
Range Sensing

The method of triangulation applies trigonometric principles to determine the distance of an object from two previously known positions.

$$t = R_2 - R_1 = R_2 - d \tan \theta$$

Here d is found from the position of the light spot on the workpiece in Figure 3-117.

FIGURE 3-117



Optical Ranging Using Position Sensitive Detectors

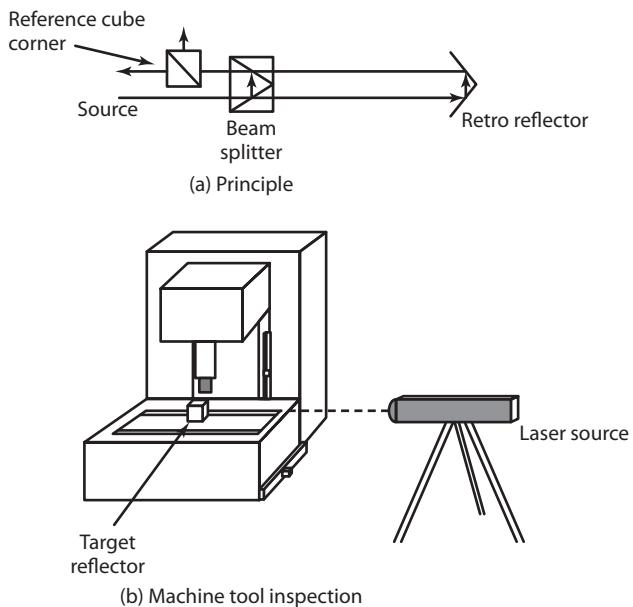
The light emitting diode and collimating lens transmit a pulse in the form of a narrow beam. After striking the object, the beam is reflected back to the detector. The received intensity is focused on the PSD. The sensor consists of a silicon device and provides position signals on a light spot traveling over its surface. The range is calculated.

Laser Interferometer

Laser interferometer (Figure 3-118) measures distance in terms of the wavelength of light by examining the phase relationship between a reference beam and a laser beam reflected from a target object.

Applications

- Range sensing techniques are used in manufacturing automation applications, such as automatic guidance systems, robot navigation, and collision avoidance.
- Optical principles are widely used for precision position measurement.
- Laser interferometers are also used for precision-motion measurement, checking of the linearity of precision-machine tool slides, and perpendicularity of machine-tool structures (mainly during installations of machine tools).

FIGURE 3-118

Features

- One limitation of light-stripe scanning is the poor depth resolution that is obtained for object surfaces that are parallel to the light stripe. It can be overcome by scanning the image in two directions, one perpendicular to the other.
- Laser interferometers have extremely high order of accuracy and resolution in linear measurements from a few millimeters to a large distance

3.9.6 Laser Interferometric Transducer

A laser interferometer is an optoelectronic instrument that measures distance in terms of the wavelength of light by examining the phase relationship between a reference beam and a laser beam reflected from a target object. It has extremely high order of accuracy and resolution in linear measurements from a few mm to a large distance. As shown in Figure 3-118, the laser produces collimated light rays of single frequency present with phase coherence. The laser beam with an optical arrangement produces the reference beam. A part of the reference beam is transmitted to the target and a part of the reference beam is sent to the interferometer. The rays reflected from the target are recombined at the interferometer. The phase difference between the reference beam from the source, and the reflected beam from the target is equal to the extra length traversed by the beam. The digitized information from the difference between the two signals provide the distance information. As shown in the bottom Figure 3-118(b), laser interferometers are also used for precision motion measurement, checking of the linearity of precision machine tool slides, and the perpendicularity of machine tool structures (mainly during installations of machine tools).

3.9.7 Fiber-Optic Devices In Mechatronics

Fiber-optic sensing is a new area in sensing and transmission that is expected to find widespread use in Mechatronics applications. Main sensing applications using fiber optics are in the domain of temperature and pressure measurement. Since light can be modulated and transmitted to large distances, even to normally inaccessible areas using fiber optic bundles, there had been a large increase in the fiber optic based sensors. Using fiber optic wave guides, light can be modulated along different paths as shown in Figures 3-119 and 3-120.

FIGURE 3-119 OPTICAL FIBER

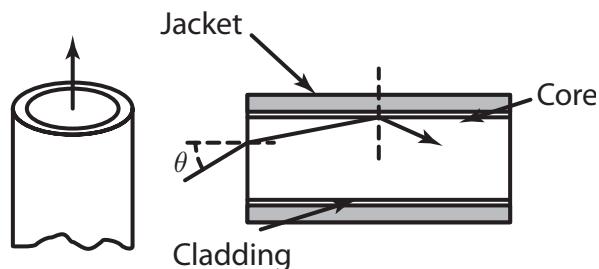
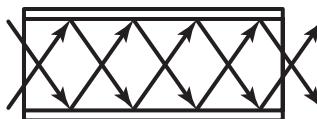
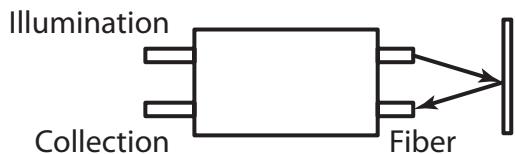


FIGURE 3-120 INTERNAL REFLECTION

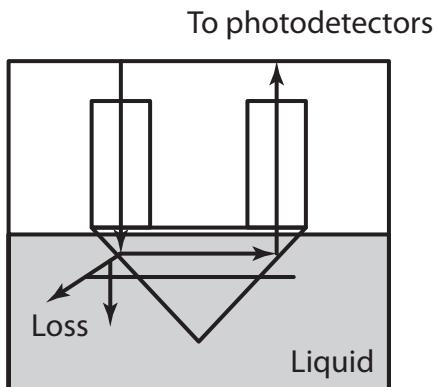


Optical fiber is basically a guidance system for light and is usually cylindrical in shape. If a light beam enters from one end face of the cylinder, a significant portion of energy of the beam is trapped within the cylinder and is guided through it and emerges from the other end. Guidance is achieved through multiple reflections at the cylinder walls. Internal reflection of a light ray is based on Snell's law in optics. If a light beam in a transparent medium strikes the surface of another transparent medium, a portion of the light will be reflected and the remainder may be transmitted (refracted) into the second medium. Light intensity, displacement (position), pressure, temperature, strain, flow, magnetic and electrical fields, chemical composition, and vibration are among the measurands for which fiber optic sensors have been developed.

Fiber bundles have highly internal reflective characteristics. The information can be transmitted either in the form of phase modulation or intensity modulation. Depending on the sensed property of light, fiber-optic sensors are also divided into phase-modulated sensors and intensity-modulated sensors. Intensity modulated sensors are simpler, more economical, and widespread in application.

FIGURE 3-121 DISTANCE SENSING

Two principles that are widely used in fiber-optic sensors are the reflective and the microbending principles. Both concepts sense displacement but can be used for other measurements, if the measurand can be made to produce a displacement. Figure 3-122 shows the schematic of a displacement sensor, used in an intensity mode. The incident light is transmitted back from the object. The analysis and comparison of transmitted and reflected intensities is done separately to give a measure of the distance. Any motion or displacement of the reflecting target can affect the reflected light that is being transmitted to the detector. The intensity of the reflected light captured depends on the distance of the reflecting target from the inspection probe. Disadvantages of this type of sensor are that they are sensitive to the orientation of the reflective surface and to the contamination.

FIGURE 3-122 LIQUID LEVEL

Figures 3-122 and 3-123 show examples of liquid level sensors. The level sensor in Figure 3-122, consists of two sets of optical fibers and a prism. When the sensor is above the liquid, most of the light is received by the receiver. When the prism reaches the liquid level, the angle of the total internal reflection changes because of the difference in the refractive index liquid and air. There is a higher loss of intensity of light that is detected at the receiver. Figure 3-123 shows another example of a level sensor. The U-shaped instrument modulates the intensity of passing light. The detector has two regions of sensitivity at the bent region of the U-shape. Sensitive liquid droplets covering the region move away from the region when the level sensor is lifted thereby providing a different output than the former position. When the sensitive regions touch the liquid, the light propagated through the fiber drops.

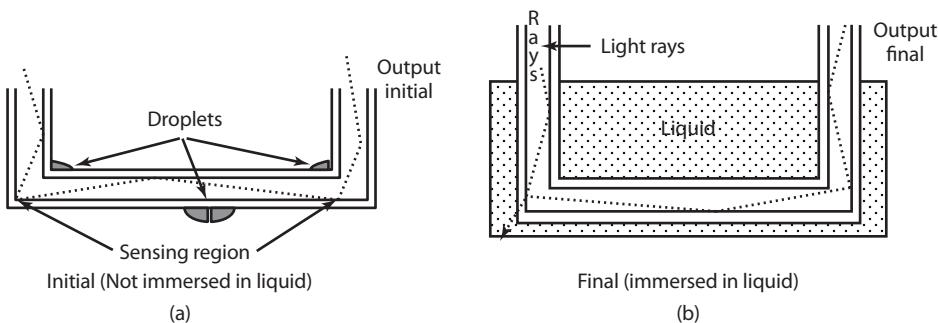
FIGURE 3-123 LIQUID LEVEL

Figure 3-124 shows the schematic of micro-strain gauges. In this case, fiber-optic bundles are squeezed between two deformers. The external force influences the total internal reflection of the fibers. Instead of reflection, light beam moves orthogonally and refracts through the fiber wall. The modulated intensity of light by the applied pressure gives a measure of the applied force. Microbend fiber-optic strain gauges have application in the areas of tactile sensing and vibration monitoring. If a fiber is bent as shown in Figure 3-124, a portion of the trapped light is lost through the wall of the fiber. The amount of the received light at the detector compared to the light source is a measure of the physical property influencing the bend.

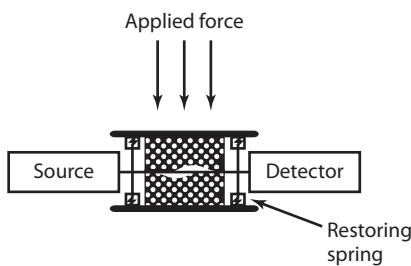
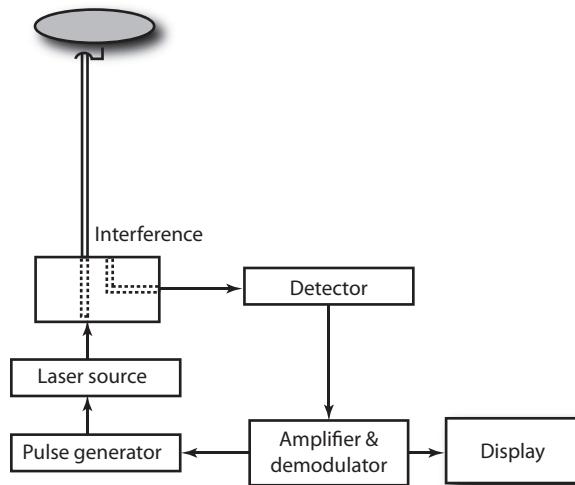
FIGURE 3-124 MICROBEND STRAIN GAUGES

Figure 3-125 shows the principle of fiber-optic temperature sensing. Such types of sensors are used in ships and large buildings where there is a need to transmit temperature data over large distances. The normal source of light is a pulse laser. The temperature is sensed by using the principle of back scattering of light. The delay occurring in the reflected laser pulses in comparison to the incident pulses is an indication of the measure of the temperature.

Several fiber-optic sensing concepts have been used in measurement of temperature. These include reflective, microbending, and other intensity- and phase-modulated concepts. In reflective sensors, the displacement of a bimetallic element is used as an indication of temperature variation. Active sensing material (such as liquid crystals, semiconductor materials, materials that

FIGURE 3-125 FIBER-OPTIC TEMPERATURE SENSOR

produce fluorescence, and other materials that can change spectral response) can be placed in the optical path of a temperature probe to enhance the sensing effect. The radiated light from a surface (which represents the surface temperature) can be collected and measured by a fiber-optic sensor called a blackbody fiber-optic sensor. Blackbody fiber optic sensors use silica or sapphire fibers, with the fiber tip coated with precious metal for light collection. These sensors can have a range of 500 to 2000°C. Fiber-optic temperature sensors have additional advantages of high resolution. Figure 3-126 presents a photograph of a fiber-optic liquid level sensor. Several

FIGURE 3-126 FIBER-OPTIC LIQUID LEVEL SENSOR

Courtesy of Gems Sensors Inc., Plainville, CT.

fiber-optic concepts are being used in design of fiber optic pressure sensors which have demonstrated high accuracy. Optical fibers have extensive application in telecommunication and computer networking, but their application as sensing devices is not that widespread. Optical sensing and signal transmission have several potential advantages over conventional electric output transducers and electric signal transmission.

3.10 Summary

Sensors are required to monitor the performance of machines and processes and to compensate for the uncertainties and irregularities of the work environment. Using a collection of sensors, we can monitor a particular situation in an assembly line, in a way that can substitute a human being. Sensors can be used to evaluate operations, conditions of machines, inspection of the work in progress, and identification of parts and tools. Sensors are also used for pre-process, post-process inspection and on-line measurements. Some of the more common measurement variables in mechatronic systems are temperature, speed, position, force, torque, and acceleration. When measuring these variables, several characteristics become important. These include the dynamics of the sensor, stability, resolution, precision, robustness, size, and signal processing. Progress in semiconductor manufacturing technology has made it possible to integrate various sensory functions. Intelligent sensors are available that not only sense information but process it as well. These sensors facilitate operations normally performed by the control algorithm, which include automatic noise filtering, linearization sensitivity, and self-calibration. The ability to combine these mechanical structures and electronic circuitry on the same piece of silicon is an important breakthrough. Many microsensors, including biosensors and chemical sensors, have the potential to be mass produced.

REFERENCES

- Smaili, A., Mirad, F., *Applied Mechatronics*, Oxford University Press, NY 2008.
- Sabri, Centinkunt, *Mechatronics*, John Wiley and Sons, Hoboken, NJ, 2007.
- Hegde, G.S., *Mechatronics*, Jones and Bartlett Publishers, Boston, MA, 2007.
- Necsulescu, Da., *Mechatronics*, Prentice Hall, NJ, 2002.
- Pawlak, Andrzej., *Sensors and Actuators in Mechatronics*, CRC-Taylor and Francis, Boca Raton, FL., 2007.
- Alciatore, David, and Histan, Michael., *Introduction to Mechatronics and Measurement Systems*, Third Edition, McGraw Hill, NY 2007.
- Rizzoni, Giorgio, *Principles and Applications of Electrical Engineering*, Third Edition, McGraw-Hill, NY, 2000.
- Aberdeen Group, *System Design: New Product Development for Mechatronics*, Boston, MA,
- January 2008 and NASA Tech Briefs, May 2009 (www.aberdeen.com).
- Brian Mac Cleery and Nipun Mathur, “Right the first time” *Mechanical Engineering*, June 2008.
- Bedini, R., Tani, Giovanni, et. al “From traditional to virtual design of machine tools, a long way to go—Problem identification and validation” Presented at the International Mechanical Engineers Conference, IMECE, November 2006.
- Pavel, R., Cummings, M. and Deshpande, A., “Smart Machining Platform Initiative.—First part correct philosophy drives technology development,” *Aerospace and Defense Manufacturing Supplement*, Manufacturing Engineering, 2008.
- Hyungsuck Cho, *Optomechatronics – Fusion of optical and Mechatronic Engineering* Taylor and Francis & CRC Press, 2006.
- Lee, Jay, “E-manufacturing—fundamental, tools, and transformation” *Robotics and Computer Integrated Manufacturing* 2003.

- Landers, R.G. and Ulsoy, A.G., "A Supervisory Machining Control Example," *Recent Advances in Mechatronics*, ICRAM '95, Turkey, 1995.
- Ohba, Ryoji., "Intelligent Sensor Technology," John Wiley & Sons. New York, NY, 1992.
- Philpott, M.L., Mitchell, S.E., Tobolski, J.F., and Green, P.A., "In-Process Surface Form and Roughness Measurement of Machined Sculptured Surfaces," *Manufacturing Science and Engineering*, Vol. 1, ASME, PED-Vol. 68-1, 1994.
- Stein, J. L. and Huh, Kunsoo, "A Design Procedure For Model Based Monitoring Systems: Cutting Force Estimation As A Case Study," *Control of Manufacturing Processes*, ASME, DSC, vol 28/PED-vol 52, 1991.
- Stein, J. L. and Tseng, Y. T. "Strategies For Automating The Modeling Process," *ASME Symposium For Automated Modeling*, ASME, New York, 1991.
- Shetty, D., and Neault, H., "Method and Apparatus for Surface Roughness Measurement Using Laser Diffraction Pattern," United States Patent, Patent Number: 5,189,490, 1993.
- NI LabVIEW-SolidWorks Mechatronics Toolkit, <http://www.ni.com/mechatronics/>
- Shetty, D., "Design For Product Success" *Society of Manufacturing Engineers*, Dearborn, MI, 2002.
- Sze, S.M., *Semiconductor Sensors*. John Wiley & Sons, Inc., 1994.
- Ulsoy, A.G., and Koren, Y., "Control of Machining Processes," *Journal of Dynamic Systems, Measurement, and Control*, Vol. 115, pp. 301–308, 1993.
- Bolton, W., "Programmable Logic Controllers, Second Edition," Newnes, Woburn, MA, 2000.
- Bolton, W., *Mechatronics-A Multidisciplinary Approach*, Fourth Edition, Prentice Hall, NJ, 2009.
- Pallas-Aveny, R., Webster, J., *Sensor and Signal Conditioning*, John Wiley & Sons, NY, 1991.

PROBLEMS

Errors and Sensitivity Analysis:

- 3.1. A torque transducer is used to measure the power of a rotating shaft. During the mode of measurement, the following parameters are monitored.
- Speed of rotation of the shaft during the time t , (R)
 - Force at the end of the torque arm, (F)
 - Length of the torque arm, (L)
 - Time (t)

The errors in each of the measurements are

- Shaft speed, $R = 2502 \pm 1$ revolutions
- Force on the arm, $F = 55.02 \pm 0.18$ N
- Length of the arm, $L = 0.0397 \pm 0.0013$ m
- Time in seconds, $t = 30 \pm 0.50$ s

The power is computed using the equation

$$\text{Power} = \frac{2 \cdot \pi \cdot R \cdot F \cdot L}{t}.$$

Determine the absolute error in the measurement of torque.

- 3.2. The discharge coefficient, C_q , of an orifice can be found by collecting water that flows during a timed interval when it is under constant head, h . The following formula is used to measure the discharge coefficient.

$$C_q = \frac{W}{(t)(\rho)(A)\sqrt{2gh}}$$

where

$$\begin{aligned}W &= 200 \pm 0.23 \text{ kg} \\t &= 500 \pm 2 \text{ s} \\\rho &= 1000 \text{ kg/m}^3 \\d &= 1.25 \pm 0.0025 \text{ cm} \\g &= 9.81 \pm 0.11 \text{ m/s}^2 \\h &= 3.66 \pm 0.003 \text{ m}\end{aligned}$$

Find C_q and its component error.

- 3.3. The resistance of certain length of wire R is given by $R = 4\rho l \pi d^2$ where

$$\begin{aligned}\rho &= \text{resistivity of the wire in } \Omega\text{-cm} \\l &= \text{length of the wire in cm} \\d &= \text{diameter of the wire in cm}\end{aligned}$$

Determine the nominal resistance and the uncertainty in resistance of the wire with the following data.

$$\begin{aligned}\rho &= 45.6 \times 10^{-6} \pm 0.15 \times 10^{-6} \Omega\text{-cm} \\l &= 523.8 \pm 0.2 \text{ cm} \\d &= 0.062 \pm 1.2 \times 10^{-3} \text{ cm}\end{aligned}$$

- 3.4. Calculate the power consumption in an electric circuit. The voltage and current are measured to be, $V = 50 \pm 1 \text{ V}$, $I = 5 \pm 0.2 \text{ A}$. What is the maximum possible error?
- 3.5. This example is about an explosive detonation manufacturer. The shell is filled with explosives. A pressure of 35,000 kPa (absolute) is exerted as shown. The formula for hoop stress is given as

$$\sigma = \frac{pr}{t}$$

Find the hoop stress on the wall of the shell and component error if

$$\begin{aligned}\text{Pressure exerted is} &= 35,000 \pm 70 \text{ kPa (absolute)} \\\text{Shell Radius} &= 0.287 \pm 0.007 \text{ cm} \\\text{Shell Thickness} &= 0.028 \pm 0.0001 \text{ cm}\end{aligned}$$

- 3.6. The mass moment of inertia for a sphere is given by

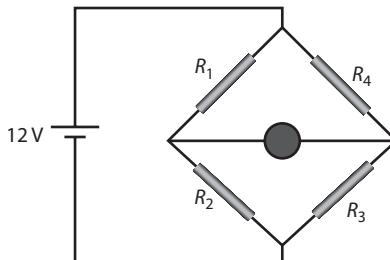
$$I_{xx} = I_{yy} = I_{zz} = \frac{2mr^2}{5}$$

where

$$\begin{aligned}m &= \text{mass of the sphere in kg} \\r &= \text{radius of the sphere in mm} \\m &= 5 \pm 0.04 \text{ kg}; r = 100 \pm 0.2 \text{ mm}\end{aligned}$$

Calculate the absolute error in the measurement of the mass.

- 3.7. Choose the appropriate definitions from the following list for the sentences.
- a. *null-type device*
 - b. *amplifier*
 - c. *drift*
 - d. *transducer*
 - e. *precision*
 - f. *accuracy*
 - g. *calibration*
 - h. *resolution*
 - i. *linearity*
 - j. *backlash*
 - k. *relative error*
 - l. *noise*
- () Device whose output is an enlarged reproduction of the essential features of the input wave and which draws power from a source other than the input signal
- () Measure and generates an opposing effect to maintain zero deflection
- () A device that converts input energy into a form of an output with different form of energy
- () Ratio of difference between measured value and true value of the measurand
- () Smallest increment in measurand that can be detected with certainty by the instrument
- () Ability of the instrument to give identical output measurements when repeat measurements are made with the same input signal
- () Gradual departure of the instrument output from the calibrated value
- () Maximum distance or an angle, any part of the mechanical system can be moved in one direction without causing the motion of the next part
- () Characteristic of the instrument whose output is a liner function of the input
- 3.8. The voltmeter scale has 100 divisions. The scale can be read to 1/5 of a division. Calculate the resolution of the instrument in mm.
- 3.9. A rotary variable differential transformer (RVDT) has a specification on ranges and sensitivities.
- Range $\pm 30^\circ$, linearity error $\pm 0.5\%$ full range
Range $\pm 90^\circ$, linearity error $\pm 1.0\%$ full range. Sensitivity 1mV/V input per degree
- What is the error reading in 50° due to non linearity if the RVDT is used in $\pm 90^\circ$ range?
- 3.10. What will be the change in resistance of a strain gauge, with a gauge factor of 4 and resistance of $50\ \Omega$ if the gauge is subjected to a strain of 0.002?
- 3.11. A pressure gauge uses four strain gauges to monitor the displacement of a diaphragm. Four active gauges are used in a bridge circuit (Figure P3-11) The gauge factor is 2.5 and resistance of gauges $100\ \Omega$. Because of the differential pressure on the diaphragm, gauges R_1 and R_3 are subjected to tensile strain of $(2)(10)^{-4}$ and gauges R_2 and R_4 are subjected to compressive strain of $(2)(10)^{-4}$.
The supply voltage to the bridge is 12 V. What will be the offset voltage?
- 3.12. A force of 5400 N is exerted on an aluminum rod, whose diameter is 6.2 cm and length 30 cm. Calculate the stress and strain in the beam if the Young's modulus of aluminum is 70 GN/m^2 . A strain gauge with a gauge factor of 4 and resistance of $350\ \Omega$ is attached to the rod. Calculate the change in resistance. If the strain gauge is used in a bridge circuit and all other resistances are $350\ \Omega$, find the offset voltage of the bridge. Supply voltage of the bridge is 10 V.

FIGURE P3-11 BRIDGE CIRCUIT

- 3.13. A Resistance-wire strain gauge uses soft iron wire of small diameter. Gauge factor is +4.2. Neglect piezo-resistive effect. Calculate Poisson's ratio.
- 3.14. A compressive force is applied to a structure, the strain = 5 microstrains. Two separate strain gauges are attached to the structures, one is a nickel wire strain gauge of gauge factor = -12.1 and another is a nicrome wire strain gauge of gauge factor = 2. Calculate the value of resistance of the gauges after they are strained. The resistance of strain gauge = $120\ \Omega$.
- 3.15. A resistance wire strain gauge with a gauge factor = 2 is bonded to a steel structure member subjected to a stress of 100 MN/m^2 . Modulus of elasticity of steel is 200 GN/m^2 . Calculate the percentage change in value of the gauge resistance due to the applied stress.
- 3.16. A strain gage has a resistance of $250\ \Omega$ and a gage factor of 2.2. It is bonded to an object to detect movement. Determine the change in resistance of the strain gage if it experiences a tensile strain of 450×10^{-6} due to the change in size of the object. Also, if the relationship between change in resistance and displacement is $0.05\ \Omega.\text{mm}^{-1}$, determine the change in the size of the object.
- 3.17. A steel bar with modulus of elasticity 200 GPa and diameter 10mm is loaded with an axial load of 50 kN. If a strain gage of gage factor 2.5 and resistance $120\ \Omega$ is mounted on the bar in an axial direction, first find the change in resistance. Assuming this change in resistance is in positive direction, let us connect the strain gage to one branch of a wheatstone bridge (R_1) with the other three legs having the same base resistance ($R_2 = R_3 = R_4 = 120\ \Omega$). Input voltage to the bridge is 12 V. What is the output voltage of the bridge in the strained state?
- 3.18. This is an example of a sensing operation during the process of work-handling in a robot manipulator. Strain gauges can be used to measure the force acting on the object while the object is gripped. Strain gauges are mounted on the fingers of a gripper. Strain gauges 2 and 3 are attached inside of the finger. Strain gauges 1 and 4 are attached to the outside of the finger. When the object is grasped, gripping force causes strain gauges 2 and 3 to stretch and 1 and 4 to compress. The resistance of the gauges 2 and 3 increase, while the resistance of gauges 1 and 4 decrease. Suppose the strain gauges are used as force-sensors, what is the bridge output when there is no gripping force? What is the output voltage for a gripping force that causes a strain of $3000\ \mu\text{m}$. (Let us assume the supply voltage to be 12V; strain gauges have unstrained resistance of $1000\ \Omega$. Use the formula, $\Delta R = 2R_{nom} \cdot strain$.)
- 3.19. (a) What will be the change in resistance of a strain gauge, with a gauge factor of 2 and resistance of $100\ \Omega$, if the gauge is subjected to a strain of 0.005 ? (b) An angular incremental encoder is used with a 80 mm radius tracking wheel. This is used to monitor linear displacement. The angular

encoder provides 128 pulses per one rotation. What will be the number of pulses for a linear movement of 250 mm

- 3.20. Strain is monitored in a cantilever beam using strain gauges of resistance $1\text{ K}\Omega$, GF = 2 and temperature Coefficient = $10 - 5/\text{ }^{\circ}\text{C}$ at room temperature. It is mounted on beam and connected to the bridge circuit.
 - Calculate the change in resistance of the gauge if the gauge is strained 0.1% (Use strain 5 .0011);
 - Calculate the change in effective strain indicated when the room temperature increases by $10\text{ }^{\circ}\text{C}$;
 - Suggest a way of reducing this temperature effect.
- 3.21. A resistance transducer has a resistance of $250\ \Omega$ and a gage factor of 2.2. It is bonded to an object to detect movement. Determine the change in resistance of the strain gauge if it experiences a strain of 450×10^{-6} due to the change in the size of the object. Also if the relationship between the change in the resistance and displacement is $0.05\ \Omega$ per mm, determine the size of the object.
- 3.22. A strain gage bridge has a strain gage of resistance $R = 200\ \Omega$ and gage factor $G = 1.9$. R_2 , R_3 , and R_4 are fixed resistors also rated at $200\ \Omega$. The strain gage experiences a tensile strain of 400 microstrain due to the displacement of an object. Determine the change in resistance ΔR of the strain gage. If the input voltage is V_i volts then determine the change in output voltage ΔV_o

UNITS

$$1\text{Picofarad (pF)} = 10^{-12}\text{ f}, 1\text{ Nanofarad (nF)} = 10^{-9}\text{ f}$$

- 3.23. A capacitance transducer consists of two plates of diameter 2 cm each, separated by an air gap of 0.25 mm. Find the displacement sensitivity
- 3.24. A capacitance transducer has two plates, with 12 cm^2 area and are apart by 0.12 cm. The plates are in vacuum. Given the permittivity of vacuum is $8.85 \times 10^{-12}\text{ F/m}$, calculate the capacitance. What would happen to the capacitance if one of the plates were moved 0.12 cm further away from the other plate?
- 3.25. A transducer using the capacitance principle consists of two concentric cylindrical electrodes. The outer diameter of inner cylinder is 4 mm. The inner diameter of the outer electrode is 4.2 mm. The length of the electrode is 0.03 m. Calculate the change in capacitance if the inner electrode is moved through a distance of 1.5 mm.
- 3.26. A parallel plate Capacitance transducer uses plates of area 500 mm which are separated by a distance of 0.2 mm. (a) calculate the value of capacitance when the dielectric is air having a permitivity of $8.85 \times 10^{-12}\text{ F/m}$. (b) A linear displacement reduces the gap length to 0.18 mm. Calculate the change in capacitance. (c) Calculate the ratio of per unit change of capacitance to per unit change in displacement. (d) Suppose a mica sheet of .01 mm thick is inserted in the gap, Calculate the value of original capacitance and change in capacitance for the same displacement. The dielectric constant of mica is $8[C = A/d]$.
- 3.27. A quartz PZT crystal having a thickness of 2 mm and voltage sensitivity of 0.055 Vm/N is subjected to a stress of 1.5 MN/sq.m . Calculate voltage output and charge sensitivity.
- 3.28. A ceramic pickup has a dimension of $5\text{ mm} \times 5\text{ mm} \times 1.25\text{ mm}$. The force acting on it is 5 N. The charge sensitivity of the crystal is 150 pC/N , its permitivity $12.5 \times 10^{-9}\text{ F/m}$. If the modulus of elasticity of the crystal is $12 \times 10^6\text{ N/m}^2$, calculate the strain, the charge, and the capacitance.
- 3.29. A piezoelectric crystal has a dimension of 100 mm^2 . Its thickness is 1.25 mm. It is held between two electrodes for measuring the change of force across the crystal. Young's modulus of the crystal is 90 GN/m^2 . Charge sensitivity is 110 pC/N . Permitivity is $(\epsilon_0\epsilon_r)$ 1200. The connecting cable has a capacitance of 250 pF, while the oscilloscope for display has a capacitance of 40 pF. What is the resultant capacitance?

- 3.30. Piezoelectric crystal of 1 cm^2 area, 0.1 cm thick has been subjected to a force. Two metal electrodes measure the changes in the crystal. Young's modulus of the material = $9 \times 10^{10} \text{ Pa}$. Charge sensitivity $2 \mu\text{C/N}$, Relative permittivity is 5; the applied force is 0.01 N
- Find the voltage across the electrodes.
 - Find the change in crystal thickness

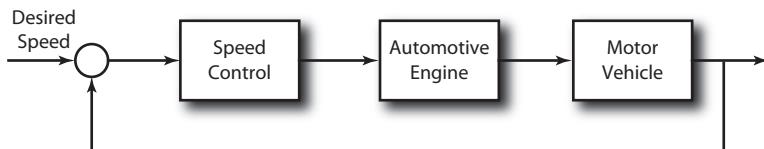
$$\text{Output Voltage} = \frac{gtF}{A}; g = \frac{d}{\epsilon_r \epsilon_0} \text{ Vm/N}$$

- 3.31. The output of an inductance type transducer (such as LVDT) is connected to a 5 V voltmeter. An output of 2 mV appears across the terminals of the transducer when the core of the LVDT moves through a distance of 0.1 mm . Calculate the sensitivity of LVDT.
- 3.32. In a resistance temperature detector (RTD) using platinum and nickel, the temperature coefficient at 20°C is $0.004/\text{C}$ and resistance $R = 106 \Omega$. Find the resistance at 25°C .
- 3.33. RTD of Problem 3.32 is used in a bridge circuit. If $R_1 = R_2 = R_3 = 100 \Omega$, Supply voltage is 10 V . Calculate the voltage the detector must resolve to define 1°C change in temperature.

System:

- 3.34. A steel mill has a production set up where metal sheets are rolled for desired thickness as they emerge from the production sequence. It is a continuous, real-time production and measurements have to be made on-line. Suggest a sensor that can do the job. The final output should be electrical.
- 3.35. Figure P3-35 shows a block diagram of an automotive cruise control system. This helps the driver in monitoring and controlling the speed.

FIGURE P3-35 AUTOMOTIVE SPEED CONTROL SYSTEM



- Draw similar diagrams for the following applications by showing the modules of instrumentation system
- Automatic coffee maker for home use
 - Motion of axes in a machine tool
- 3.36. A hospital is interested in developing an instrument to measure the force exerted by the human finger. This instrument will be useful in the rehabilitation department. How will you approach the design of such an instrument? Identify the type of sensor, explain its principle with a possible sketch. How will you proceed with the data acquisition and display concept?
- 3.37. The automatic control system for the temperature of a bath of liquid consists of a reference voltage fed into a differential amplifier. This is connected to a relay, which then switches on or off the electrical power to a heater in the liquid. Negative feedback is provided by a measurement system, which feeds a voltage into the differential amplifier. Sketch a block diagram of the system and explain how the error signal is produced.

System:

- 3.38. Indicate **True** or **False** or the correct answer.
- Condition monitoring means monitoring the condition of a machine when it is not running (T or F).
 - Eddy-current type of transducer produces an output proportional to velocity (T or F).
 - A common LVDT is
 - A differential transformer
 - A mechanical position-to-electrical transducer sensor
 - Inductive electromechanical transducer
 - All the above
 - A capacitance transducer has two plates of area 5 cm^2 each, separated by an air gap of 1 mm thickness. Value of capacitance is 442 pF. (T or F).
 - Mechatronic Supervisory control system requires:
 - A digital computer monitoring the system performance
 - Individual controllers actually controlling each of the processes
 - The controllers get the set point from the computer
 - All of the above
 - A supervisor in the loop
 - Which parameter the bonded strain gauge measures?
 - Deformation
 - Torque
 - Force
 - Pressure
 - Stress
 - Which of the following parameters can a proximity sensor be used to measure?
 - Speed of rotation of a shaft
 - Closeness of an object
 - Deformation of a metal piece
 - Relative position of two linear motion surfaces
 - Instantaneous position of a rotating shaft
 - Which of the following phenomenon is commonly used in industry to sense very small changes in the physical dimensions of a load (force) column?
 - The proportionality between liquid level and pressure.
 - The attenuation of nuclear radiation by solid materials.
 - The variation of resistance of a wire as it is deformed.
 - The sensitivity of hair to moisture.
 - The principle that, if hydraulic flow-velocity is high, the corresponding pressure will be low, and vice versa.
 - Select the right answer: Rotameter is a
 - Drag-force flow meter
 - Variable-area flow meter
 - Variable-head flow meter
 - Rotating propeller-type flow meter
 - Rotating speed indicator
 - Turbine flow meters are primarily used to measure the flow of fluids which are
 - Corrosive
 - Chunky
 - Viscous
 - Petrochemical
 - For all liquids mentioned above

- k. The type of electrical output should be expected from a digital shaft angle encoder?
- A series of digital pulses over a single pair of output wires.
 - Several parallel wires, each one with a digital voltage level, which must be interpreted together to get the shaft angle.
 - A variable resistance analog signal.
 - A bipolar dc voltage.
- l. Which of the following statements describe properties inherent in an open loop control system?
- Output has no effect on input.
 - Inherently stable.
 - Controller has no way of knowing if its command was executed.
 - Controller does not care whether its command was executed.
 - All of the statements above describe an open loop control system.
- 3.39. Make a table listing in one vertical column each of the following sensors: Pneumatic, LVDT, Eddy Current, Hall Effect. Then make four adjacent vertical columns, labeling them: Variable Measured, Principle of Operation, Advantages/ Disadvantages. Attempt to fill every blank space in the table.
- 3.40. Identify the sensor, signal conditioner, and display elements of a measurement system such as a mercury-in-glass thermometer. Identify the input and output parameters

CHAPTER 4

ACTUATING DEVICES



- 4.1 Direct Current Motors
 - 4.1.1 Mathematical Model of a DC Motor
 - 4.1.2 Brushless DC Motors
 - 4.1.3 AC Motors
- 4.2 Permanent Magnet Stepper Motor
 - 4.2.1 Modeling Approach
 - 4.2.2 Drive Equations and Block Diagram Model
 - 4.2.3 Motor Equations and Block Diagram Model
 - 4.2.4 Position System Using Stopper Motor
- 4.3 Fluid Power Actuation
 - 4.3.1 Control Systems in Fluid Power
 - 4.3.2 Fluid Power Actuators
- 4.4 Fluid Power Design Elements
 - 4.4.1 Fluid Power Energy-Input Devices
 - 4.4.2 Energy Modulation Devices (Valves)
 - 4.4.3 Energy-Output Devices
 - 4.4.4 Control Modes of Fluid Power Circuits
 - 4.4.5 Other Electric Components in Fluid Power Circuits
- 4.5 Piezoelectric Actuators
- 4.6 Summary
- References
- Problems

Mechatronic systems employ actuators or drives that are part of the physical process being monitored and controlled. Actuation is the result of a direct physical action upon the process, such as removing a workpiece from a conveyor system or the application of a force. It has a direct effect upon the process. Actuators take low power signals transmitted from the computer and produce high power signals which are applied as input to the process. There are many types of actuating devices, some of the most common ones include solenoids, electrohydraulic actuators, DC or AC motors, stepper motors, piezoelectric motors, and pneumatic devices.

Electrical actuators convert electrical command signals into mechanical motions. In this chapter, emphasis is placed on DC motors, stepper motors, and fluid power devices (electrohydraulic) because of their popularity in mechatronics. Although the main focus in this chapter is on DC motors, it should be noted that AC motors are also widely used for servomechanism.

4.1 Direct Current Motors

The major factors in selecting an actuator for mechatronic applications are

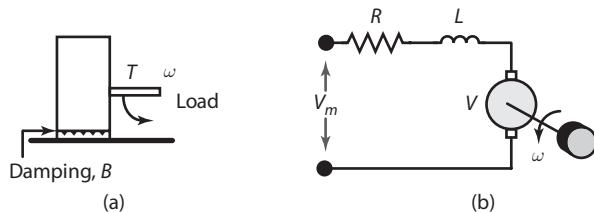
- Precision
- Accuracy and resolution
- Power required for actuation
- Cost of the actuation device

The most popular actuators in mechatronic systems are direct current (DC) motors. DC motors are electromechanical devices that provide precise and continuous control of speed over a wide range of operations by varying the voltage applied to the motor. The DC motor is the earliest form of electric motor.

The desirable features of DC motors are their high torque, speed control ability over a wide range, speed-torque characteristics, and usefulness in various types of control applications. DC motors are well suited for many applications, including manufacturing equipment, computer numerically controlled systems, servo valve actuators, tape transport mechanisms, and industrial robots.

The DC motor converts direct-current electrical energy into rotational mechanical energy. It makes use of the principle that a wire carrying a current in a magnetic field experiences a force. The windings wrapped around a rotating armature carries current. The armature is the rotating member (rotor), and the field winding is the stationary winding (stator). The rotor has many closely spaced slots on its periphery. These slots carry the rotor windings. The rotor windings (armature windings) are powered by the supply voltage. An arrangement of commutation segments and brushes ensures the transfer of DC current to the rotating winding. A schematic of a DC motor is shown in Figure 4-1.

FIGURE 4-1 (A) CONVENTIONAL DC MOTOR DIAGRAM (B) LOADING



4.1.1 Mathematical Model of a DC Motor

The behavior of DC motors can be explained by two fundamental equations. These equations are known as torque and voltage equations. Equations 4-1 and 4-2 present the torque equation and voltage equations, respectively.

$$\text{Torque equation:} \quad T = k_t i \quad (4-1)$$

$$\text{Voltage equation:} \quad V = k_e \theta \quad (4-2)$$

where

T = motor torque in N-m (newton-meters)

V = induced voltage in V (volts)

i = current in the armature circuit in A (amperes)

θ = rotational displacement of the motor shaft in rad (radians)

k_t = torque constant in Nm/A

k_e = voltage constant in V/(rad/sec)

When an input voltage, V_m is applied to the armature, the voltage equation is influenced by the drop in the voltage because of the voltage drop, RI , across the armature resistance.

$$V_m = R_a i + L_a \frac{di}{dt} + V \quad (4-3)$$

where

V_m = voltage at the armature terminal in volts (V)

R_a = armature resistance in ohms (Ω)

L_a = armature inductance in henry (H)

i = armature current in ampere (A)

The inductance of the winding is usually neglected. This is because it represents a fraction of the armature flux that is not linked to the stator and not used in the generation of torque. The DC servo motor drives a mechanical load which consists of dynamic and static components. The primary loads on the motor are inertia and friction, and the varying torque is represented by Eq. 4-4.

$$T = J\ddot{\theta} + B\dot{\theta} + T_L \quad (4-4)$$

where

J = the moment of inertia of the rotor

B = the viscous damping coefficient

T_L represents the load on the motor

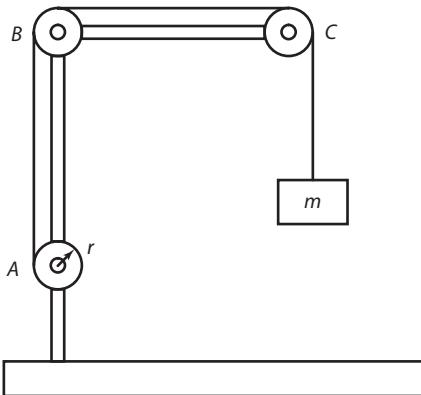
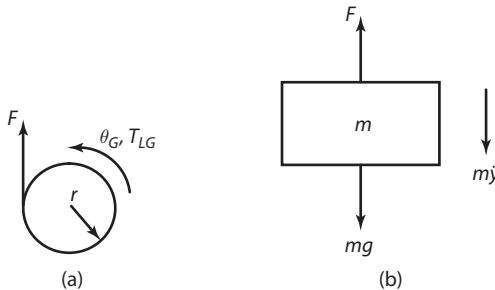
DC motors are capable of producing high rotational velocities and comparatively low torque. When the DC motors are used as actuators, a gearing arrangement is normally utilized to account for decreased speed and increased torque. DC motors provide torque which is proportional to the armature current. A DC source capable of supplying positive and negative currents is normally used in practice. A generally used arrangement of the DC motor is through DC coupled push-pull amplifiers. The selection of the DC motor depends upon its application. DC servo motors are used in numerically controlled machine tools and robot manipulators.

EXERCISE 4.1 Displacement of Permanent Magnet DC Motor

A permanent magnet (PM) DC gear motor is used to lift a mass, as shown in the Figure 4-2. Develop a mathematical relationship between the voltage applied to the motor and the rotational displacement of the motor shaft which is also a measure of the linear displacement of the mass. Assume that the string is inextensible, and also neglect the friction between the string and the pulleys.

Solution

In Figure 4-2, pulley A is coupled to the geared PM DC motor, while pulley B and C are idlers supporting the string. When pulley A rotates by an angle θ_G in the counterclockwise direction, the mass m will move up by a distance of $y = r\theta_G$. Figure 4-3(a) and (b) shows the free-body diagram of pulley A and mass m , respectively.

FIGURE 4-2 PERMANENT MAGNET DC GEAR MOTOR SYSTEM**FIGURE 4-3 PERMANENT MAGNET DC GEAR MOTOR SYSTEM FREE-BODY DIAGRAM**

For the moving mass using Newton's Law, we get

$$F = m\ddot{y} + mg = mr\ddot{\theta}_G + mg \quad (4-5)$$

For the rotating pulley after neglecting its inertia and friction losses, we get

$$T_{LG} = Fr = mr^2\ddot{\theta}_G + mgr \quad (4-6)$$

Hence, the load on the motor considering the gear ratio, G , will be

$$T_L = \frac{T_{LG}}{G} = \frac{mr^2\ddot{\theta}_G}{G} + \frac{mgr}{G} \quad (4-7)$$

Now, the relationship between the angular displacement of the motor shaft and gear output shaft is

$$\theta_G = \frac{\theta}{G} \quad (4-8)$$

Hence, from Equations 4-7 and 4-8, we get

$$T_L = \frac{mr^2\ddot{\theta}}{G^2} + \frac{mgr}{G} \quad (4-9)$$

From Equations 4-4 and 4-9, we get

$$T = J\ddot{\theta} + B\dot{\theta} + \frac{mr^2\ddot{\theta}}{G^2} + \frac{mgr}{G} \quad (4-10)$$

From Equations 4-1 and 4-10, we get

$$i = \frac{T}{k_t} = \frac{J\ddot{\theta}}{k_t} + \frac{B\dot{\theta}}{k_t} + \frac{mr^2\ddot{\theta}}{k_t G^2} + \frac{mgr}{k_t G} \quad (4-11)$$

Hence,

$$\frac{di}{dt} = \frac{J\ddot{\theta}}{k_t} + \frac{B\dot{\theta}}{k_t} + \frac{mr^2\ddot{\theta}}{k_t G^2} = \left(J + \frac{mr^2}{G^2} \right) \frac{\ddot{\theta}}{k_t} + \frac{B\dot{\theta}}{k_t} \quad (4-12)$$

Substituting Equations 4-2, 4-11, and 4-12 in Equation 4-3, we get

$$V_m = R_a \left(\frac{J\ddot{\theta}}{k_t} + \frac{B\ddot{\theta}}{k_t} + \frac{mr^2\ddot{\theta}}{k_t G^2} + \frac{mgr}{k_t G} \right) + L_a \left[\left(J + \frac{mr^2}{G^2} \right) \frac{\ddot{\theta}}{k_t} + \frac{B\dot{\theta}}{k_t} \right] + k_\theta \ddot{\theta} \quad (4-13)$$

For analysis, both torque constant and voltage constant can be assumed to be equal to k , hence Equation 4-13 reduces to

$$V_m - R_a \frac{mgr}{kG} = \frac{1}{k} \left[\left(J + \frac{mr^2}{G^2} \right) L_a \ddot{\theta} + \left(JR_a + BL_a + R_a \frac{mr^2}{G^2} \right) \dot{\theta} + (BR_a + k^2) \ddot{\theta} \right] \quad (4-14)$$

Equation 4-14 gives the required mathematical relationship between the voltage applied to the motor, V_m , and the rotational displacement of the motor shaft, θ , where the term $R_a \frac{mgr}{kG}$ is the voltage required to balance

the constant torque developed due to the gravitational force, mg . (Voltage = resistance \times Current; Current = torque/motor constant, and Torque = mgr/G)

EXERCISE 4.2 Simulation of Angular Displacement of the Motor

Simulate the response of the system described in Figure 4-2 for a constant input voltage of 10 V DC using MATLAB. Use the data given for a Shayang gear motor model number IG420049-SY3754.

Armature resistance, $R_a = 20.5 \Omega$

Armature inductance, $L_a = 168 \mu H$

Motor constant, $k = 0.032 \text{ Nm/A}$ (or $V/\text{rad/sec}$)

Gear ratio, $G = 49$

Mass, $m = 1.125 \text{ KG}$

Radius of the pulley, $r = 0.022 \text{ m}$

Solution

After neglecting rotor inertia and damping losses in the motor, Equation 4-14 reduces to

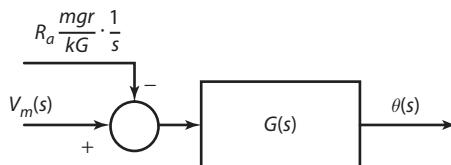
$$V_m - R_a \frac{mgr}{kG} = \frac{1}{k} \left(\frac{mr^2}{G^2} L_a \ddot{\theta} + R_a \frac{mr^2}{G^2} \dot{\theta} + k^2 \ddot{\theta} \right) \quad (4-15)$$

At zero initial condition applying the Laplace transform to Equation 4-15, we get

$$\frac{\theta(s)}{V_m(s) - R_a \frac{mgr}{kG} \frac{1}{s}} = G(s) = \frac{k}{\frac{mr^2}{G^2} L_a s^3 + R_a \frac{mr^2}{G^2} s^2 + k^2 s} \quad (4-16)$$

Equation 4-16 represents the open-loop transfer function of the system, which can be represented using a block diagram, as shown in Figure 4-4.

FIGURE 4-4 PM DC GEAR MOTOR SYSTEM OPEN-LOOP BLOCK DIAGRAM



MATLAB Code

```

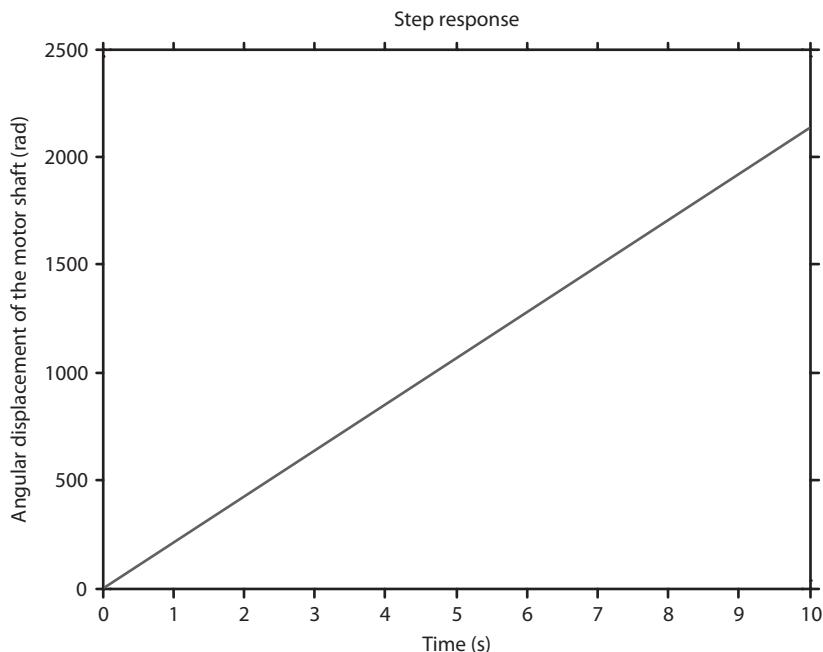
clear
clc
Ra = 20.5; %Armature Resistance, ?
La = 168E-6; %Armature Inductance, H
k = 0.032; %Motor Constant, Nm/A (or V/rad/sec)
G = 49; %Gear Ratio
m = 1.125; %Mass, KG
r = 0.022; %Radius of the pulley, m
g = 9.81; %Acceleration due to gravity, m^2/sec
Vm = 10; %Input voltage to the motor
%Vm(s)=Vm/s, constant input and
%hence, Vm(s)-(Ra(mgr/kG))/s=(Vm-Ra(mgr/kG))/s=Vrm/s, where
Vrm = Vm-Ra*((m*g*r)/(k*G));
Gs = tf(k,[m^2*La/G^2 Ra*m^2*r^2/G^2 k^2 0]);
t = 0:0.01:10;
U = Vrm*ones(size(t));
lsim(Gs,U,t)
ylabel('Angular displacement of the motor shaft (rad)')

```

Result

Figure 4-5 shows the response of the system for a constant voltage of 10 V DC. As seen from the figure, if 10 V is constantly applied to the motor, the motor shaft will move by 2130 rad in 10 sec (i.e., the mass will move by $0.022 \times 2130 = 46.86$ m).

As expected, the result shows that, with a constant voltage given to the motor, it will continue to rotate. However, to lift the mass to a specified height, we would need a controller that would monitor the angular

FIGURE 4-5 STEP RESPONSE OF THE OPEN-LOOP SYSTEM

displacement of the motor shaft and develop a controlled input voltage to the motor that would take the mass to the specified height. The design of one such controller is explained in Chapter 6.

4.1.2 Brushless DC Motors

A major maintenance problem in conventional DC motors is brush arcing. The magnetic polarity of the stator is fixed, and the polarity of the rotor is switched mechanically to get proper direction of motor torque. The armature voltage is supplied by a pair of brushes that maintain contact with split slip-ring commutation. Brushes are the weak factors in DC motors, and they generate excessive noise, contact bounce, and maintenance problems due to rapid wear out. Brushless DC motors prevent brush arcing by putting the permanent magnet in the rotor and energizing the stator through angular positions.

Modern, brushless DC motors use solid-state switching for commutation. In these motors, electrical commutation duplicates mechanical brush commutation. In brushless DC motors, the polarity of the rotor unit, which is a permanent magnet, is fixed relative to the rotor itself, and the polarity of the stator is switched by electronic means to achieve the same objective. Since the electrical commutation simulates the mechanical commutation in conventional systems, brushless DC motors exhibit similar torque speed characteristics.

The advantages of brushless DC motors are high reliability and the ability to generate relatively high torque at speeds up to 100,000 rpm. Brushless DC motors are used in general-purpose applications, as well as in servo systems for motion control applications. Motors in the range up to 1 hp and operating at speeds up to 7,200 rpm are used in computer peripherals and also are used as drivers for fluid power devices.

4.1.3 AC Motors

Alternating current motors have become popular in many machine tools. AC motors operate without brushes. They are more reliable, rugged in construction, and have less maintenance. AC motors are classified as single phase and polyphase and again are subdivided into induction and synchronous motors. The velocity of the AC synchronous motor is controlled by the variable frequency supply. The main advantage of the AC motor over the DC motor is its interfaceability with the AC signals of synchro resolvers and other AC transducers. The popularity of alternating current motors (AC) is due to the following reasons.

- Most of the power-generating systems produce alternating current.
- AC motors cost less than (direct current) DC motors.
- Some AC motors do not use brushes and commutators. This eliminates many problems of maintenance and wear. It also eliminates the problem of dangerous sparking.

The AC motor is particularly well-suited for constant-speed applications. This is because its speed is determined by the frequency of the AC voltage applied to the motor terminals. The DC motor is better-suited for applications that require variable speeds. An AC motor can also be made with variable speed characteristics but only within certain limits. AC motors are available in different sizes, shapes, and ratings for many different types of jobs. Based on the power requirements, they can be classified as single phase and polyphase which are further subdivided into induction and synchronous motors based on rotor magnetic field, which is either induced in the rotor by the stator field (as in case of induction motor) or provided by a separate DC current source.

4.2 Permanent Magnet Stepper Motor

In recent years, the stepper motor has emerged as a cost-effective alternative to DC motors in motion-control applications. The stepper motor is an actuator which translates electrical pulses into precise, equally spaced, angular movements of the rotor in the form of steps. The rotor is positioned by magnetically aligning the rotor and stator teeth, which occur when the air gap between the two sets of teeth is minimized and aligned.

Stepper motors are categorized according to their type. Two basic types of motors are

1. Variable reluctance (VR) stepper motors.
2. Permanent magnet (PM) stepper motors.

In *VR motors*, the stator windings are excited in a sequence that will cause the rotor to align to a position that minimizes magnetic reluctance between the stator and rotor. In *PM motors*, the excitation pattern is provided by the permanent magnets. Permanent magnet motors have a smaller step than variable reluctance motors—typical values being 1.8° versus 15° , which makes them more suitable for accurate positioning applications; however, the torque per unit volume of the PM motor is considerably lower than that of the VR motor.

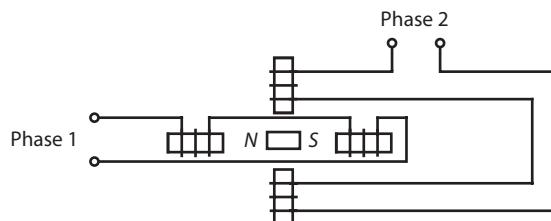
Typical torque ranges for PM motors are usually under 3.5 N·m and for VR motors under 14 N·m. This limits the range of applications for PM motors to a lower torque region than that of VR motors. As a result, PM motors are available in smaller standard sizes (commercially known as size 23 or size 34). For example, a four-phase, size 23 motor typically produces under 0.7 N·m of torque with a speed range of up to 30,000 steps per second (sp), whereas a size 34 motor produces roughly three times the torque at one third of the speed.

For a majority of actuation applications, stepper motors provide a low-cost alternative. The major component of cost is the drive circuit. They are extremely well suited for use in open-loop

applications due to their accuracy and noncumulative position-error characteristics. Since the stepper motor is inherently a discrete device, it is easy to control from a digital computer algorithm, stability is rarely an issue, and the brushless design results in less wear. Compared to DC servo motors, stepper motors produce considerably less torque, lower speeds, and higher vibrations; however, for many applications, their benefits outweigh their drawbacks.

The operating principle of a permanent magnet stepper motor is illustrated in Figure 4-6. The stepper motor consists of a stator with a number of poles. Four such poles are shown in the figure. Each pole is wound with a field winding—the coils on the opposite pairs of poles being in series. The stator shown here has two sets of windings showing phase 1 and phase 2. Each pole in the stator is separated by the adjacent pole by 90° . The rotor has a two-pole permanent magnet. Current is supplied from a DC source to the windings through switches in an appropriate sequence. The rotor will move to line up with the stator.

FIGURE 4-6 STEPPER MOTOR PRINCIPLE



4.2.1 Modeling Approach

This section presents the modeling and simulation for an eight-wire (four-phase), size 23 PM stepper motor with a resolution of 1.8° per step and a 0 to 1000 step per second, sps, speed range. The motor is directly attached to a load having a total inertia value (including the rotor mass) of 0.04 kg-m^2 and a total viscous damping factor of 0.5 Nm/rad/s . The motor is driven by a four-phase driver which produces 20-volt and 2-amp maximum pulses to each of the four-phases sequentially.

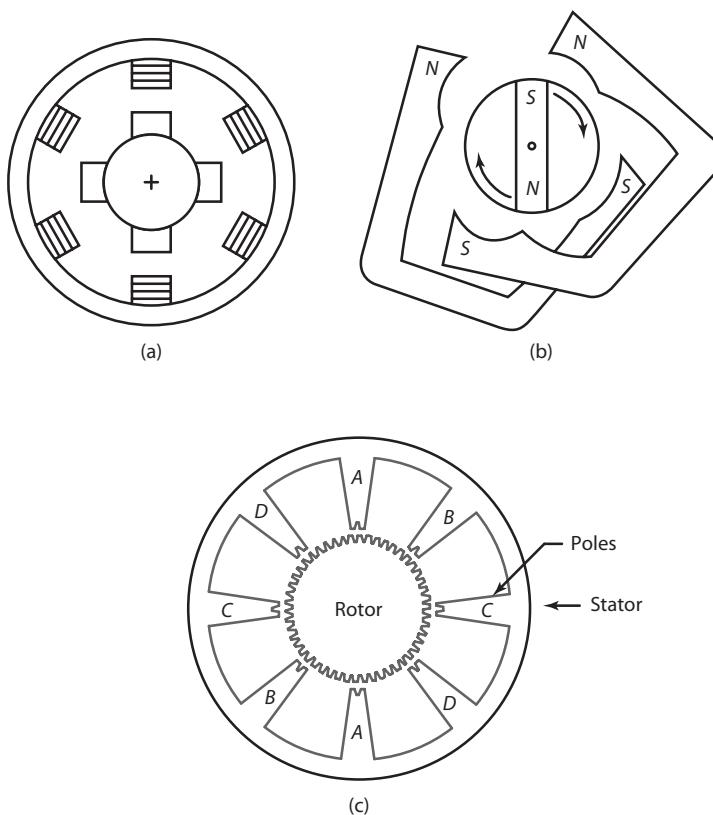
The dynamic performance of the stepper motor system (drive, motor, and load) is simulated in three operating ranges: single step, low speed, and high speed. A four-phase, 1.8° PM stepper motor has eight stator poles with two or more teeth per pole and a 50 tooth rotor. Each pole has one winding which produces a magnetic flux into or out of the rotor, depending on the direction of the current flow.

The stator–rotor configuration is presented in Figure 4-7. The four-pole pairs (phases) are labeled A, B, C, and D. From Figure 4-3, it can be seen that clockwise phase excitation (A, B, C, D) results in counterclockwise rotor motion and counterclockwise phase excitation (A, D, C, B) in clockwise rotor motion. Since all four phases are identical, the electromagnetic torque produced by one phase is first modeled.

The total electromagnetic torque produced by the four phases is obtained by copying the one-phase model three times and summing the four individual phase torque's.

The electromagnetic torque produced by the motor is applied directly to the load with no gear reduction present. The load is modeled as a lumped inertia damper which includes the motor contributions as well as those of the load. The load model is forced by the difference between the applied electromagnetic torque from the motor and the reaction torque from the load. The load model produces two outputs: rotor speed and rotor angle, which are fed back and used in the motor model.

The drive circuit is modeled as a pulse generator with four sequentially triggered phases so that only one phase is on at any given time. The drive circuit model assumes ideal switching between phases and does not model the L/R time constants or the transistor switching behavior. The model

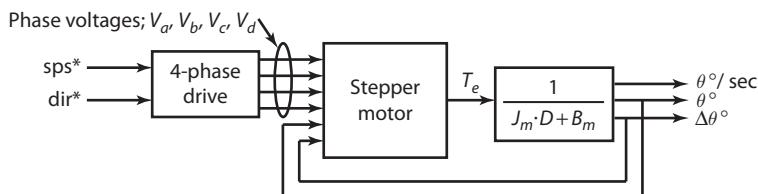
FIGURE 4-7 STEPPER MOTOR CONFIGURATION

is suitable for use in the present application; however, more detail could be easily included if required. The drive circuit has two command inputs, a step per second command, sps^* , and a direction command, dir^* . It produces four voltage outputs, one for each phase of the motor.

The top level block diagram of the stepper motor system is presented in Figure 4-8. The system consists of three components; the drive, the stepper motor, and the load.

The sps^* command is selectable in the 0 to 1000 sps range and the dir^* command is also selectable and has two states, 1 or -1 , where 1 forces clockwise rotor rotation and -1 forces counterclockwise rotation.

The digital motion control of the stepper motor requires that the number and the frequency of pulses are calculated by the computer and sent to the stepper motor to produce the required motion.

FIGURE 4-8 STEPPER MOTOR SYSTEM TOP-LEVEL BLOCK DIAGRAM

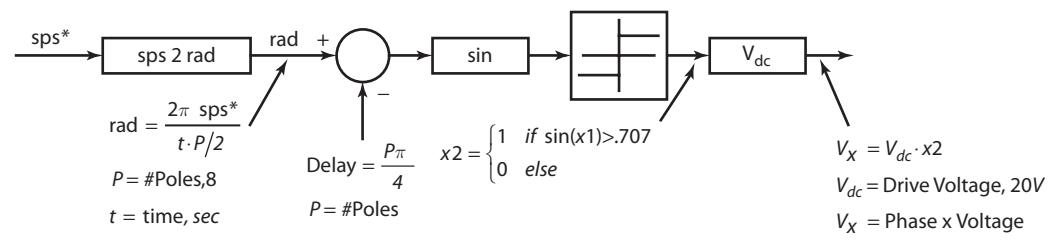
4.2.2 Drive Equations and Block Diagram Model

For a specified steps per second, one phase of the four-phase drive model produces an on-voltage pulse at a rate of $sps^*/4$ times per second. The division by four accounts for the number of phases. The duration of time that the pulse is “on” is $1/sps^*$ seconds.

For example, the phase voltage corresponding to an $sps^* = 8$ of phase A during a 1-second time span is high between times 0 and 0.125 seconds and between 0.5 and 0.625 seconds. The phase voltages for phases B, C, and D are identical in shape but are delayed by $1/sps^*$, $2/sps^*$, and $3/sps^*$ seconds, respectively.

Figure 4-9 presents the block diagram used to model the drive circuit behavior. The drive model produces positive, valued, and sequential pulses which will move the rotor in one direction. To achieve bidirectional movement, the phase voltage signal, V_x , is multiplied by the direction reference, dir^* .

FIGURE 4-9 DRIVE CIRCUIT MODEL



4.2.3 Motor Equations and Block Diagram Model

The PM motor consists of four identical phases allowing the motor model to be developed based on a model of one phase which is then tripled for the remaining three phases. The one phase model operates as follows.

As a voltage pulse occurs from the drive circuit, the stator winding produces a current due to the difference between the voltage pulse and the back emf voltage. Neglecting the mutual inductance, the winding is modeled as a self inductance (due to changes in the phase current) and a resistance.

The resulting phase-current model is represented by Equation 4-17. For brevity, the time dependence has been dropped on the current and voltage signals.

$$i = \frac{1}{R + L \cdot D} \cdot (V_x - V_{bemf}) \quad (4-17)$$

where

i = phase current, amps (A)

R = phase resistance, ohms (Ω)

L = Phase inductance, Henry (H)

D = derivative operator

V_x = supply voltage from driver, volts DC (V)

V_{bemf} = back emf voltage, volts (V)

The rotor motion creates a flux linkage in the windings. This causes a back emf voltage, which is proportional to the rotor speed and varies periodically with the rotor position according to Equation 4-18.

$$V_{\text{bemf}} = -K_{\text{bemf}} \cdot \dot{\theta} \cdot \sin(r \cdot \Delta\theta) \quad (4-18)$$

where

V_{bemf} = back emf voltage, volts (V)

K_{bemf} = back emf constant, volts/radians (V/rad)

$\dot{\theta}$ = rotor speed, radian/second (rad/s)

r = number of rotor teeth

$\Delta\theta$ = delta rotor angle, radian, range: 0 to 1.8°

The self inductance, L , used previously, also varies with the delta rotor position. The variation is periodic and represented by Equation 4-19.

$$L = L_1 + L_2 \cdot \cos(r \cdot \Delta\theta) \quad (4-19)$$

where

L = phase self inductance, Henry (H)

L_1, L_2 = constants, Henry (H)

r = number of rotor teeth

$\Delta\theta$ = delta rotor angle, radian, range: 0 to 1.8°

Similar to a DC motor, torque in a PM stepper motor is proportional to the phase current by a torque constant due to the constant flux from the permanent magnet; however, it differs due to its dependence on the flux produced by the phase current, which varies periodically with the rotor position.

Equation 4-20 presents the electromagnetic torque equation for the PM stepper motor.

$$Te = -K \cdot i \cdot \sin(r \cdot \Delta\theta) \quad (4-20)$$

where

T_e = electromagnetic torque, Nm

K = torque constant, Nm/A

i = phase current, amps

r = number of rotor teeth

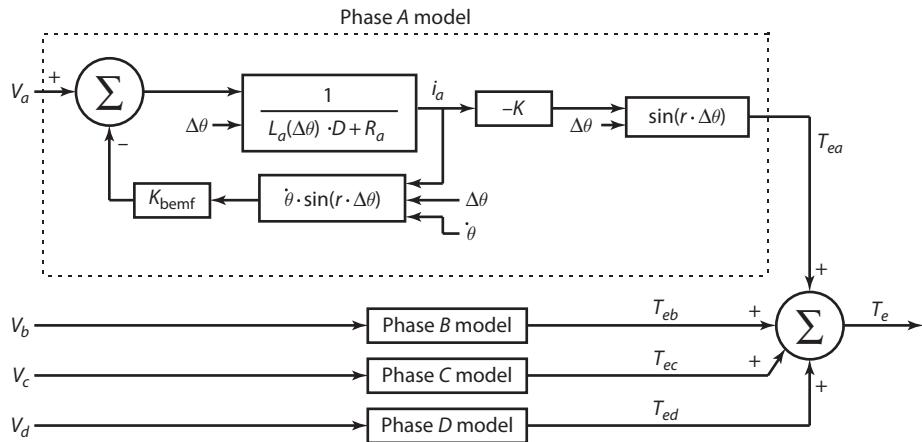
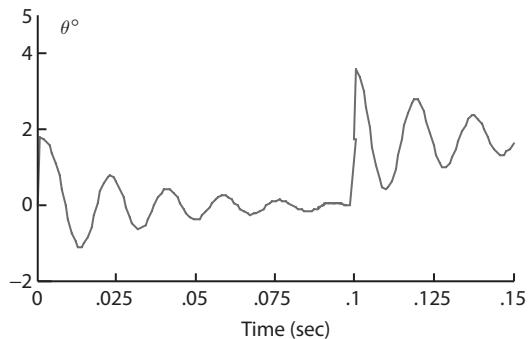
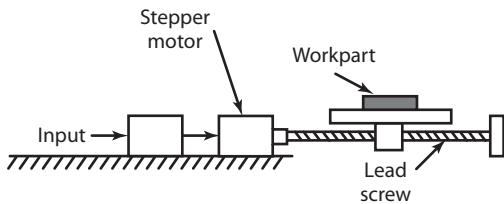
$\Delta\theta$ = delta rotor angle, radian, range: 0 to 1.8°

The complete block diagram model for the four-phase PM motor is presented in Figure 4-10. The contents of the phase B , C , and D blocks are identical to that of the phase A model.

The contents of the phase B , C , and D model blocks are copies of the phase A model. They are represented as top-level blocks here for brevity.

A typical small-signal angle response for this stepper motor is presented in Figure 4-11.

Figure 4-12 illustrates the angular motion of the rotor as it travels over a 1.8° interval. The ringing effect (a common feature of the stepper motor response) can sometimes be attenuated electrically or by the load; however, it is difficult to completely remove. Therefore, when applying a stepper motor actuator, you should expect this ringing behavior and factor it into the system design.

FIGURE 4-10 BLOCK DIAGRAM MODEL OF FOUR-PHASE PM STEPPER MOTOR**FIGURE 4-11 FOUR-PHASE PM STEPPER MOTOR MODEL RESPONSE****FIGURE 4-12 MOTOR AND LEAD SCREW ARRANGEMENT IN A POSITIONING SYSTEM**

4.2.4 Positioning System Using Stepper Motor

A positioning system normally uses a stepping motor and a lead screw arrangement. In a computer numerically controlled (CNC) machine tool, the stepping motor is driven by a series of electrical pulse signals that are transmitted from the input module. Each pulse causes the motor

to rotate a fraction of one revolution, called the step angle. The allowable step angles must conform to the relationship

Step Angle

$$\theta = 360/n_s$$

where

θ = step angle, degrees

n_s = the number of step angles for the motor

Angle of Rotation If the motor is directly connected to the screw without a gear box, the angle of rotation of the leadscrew is given by

$$A = n_p \theta$$

where

A = angle of leadscrew rotation, degrees

n_p = number of pulses received by the motor

θ = step angle, here defined as degrees/pulse

Distance Moved The movement of the table in response to the rotation of the lead screw is calculated from

$$S = pA/360$$

where

S = position relative to the starting position, mm

p = pitch of the lead screw, mm/rev

$A/360$ = the number of revolutions (and partial revolutions) of the lead screw

Number of Pulses From the above equations, the number of pulses required to move a predetermined position can be found by

$$n_p = 360 S/p\theta$$

Rotational Speed The pulses are transmitted at a certain frequency, which drives the worktable at a specific velocity. The speed of the leadscrew depends on the frequency of the pulses

$$N = 60f_p/n_s$$

where

N = rotational speed, rev/min

f_p = pulse frequency (pulses/sec)

For a two-axis table with continuous path control, the relative velocities of the axes are coordinated to achieve the desired travel direction.

The table travel speed in the direction of lead-screw axis is determined by the rotational speed as

$$v_t = N \cdot p$$

$$f_r = N \cdot p$$

where v_t is the table travel speed in mm/min, which also can be considered as feed rate (f_r) and p is the pitch of the leadscrew (mm/rev).

EXAMPLE 4.3

A machine table driven by closed-loop positioning system consists of a servo motor, lead screw, and an optical encoder. The lead screw has a pitch of 0.500 cm and is coupled to the motor shaft with a gear ratio of 4:1 (four-turns of motor for one turn of lead screw). The optical encoder generates 150 pulses/rev of the lead screw. The table has been programmed to move a distance of 7.5 cm at a feed rate of 40 cm/min. Determine the following.

- How many pulses are received by the control system to verify that the table has moved exactly 7.5 cm?
 - Pulse rate.
- (Note that pitch is the axial distance traveled for one revolution of the screw.)

Solution

Lead-screw pitch = 0.5 cm/rev.

Motor rpm = 4 * lead screw rpm

Lead screw generates 150 pulses/rev

Distance to be moved, S = 7.5 cm

Feed rate = 40 cm/min.

Time required to travel 7.5 cm (t) = 0.188 min

If the lead-screw pitch is 0.5 cm and the distance traveled is 7.5 cm, it will cause 15 revolutions of the screw. Each revolution of the screw generates 150 pulses. Thus,

$$7.5 \text{ cm}/0.5 = (15 \text{ rev})^* (150 \text{ pulses}/\text{rev}) = 2250 \text{ pulses}$$

$$\text{Pulse rate} = 2250 \text{ pulses}/0.188 \text{ min} = 12000 \text{ pulses}/\text{min} \text{ or } 200 \text{ pulses/sec}$$

4.3 Fluid Power Actuation

The field of mechatronics has benefited by the developments in fluid power actuators. Fluid power actuators in the form of totally integrated packages with intelligent controls, energy-efficient power sources, and computer-controlled sensing devices are currently in use. In most of the applications, the control speed is of main concern, which (to a large extent) is achieved by developments in electrohydraulic servo valves, programmable controllers, interface components, and systems with hardware-in-the-loop. Modern control systems have contributed to flexibility in controlling fluid power elements. The development of electrical torque motors for electrical servo valves has addressed the need of converting electrical signals into hydraulic signals. Fluid power systems are extensively used for driving high-power machine tools, such as robots, as they can deliver a higher amount of power while being relatively small in size.

The three main components of a fluid power control system are

- Fluid power actuator
- Servo valve
- Load

A valve can be actuated by electromechanical actuators, such as solenoids and torque motors. For on/off applications, solenoids are preferable, whereas for continuous control, torque motors are used.

4.3.1 Control Systems in Fluid Power

Figure 4-13 presents a basic diagram of a computer controlled fluid power system that displays the components of sensing, controlling, and actuating operations.

A fundamental component in a fluid power system is the valve, which is the actuator mechanism. The valve can be positioned manually or automatically. The mechanism shown is a double-acting actuator, where the fluid pressure acts on both sides of the piston. The fluid flow at the ports of the actuator is regulated by a servo valve.

Spool valves are extensively used in fluid power systems. Input displacement applied to the spool rod through an electrically operated torque motor can regulate the flow rate to the main fluid power actuator by sending an appropriate pressure difference across the actuator lines. The spool movements in the valve assembly are limited to very small displacements. In the null position, the input line is blocked so that equal pressure exists on both sides of the actuator piston. When the valve stem is moved to the right, oil at pressure P_S enters the actuator cylinder to the left of the piston.

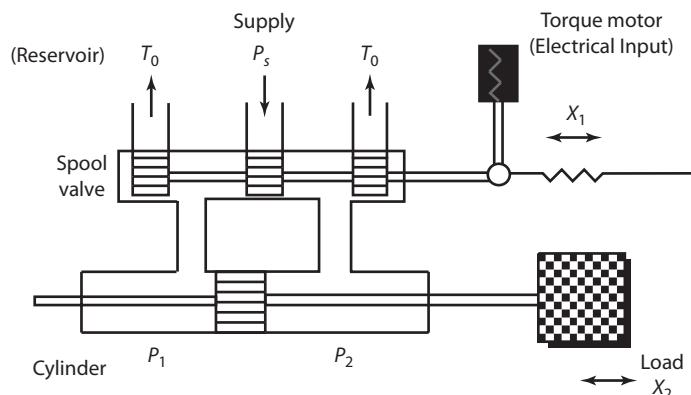
Assuming incompressibility of oil, it follows that the flow rate of oil is proportional to the movement of the valve to the left of the actuator piston. Referring to the Figure 4-13, the pressure difference across the piston for displacement to the right is given by Equation 4-21.

$$P_d = P_1 - P_2 \quad (4-21)$$

Consequently, it creates a force on the piston:

$$F = AP_d = A(P_1 - P_2) \quad (4-22)$$

FIGURE 4-13 VALVE ACTUATOR MECHANISM



Flow rate, q , into the left side of the piston obeys

$$q = k_1 x_1 - k_2 P_d \quad (4-23)$$

Here x_1 is the movement of the valve about the null position, and k_1 and k_2 are valve constants.

Equation 4-23 states that the flow rate increases as the valve stem exposes more of the hydraulic fluid pressure line to the chamber but decreases as the back pressure increases. The fluid flowing into the left must be balanced by the movement of the piston to the right.

$$q = A \frac{dx_2}{dt} = k_1 x_1 - k_2 P_d \dots (a) \quad (4-24)$$

$$P_d = \frac{1}{k_2} \left(k_1 x_1 - A \frac{dx_2}{dt} \right) \dots (b)$$

$$F = AP_d \dots (c)$$

$$F = \frac{A}{k_2} \left(k_1 x_1 - A \frac{dx_2}{dt} \right) \dots (d)$$

The load is balanced by the force of the piston. Inertia of the moving parts of the actuator is modeled as mass, M , and the equivalent viscous damping constant as, f .

$$F = M \frac{d^2 x_2}{dt^2} + f \frac{dx_2}{dt} \quad (4-25)$$

Equating Equations 4-24(d) and 4-25, we get

$$M \frac{d^2 x_2}{dt^2} + f \frac{dx_2}{dt} = \frac{A}{k_2} \left(k_1 x_1 - A \frac{dx_2}{dt} \right) \quad (4-26)$$

$$M \frac{d^2 x_2}{dt^2} + \left(f + \frac{A^2}{k_2} \right) \frac{dx_2}{dt} = \left(A \frac{k_1}{k_2} \right) x_1$$

Taking the Laplace transform, we have

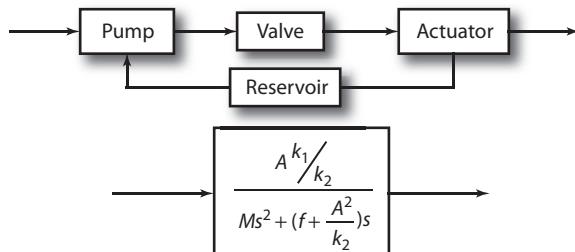
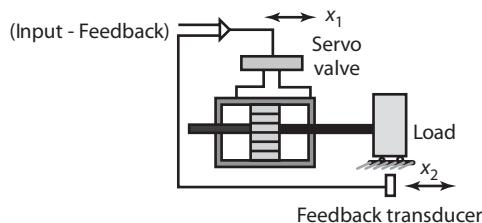
$$\left[Ms^2 + \left(f + \frac{A^2}{k_2} \right) s \right] x_2 = A \left(\frac{k_1}{k_2} \right) x_1 \quad (4-27)$$

$$\frac{x_2}{x_1} = \frac{A \frac{k_1}{k_2}}{Ms^2 + \left(f + \frac{A^2}{k_2} \right) s}$$

The relationship between input and output is described by a second-order differential equation.

Figure 4-14 shows the block diagram of the combined valve actuator system against a load. The values of k_1 and k_2 can be found from the linearized valve characteristics which are predetermined.

Figure 4-15 shows a fluid power system using a position feedback. If the input is moved by a certain amount, the amplifier is driven by the corresponding voltage, and the amplifier voltage excites the solenoid valve winding, which causes the valve stem to move by that amount. The movement of the valve causes the load to move by an amount x_2 . This movement causes the feedback potentiometer to move a distance x_2 .

FIGURE 4-14 BLOCK DIAGRAM OF THE COMBINED SYSTEM**FIGURE 4-15 FLUID POWER ACTUATOR AND SERVO WITH POSITION FEEDBACK**

At this point, the valve is returned to the null position and the motion ceases. Using the information in the previous equations, an overall transfer function of the system can be derived, and the system can be modeled for appropriate damping characteristics.

Fluid power systems can be used in position-control modes or velocity-control modes. The modeling procedure described is for a position-control system with the feedback transducer moving the same electrical distance as the command transducer and the load following it. In a velocity-control system, if the fluid power actuator slows down because of an increase in load, the tachometer voltage is reduced, thereby nullifying the command voltage. When higher speed is desired, the command voltage is increased. The higher command voltage then produces more flow to overcome internal leakage of the hydraulic components. If the speed of the load is decreased, the voltage from the electrical control is reduced. This reduces the amplifier error signal and input to the torque motor. This action results in a proportionate valve opening and decreases the fluid flow.

The servo valve is critical to the proper operation of the system. The dynamic performance depends on the time response of the servo valve. This information is available to the designer as a plot of valve response against signal frequency. For fluid power system design, the general procedure is to use well established linear-analysis methods to calculate system characteristics. The information obtained using transfer functions provides performance values at a particular operating point.

Nonlinear operation is prevalent in fluid power. Nonlinearities occur due to resolution errors and hysteresis. These are usually the major causes of position inaccuracy. Digital simulation allows the use of mathematical models of nonlinear differential equations, nonlinear friction, switching functions, as well as other motion profiles as inputs and the outputs (such as position, velocity, pressure, and flow) from the beginning to the end of the cycle.

4.3.2 Fluid Power Actuators

The fluid power actuator is either a fluid power cylinder for linear motion or a rotary-type motor for angular motion. Fluid power actuators make use of incompressible fluids and are capable of providing a high horsepower-per-unit volume ratio. Earlier, Figure 4-13 provided a simple sketch of the hydraulic actuating system. The double acting hydraulic piston is the principal moving part in the hydraulic system. Fluid can flow into the left side and can exit out of the right side or vice versa, resulting in a movement of the piston to the right or left respectively.

As shown in Figures 4-13 and 4-14, the control over the direction of fluid flow is accomplished by the servo valve. A high-precision electric motor moves the valve piston incrementally, allowing the fluid to flow from the source to the actuator through one port and returning to the valve through another port. The ideal hydraulic rotary actuator provides shaft torque, T , proportional to the differential pressure, ΔP , across the servo valve.

$$T = kD\Delta P \quad (4-28)$$

where

k = proportionality constant relating torque and differential pressure

D = displaced volume measured in mm^3

Fluid power actuators are used for precise linear motion. They often can be applied more easily than electrically operated actuators. Their prime applications are in automobiles, ships, elevators, and airplanes. Fluid power drives have a substantially higher power-to-weight ratios, resulting in higher machine-structure-frame resonant frequencies for a given power level. Fluid power systems can be directly coupled to the load without the need for intermediate gearing. Since the fluid power actuators use the hydraulic power of a pressurized liquid, they are capable of providing very high forces (and torque) at high power levels. Fluid power actuating systems are much stiffer than electrical actuation, resulting in greater accuracy and better frequency response. Fluid power drives give smoother performance at lower speeds and have a wide response range.

Figure 4-16 shows the photograph of a digital hydraulic linear positioner. This actuator uses a stepper-motor controlled digital spool valve and a magnetostrictive linear displacement transducer to monitor the position of the actuator.

FIGURE 4-16 DIGITAL HYDRAULIC LINEAR POSITIONER



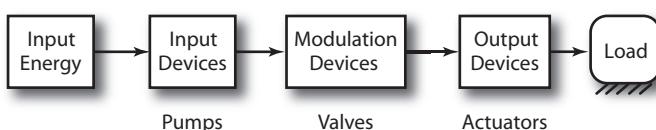
Victory Controls, LLC.

4.4 Fluid Power Design Elements

The basic fluid power system consists of a source of input energy and a suitable device for energy input, energy output, and energy modulation. Transmitting fluid power requires a pump to convert the mechanical energy into fluid energy. Proper devices are needed for the modulation of fluid actuators. The primary source of input energy is quite often an electric motor, an internal combustion engine, or another type of mechanical device that can supply force and motion to operate the pumps. The pump supplies hydraulic fluid or pneumatic pressure to the system. In other words, fluid power can be defined as the power transmitted and controlled through the use of a pressurized fluid.

The block diagram for a fluid power control system is shown in Figure 4-17. A fluid power system consists of three devices:

FIGURE 4-17 FLUID POWER SYSTEM



- Energy input device
 - Energy modulation device
 - Energy output device

The following sections present a description of each device.

4.4.1 Fluid Power Energy-Input Devices

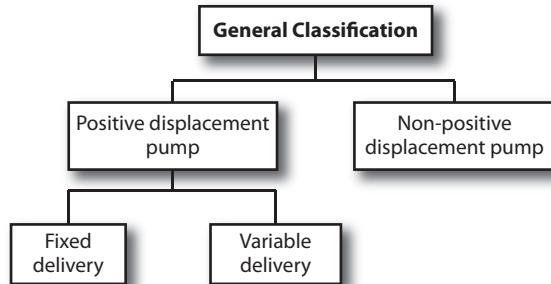
The input devices, such as pumps, are the primary source of fluid power energy creation. The hydraulic pumps are used as devices that convert mechanical force and motion into actuating power using fluid power circuits. Hydraulic pumps create flow of the fluid under consideration and develop pressure. Pressure is the direct result of resistance to flow encountered by the fluid. The pressure can be varied by providing a different load to the system or by pressure regulating devices. The basic classification of fluid power pumps is typically

- General classification
 - Classification based on design features.

A general classification of fluid pumps is as a positive displacement and non-positive displacement, as shown in Figure 4-18. The classification is based on the displacement of the fluid. *Displacement* is the actual volume of fluid displaced during a cycle of the fluid power pump.

Positive-Displacement Pumps A positive-displacement pump has a small clearance between the stationary and rotating parts. The positive-displacement pump is able to push a definite volume of fluid for each cycle of pump operation at any resistance encountered. Because of its simplicity of use, positive-displacement-type fluid pumps are increasingly used in the fluid power industry. The further subdivision of positive-displacement pump is as (i) fixed-delivery and (ii) variable-delivery types.

The *fluid delivery* of a positive-displacement pump depends on the working relationships of internal elements. Volumetric output of the fluid remains constant for a given speed of the pump.

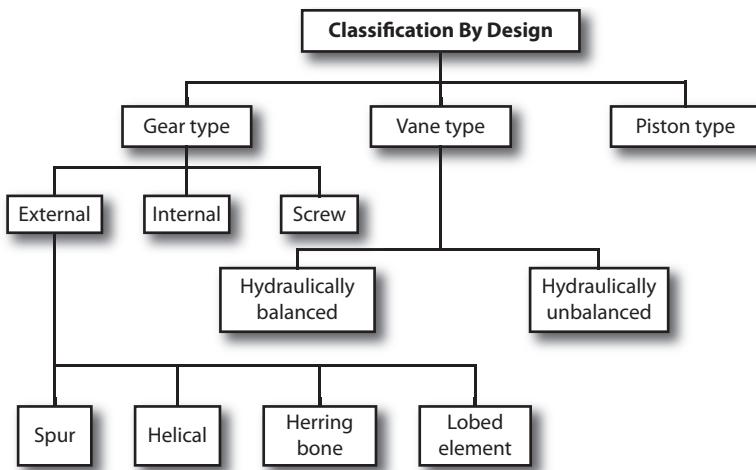
FIGURE 4-18 CLASSIFICATION BASED ON DISPLACEMENT

Reference: Henke.

Only by varying the speed of the pump can the output of the pump be changed. However, the fluid delivery in a variable pump can be changed by altering the physical relationship of the pump elements, keeping the speed at a constant level.

Non-Positive-Displacement Pumps A non-positive-displacement pump has a large clearance between the rotating and stationary parts. The total volume of the fluid displaced from the pump depends on its speed and resistance faced at the discharge side of the pump unit. In applications which deal with a low-pressure and high-volume flow situation, non-positive-displacement pumps are used.

Classification of Pumps by the Design Features Another classification of pumps is according to the specific design of the element used to create flow of the fluid, as shown in Figure 4-19. Most pumps used in fluid power applications are of the rotary type, in which a rotating assembly of components

FIGURE 4-19 CLASSIFICATION BASED ON DESIGN FEATURES

Reference: Henke.

carries the fluid from the inlet side to the outlet side. Continuous rotary motion of the rotating assembly causes the rotary pump to operate. The three most common pumping mechanisms used in rotary pumps are (a) gear-type pumping mechanism, (b) vane-type mechanism, and (c) piston-type mechanism.

Rotary Pumps with Gear-Type Mechanism The design of rotary gear pumps consists of the meshing of two or more gears which are engaged in a closely fitted housing. Gear pumps normally have a flow rate of around 0.7 m/min and a delivery pressure of up to 217 atm. The gear pumps can be categorized into the following types:

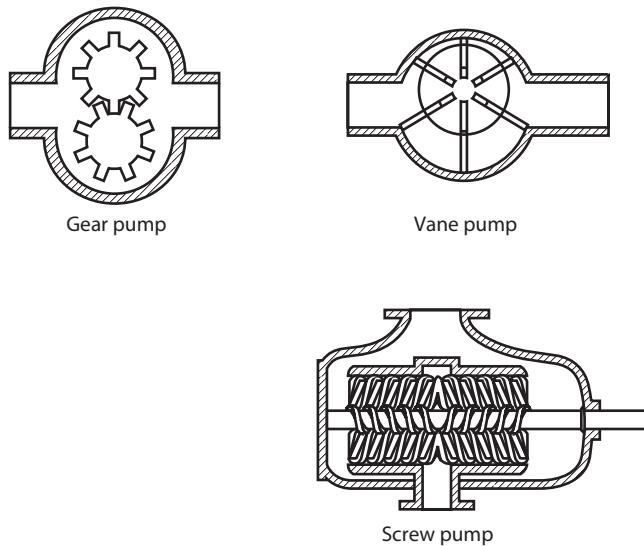
- External gear pump
- Planetary or internal gear pump
- Screw pump (with axial flow)

External Gear Pumps External gear pumps are designed with two gear combinations: one gear mounted on the drive shaft, while the second gear is mounted on the driven shaft. The gears are designed to rotate in opposite directions and mesh at a point in the housing between the inlet and outlet ports. The pumping action of the external gear pump is caused by the rotation of the gears. As the gears in contact rotate, the spaces between the teeth fill up with fluid which is carried around in small quantity between the gear teeth and the pump casing. As the pumping action continues, the gears mesh again, and the fluid is squeezed out to be discharged from the pump.

The different gear configurations used in external gear pumps are (a) spur gears, (b) helical gears, (c) herringbone gears, and (d) gear pump with lobed elements. Pumps with helical and herringbone gear features have smoother and quieter operation than the spur gears. The external gear pumps are also designed to deliver larger quantities of fluid with less fluctuation. The lobe-shaped rotating element is a modification of the external gear pump.

Internal Gear Pumps (Planetary) The internal gear pump is a modification of the external gear pump and uses two gears. The spur gear is mounted inside a larger ring where the smaller spur gear is in mesh with one side of the larger gear. It is kept apart by a separator on the other side. As in the external gear pumps, the fluid moves from the suction port to the discharge port by the entrapment action between the meshed teeth of the rotating gears. Input energy can be applied either to the inner ring gear or to the outer ring gear. It is also to be noted that the direction of rotation of both gears is the same. Another form of the internal gear pump is the Gerotor type pump. There is a special tooth form on the inner gear. The inner gear is the driver and has one tooth less than the outer gear. The two gears are sized in shape so that part of the periphery of the inner gear maintains contact with the surface of the outer gear element at all times. Another requirement is that there should be a seal between the inlet and outlet port. The volume of fluid delivered by the pump is a function of the space formed in the external rotor. A smooth fluid discharge is possible by the gradual opening and closing of the extra tooth space. Internal gear pumps are normally silent in operation.

Screw Pumps Two basic types of screw pumps used in the industry are the single screw pump and the multiple screw pump. A single screw pump consists of a screw (helical gear) that rotates eccentrically in an internal container. Multiple screw pumps consist of two or more screws that mesh as they rotate in a closed casing. When the driver rotates, a volume of fluid from the inlet is trapped between the contact points of the screws and the space between the screws and the outer casing. This rotation of the screws makes the fluid volume, which is trapped, move linearly along the screw axis until it is pushed through the outlet of the pump. It is obvious that the flow through a screw pump is in the direction of the driving screw. The output of the screw pump is normally smooth, non-pulsating, and with a very low noise level. Figure 4-20 provides an example of the constructional details of pumps.

FIGURE 4-20 EXAMPLES OF FLUID PUMPS

Pumps with Vane-Type Mechanisms Two different types of rotary vane pumps are commonly used in fluid power systems: (a) hydraulically unbalanced type and (b) hydraulically balanced type. These pumps consist of a cylindrical motor fitted with movable vanes which extend out from the outer boundary of the rotor. The main rotor rotates in an oval shaped inner area of the pump housing. When the vanes rotate and start moving from the point of minimal clearance between the rotor and the housing, fluid is sucked from the intake port of the pump section and discharged into a variable space between the rotor and the housing. As the vanes rotate and pass through the point of largest clearance between the rotor and the housing, the fluid is compressed and later discharged into the outlet port side of the vane pump.

In the unbalanced vane pump, the rotor revolves with the shaft mounted eccentrically in relation to the vane track housing. The suction action causes a large unbalanced load, because the suction port is almost diametrically opposite the discharge port. The mere existence of this unbalanced load causes the shaft and bearing to be sufficiently strong to prevent component failure.

The balanced vane pumps differ from the unbalanced type in design features. In balanced vane pumps, there are two intake and two outlet ports diametrically opposite each other. This kind of design arrangement of pressure ports opposite each other causes a balanced condition. In a balanced vane pump, the vanes are hydraulically balanced by the discharge pressure and held against the vane track by the centrifugal force.

Pumps with Piston-Type Mechanisms Piston pumps have a special feature where a number of small pistons reciprocate at high speeds. The fluid pressure generated is usually in excess of 200 atm. The main difference between the axial and rotary piston pumps is the operating position and the shape of their pistons. Piston pumps convert rotary shaft motion into a radial reciprocating motion.

Rotary Piston Pumps Radial piston pumps have a cylindrical element that rotates about a stationary central pintle element. The cylindrical element contains seven or more radial bores fitted with pistons that reciprocate in or out as the cylinder rotates. The central pintle also includes inlet and outlet ports

that connect with the inner openings of the cylinder bore so that the pintle can direct the flow in and out of the cylinder. A rotor and its supporting members move eccentrically with respect to the cylinder block.

When the driver rotates the cylinder block, the individual pistons travel outward while the cylinder bores pass the inlet ports of the pintle, drawing in fluid. When the pistons pass the maximum point of eccentricity, they are moved inward by the reaction ring. This causes the fluid to enter the discharge side of the pintle. The stroke of each piston can be changed by the eccentricity of the rotor with respect to the pump shaft. The degree of eccentricity between the cylinder and the rotor governs the rate of delivery of the fluid pump.

Axial Piston Pumps Axial piston pumps have pistons that move axially in the cylinder barrel. The cylinder block in the pump has a series of cylindrical bores with pistons that move in and out. The drive shaft causes the pistons and the cylinder block to rotate at the same speed. As the block rotates, each piston element moves in and out of its cylinder—the length of stroke depending on the angle of the cylinder block with reference to the drive plate. When each piston starts reciprocating, fluid is drawn into the cylinder bore through the valve plate. On the return stroke of the piston, fluid is forced out through the valve plate under pressure due to restriction of flow. A number of alternate design features exist in axial piston pumps. The bent axis (fixed delivery type) has a fixed angle of the cylinder block with respect to the housing. The bent-axis variable displacement pump has a cylinder block mounted in a yoke that can be positioned at various angles. The pump displacement is determined by the relative position of the cylinder block and the drive shaft. In the case of the in-line axial piston pump, the cylinder block is parallel to the drive shaft. The stroke length of the piston is determined by the angular position of the swash plate. In-line axial piston pumps are available in fixed and variable displacement models. The variable displacement swash-plate models have the swash plate mounted so that its angle can be altered. A fixed displacement pump has a swash plate mounted at a fixed angle within the housing.

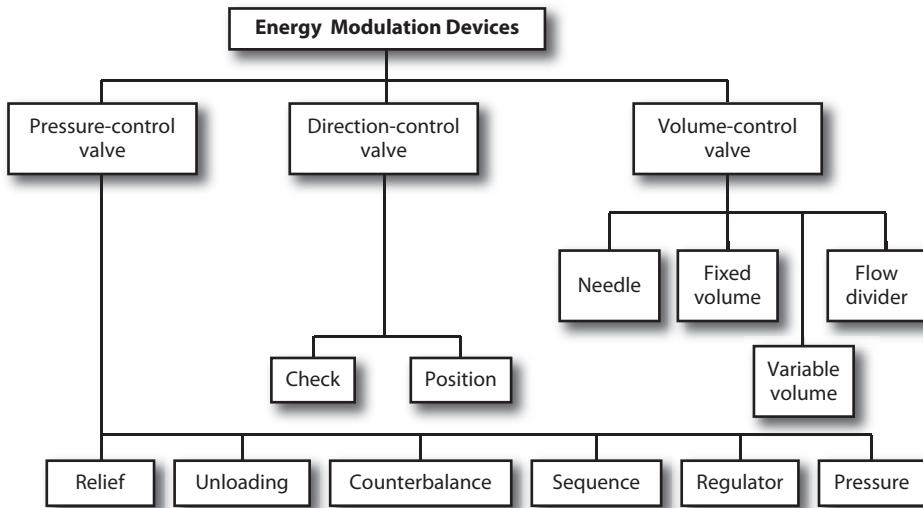
4.4.2 Energy Modulation Devices (Valves)

The energy modulation devices in fluid power systems control pressure, direction, and the rate of flow of fluid. Their control functions in a fluid power circuit are restricting or directing the rate of flow of fluid within a circuit and modifying the energy or pressure level of a fluid flow by means of regulating either flow or pressure. In general, all fluid power control valves are combinations of the basic control configurations. Those valves in a circuit that regulate pressure or create required pressure conditions are referred to as pressure-control valves. Those valves that direct, divert, combine, or restrict flow in a circuit are called directional-control valves. Volume-control valves are those valves which regulate the amount of fluid flow. Valves are usually named according to their construction, which can vary from a simple ball and seat to a multi-element, spool-type valve coupled with electrical controls. The circuit control features can vary with the nature of application. Various classifications of energy modulation devices are shown in Figure 4-21.

Pressure-control valves Pressure-control valves are controlled and modulated by pressure of the fluid in the fluid power circuit. These control valves either limit the pressure to various parts of the circuit or direct fluid to different parts of the circuit whenever the pressure level in one part reaches a predetermined set value.

Pressure-control valves are classified as

1. Relief valves
2. Unloading valves

FIGURE 4-21 ENERGY MODULATION DEVICES

3. Sequence valves
4. Counterbalance valves
5. Regulator valves
6. Pressure switch

Relief Valves *Relief valves* mainly protect a fluid power circuit from maximum pressure. The primary use of a relief valve is to limit the maximum pressure in any part of the fluid power circuit. Relief valves can be considered safety valves and have to be large enough to handle the entire pump-output volume flow. The two types of relief valves are simple and compound.

Simple Relief Valve (Direct Acting) A spring-loaded, simple, direct-acting valve is normally closed until the pressure level exceeds the preset value. When it reaches that critical pressure, it unseats the ball or poppet allowing some fluid to flow. When the line pressure drops, the valve closes. The fluid flow is restored by a direct spring-loaded ball, poppet, or spool, which actuate in order to maintain fluid flow.

Compound Relief Valves (Pilot Operated) The *compound relief valve* is a pilot-operated device and has two stages. In the closed position, the fluid at the system pressure flows through the primary inlet and exits through the primary outlet port. When the system pressure exceeds the setting of the pilot relief valve, the mechanical spring is compressed, unseating the pilot valve and permitting the pressurized fluid to return to the reservoir.

Unloading Valves The main use of an *unloading valve* is to permit a pump to operate at a minimal load. The unloading valve needs an external signal. The fluid delivery is shifted through the secondary port back to the main reservoir whenever sufficient pilot pressure is applied to move the

spool against the spring force. The displaced spool remains shifted by the pilot pressure until the pilot sensing pressure becomes less than the preset spring pressure.

Sequence Valves Quite often in fluid power circuits, it is necessary to move the actuators in a definite sequence of operations. As the name suggests, *sequence valves* are used to control the order of the flow to various parts of the fluid power circuit in a particular order. The sequencing action is caused by requiring the inlet pilot pressure to reach a set pressure level before the valve opens to let off the fluid. As long as the inlet pressure remains above the preset value of the pressure, full pressure is then available at the outlet port. The actuation of the valve is caused by fluid pressure that is generated separately.

Counterbalance Valves (Back Pressure Valves) The main use of the *counterbalance valve* is to prevent the free fall of a load held by the actuator and to develop some line of resistance. The main action of a counterbalance (back pressure) valve results in restricting fluid flow from one port to another port and to maintain a pressure level sufficient to balance a load being held by a cylinder or motor. The basic principle is that the fluid is held under pressure until pilot action overcomes the spring force setting or the counterweight in the valve. At this point, the main spool moves to bypass the return flow internally or externally to the drain.

Regulator Valves Regulator valves are also known as pressure-reducing valves. These devices provide a constant pressure at the outlet port, regardless of the pressure at the inlet port of the valve. The outlet pressure varies with the pressure at the inlet port. The *regulator valve* works by keeping a balance of the upstream pressure against both downstream and spring pressure. If the controlled pressure rises above the desired value (as preset by the spring), the diaphragm rises, thereby reducing the flow to the system and hence its pressure.

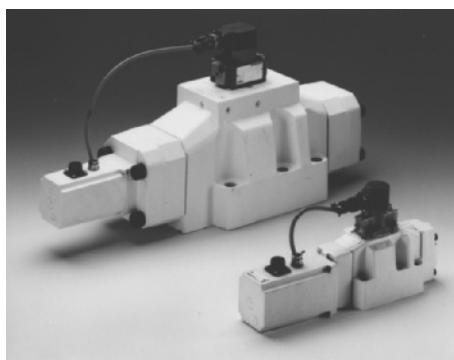
Pressure Switch In many fluid power applications, pressure switches are used whenever an electrical signal is required as the system pressure reaches a certain desired setting. There are two types of designs; (a) piston-type pressure switch and (b) bourdon-tube type switch. These switches are utilized whenever an electrical signal is required for control purposes. When the fluid system pressure reaches the pressure setting as established by the adjustable spring in the switch, an electrical signal is obtained, and the switch is actuated. The electric signal can be relayed to a solenoid valve to change the direction of flow or to actuate a pump.

Directional-Control Valves The use of the directional-control valve is to direct the flow of fluid generated by a fluid source to the various places in the system. The directional-control valve either blocks the flow completely, guides the flow to various branches where fluid power is needed to operate a fluid motor, or actuates a pilot-operated control valve. They may be used for various functions of energizing or de-energizing a fluid power circuit, to isolate a fluid power circuit from a part of the circuit, or to reverse the direction of the flow. They also can be used to combine flow from two or more branches or to separate the flow. Two main categories of directional control valves are *check valves* and *position valves*.

Check Valves for Directional Control Check valves allow free flow of fluid in one direction and restrict flow in the opposite direction. The check valve can be constructed using various blocking devices (namely a swinging disk, a spring seating disk or ball, and a gravity or self seating ball). The pilot-operated check valve allows the free flow in one direction and will only allow fluid flow in the opposite direction (normally blocked) if pilot pressure is applied at the pilot pressure port of the valve.

Position Valves for Directional Control In fluid power circuits, position valves are used to direct fluid to one or more different flow lines, and they do this by being shifted into two or more positions. Depending on the position of the valve, the interconnection of the external ports produces various combinations of flow direction. The numbers of (two-port, three-port, four-port, etc.) and the kinds of positions are added to adequately define the valve as a two-position, three-position, four-position, etc. Position determines the number of alternative flow conditions the valve can provide. These are made possible by the configuration of the spool or the passages of the valve body.

FIGURE 4-22 PHOTOGRAH OF THE PROPORTIONAL VALVE

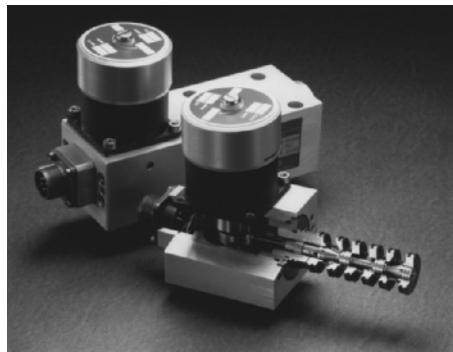


With permission from the Rexroth Corporation, Bethlehem, PA.

The control and shifting of position valves can be done by linking mechanisms, springs, cams, solenoids, pilot fluid pressure, or servomechanism.

Although the spool and piston-type position valves are often used in the fluid power industry, other types (such as the rotary and poppet position valves) are also used. A two-position, three-way sliding spool valve has three external ports used alternately to pressure and exhaust a cylindrical port. Its main use comes in controlling the speed of the fluid power cylinders. If there is a need to position the actuators at intermediate positions, a three-position, three-port-sliding spool will be needed. A two-position, four-way directional-control valve can be used to control the position of double-acting cylinders. Fluid which is at the inlet port is delivered to either of the outlet ports by the movement of the spool as per the sequence.

Figure 4-23 shows a digital valve, which has a combination of three major components: DC stepper motor, rotary-to-linear coupling and four-way spool valve. It provides a digital interface to operate linear and rotary actuators. The four-way spool valve provides the directional and proportional flow control of the fluid media. Rotary-to-linear coupling is arranged to translate the rotary action of the stepper motor into precise spool position. The stepper motor provides a digital means to position the valve spool in precise, discrete increments. Typical application of digital valves is in high-payload carriers, automation equipment, machine tools, and the plastic and textile industries.

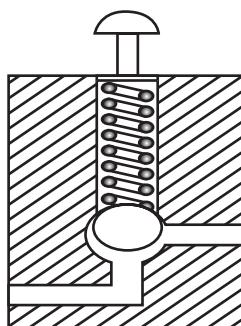
FIGURE 4-23 PHOTOGRAPH OF THE DIGITAL VALVE FOR FLUID CONTROL

Victory Controls, LLC.

Volume-Control Valves The volume-control valves are used to monitor the rate of fluid flow to various parts of a fluid power circuit. Volume-control valves (Figure 4-24) have the role of regulating the speed and functions of fluid actuators by restricting the flow of fluid. Some of the types of volume-control valves are

1. Needle valves
2. Fixed-volume, pressure-compensated valves
3. Variable-volume, pressure-compensated valves
4. Flow divider valves

Needle Valves The basic design of a needle valve is based on a long, tapered point that seats in the valve, which permits a very gradual opening and closing of the passage. The needle valve is not pressure compensated, which means that variations in pressure drop across the orifice will produce definite changes in the rate of flow through the valve.

FIGURE 4-24 SCHEMATIC OF A VOLUME-CONTROL VALVE

Fixed-Volume, Pressure-Compensated, Flow-Control Valves A fixed-volume, pressure-compensated, flow-control valve keeps a constant flow regardless of the variations in the inlet flow to the valve. If the inlet flow rate rises, the mechanism partly closes the pressure-compensated valve in order to reduce the outlet flow. Due to this mechanism, the total volume of fluid through the valve always remains fixed.

Variable-Volume, Pressure-Compensated, Flow-Control Valves A variable-volume, pressure-compensated, flow-control valve is a valve which uses an adjustable volume-control device to adjust the orifice area. Some of the components used in the valves are tapered slots or metering spools. These types of valves maintain a constant flow with varying inlet and outlet pressures.

Flow-Divider Valves The main use of the flow-divider valve is to synchronize the movements of two or more cylinders without having mechanical interconnections between them. This valve handles the flow of fluid in a line and fans out to two or more lines so that each has the same flow rate.

4.4.3 Energy-Output Devices

Fluid power energy-output devices provide either linear or rotary motion through the use of actuators, called cylinders, and fluid motors. Fluid power actuators were illustrated earlier in Section 4.3.2. Fluid actuators use hydraulic power of the order of 35 MPa. This gives the fluid actuators a capability to provide higher torques and forces at a very high power level. A fluid cylinder is a device that converts fluid power into linear mechanical force—into motion. It consists of a movable element such as a piston and piston rod, operating within a cylindrical bore. A fluid motor is a device that converts fluid power into rotary mechanical force and motion. Fluid motors and fluid pumps are similar in many respects, but the fluid motor works in a manner just opposite to the way in which pumps work. Fluid motors use the fluid delivered by a pump to provide rotating force and motion.

Fluid Cylinders The operating principle of the fluid cylinder is that the fluid entering one port drives the movable piston and rod assembly in one direction, while fluid from the other side of the piston is returned back to a reservoir. A single-acting cylinder is controlled by reversing a directional valve and permitting the flow from the pump and cylinder to return to the reservoir. A double-acting cylinder has ports that allow a fluid to enter the cylinder at either end. By forcing fluid to the cap end, the rod will extend while simultaneously discharging fluid back to the reservoir. By reversing the flow, the rod will be retracted. A cylinder can be attached to a load through a variety of mechanical linkages. The designer of a fluid power system decides the type of linkage necessary for a particular application based on design constraints, space, and applications.

Fluid Motors The actuators and motors carry out the opposite functions of fluid pumps. A rotary fluid motor is capable of converting fluid power into rotary mechanical power. A properly controlled motor can produce an output which has reversible and variable speed characteristics. The fluid under pressure acts against the area of the fluid motor in a similar manner as in the fluid cylinder and causes the rotation of the motor shaft. Rotary fluid power motors provide a higher horsepower-to-weight ratio than do other sources of power. The rotary fluid motors have good variable speed and torque characteristics. There are two general classes of fluid motors:

- Fixed-displacement motors
- Variable-displacement motors

Fixed-Displacement Fluid Motor This can deliver a constant amount of fluid for each revolution. However, it has a torque capacity proportional to the pressure applied. The speed of any fixed type of fluid motor depends on the displacement per revolution and the volume of the fluid supplied to it by the pump. Gear, vane, and piston designs are generally found in the design of fixed-displacement fluid motors.

Variable-Displacement Fluid Motor This has the volume of the fluid modulated and is built with a device that can adjust the displacement per revolution. Rotary fluid motors are also classified according to the type of internal element that is directly actuated by the flow. The three most common actuating mechanisms used in rotary fluid motors are the gear, vane, and piston.

Gear Motors Gear-type fluid motors are basically fixed-displacement units, where the speed of rotation depends on the volume of fluid delivered to the motor. Two of the most widely used fluid gear motors are the external gear type and internal gear type. The external gear design consists of a set of machined gears fitted into a closely machined housing. Both gears are driven, but only one gear is connected to the output shaft. Unlike a gear pump, a gear motor must be hydraulically balanced in order to maintain the close tolerances necessary for the fluid motor operation. Hydraulic balancing can be achieved by having passages in the core leading from the inlet and exhaust ports to points diametrically opposed to the inlet and outlet. This prevents an uneven wear and slippage of the gears. The internal gear design consists of a pair of rotating gears—one inside the other. Fluid under pressure enters one side of the motor and causes the outer and inner elements to revolve. During the rotation, as the space increases, the fluid enters from the pump. As it continues and the space decreases, the fluid is exhausted from the motor.

Vane Motors A rotary vane motor is designed so that the rotor and vane are hydraulically balanced with two inlet ports and two outlet ports diametrically opposite to each other. The design of a vane motor has spring or pressure loading to hold the vanes against the vane track at low operating speeds. There is also some oil thickness under vane tips, which is dependent on rotating speed, operating pressure, and fluid viscosity.

Piston Motors Piston-type motors can be classified as either fixed- or variable-displacement units. The two main types of rotating piston motors are axial-piston motors and radial-piston motors. Axial

FIGURE 4-25 PHOTOGRAPH OF THE RADIAL PISTON MOTOR



With permission of The Rexroth Corporation, Bethlehem, PA.

piston motors have the principle of operation of fluid entering a port which pushes against the pistons, causing the cylinder barrel and shaft assembly to turn. As these pistons exhaust the fluid, the other pistons repeat the cycle, providing continuous operation. The radial piston motor has a cylindrical barrel with an attached output shaft to transmit the force imparted to the pistons. The cylinder barrel also has a number of radial bores with each of them fitted with pistons very precisely. When the fluid enters the cylinder bore, the pistons are forced against the thrust ring which imparts a tangential force to the cylinder barrel and shaft assembly, causing it to rotate. Each piston is pushed inward by the thrust ring once it reaches the outlet port, thus pushing the fluid back toward the reservoir.

4.4.4 Control Modes of Fluid Power Circuits

The control of fluid power circuits can be classified in four basic ways. Depending on the control mode, any one or combination of the types shown can be used.

- Manual control
- Mechanical control
- Fluid control
- Electrical control

Manual Control These systems are of either the open or closed center type, which means parallel- or series-connected, respectively. Each position valve, which controls the operation of a fluid motor, is connected in parallel to the next unit. The frequently used position valves have central port openings and are arranged together by having the tank port of each valve connected to the pressure port of the next valve. The fluid delivered by the pump is bypassed to the tank whenever the fluid motors are not in operation. Central port opening valves are used in series connection if pressure distribution is uniform for all valves. Closed center-port systems are used in most applications where pump pressure has to be continuously accessible to the position valves, controlling the direction of the motoring units. In general, manual control systems have wide applications on mobile fluid power devices.

Mechanical Control Systems These are used in conjunction with manual control to produce semi-automatic operation sequences. While manual control is used to initiate the machine operation, the mechanical controls are aimed at controlling the automatic part of the cycle. Out of the above two methods to operate a machine mechanically, the first method utilizes a direct mechanical actuation of the position valves to control the actuator. The second method uses a mechanically operated pilot valve to direct the fluid flow to the main position valve. The main position valve controls the actuator.

Fluid Control This is possible by using reliable pilot fluid signals. In fluid power systems, pilot signals indicating the pressure conditions and position conditions can be reliably used to control the motor valves and other components. Sensitive fluid signals can be produced by mechanically actuated position valves and by pilot-size sequence valves. A pressure-sensitive fluid-sequence valve can not only identify the completion of a stroke of a fluid power cylinder, but it also can sense the existing loading conditions of the circuit.

Electrical Control This control of fluid power circuits can be found in a wide variety of forms depending on the individual applications. Linkages, pressure switches, limit switches, timers, and relays can be used to operate solenoids to control the position valves that direct fluid to the motor units. The electric solenoid control system gives the designer a great flexibility in use. The fluid

pressure switches are able to sense the pressure in any part of the system and operate a solenoid valve to divert flow to the tank or to another part of the circuit. Precision limit switches sense the position of the moving members of the machine and are capable of transmitting the electric signal to a solenoid valve to redirect the fluid to other parts of the system. Limit switches also can be used to initiate a sequential timing device that can hold the pressure or position for a set period of time before directing the solenoid to control the flow. Quite often, it is necessary to design a network of switching circuits to coordinate the loads and movements of all actuating units required by the machine. These fluid power circuits can have the capability of counting each operation and storing this information for later use to reset the circuit or start a new operation.

4.4.5 Other Electric Components in Fluid Power Circuits

General types of electric switches are used on electrically controlled fluid power circuits are

- Pressure switches
- Limit switches
- Selector switches
- Push-button switches
- Electric timers

Pressure Switches Used to sense the pressure in various parts of the circuit, they can perform functions similar to those of limit switches. They but do not have the exact positioning feature of the limit switch.

Limit Switches Used in fluid power circuits, these find out the position of moving members which are actuated by fluid motors. Limit switches can give a signal to stop or reverse the operation, increase or decrease the speed of travel, or initiate a new sequence of machine actuation. Limit switches are generally actuated by a roller-arm controlled motion or with a push type cam actuated motion. The switches are designed to return to the initial position by a spring action.

Selector Switches These are classified as single-type switches having two or three positions (with single- and double-throw contacts) or the multiple type. These switches also can be used to program the sequence of machine operation by interconnecting various relays to produce many combinations of fluid power operations.

Push-button Switches Generally, these operate by means of relays. Push-button switches in conjunction with solenoid valves can convert a manually controlled fluid power system into a semi-automatic system. On automatic machines, the push buttons are needed to initiate the operational sequence of the machine in the beginning.

Electric Timers Used in fluid power systems to start or stop various electric components that control the fluid power system, electric timers can coordinate the machine movements and cycle times automatically as long as the sequence of machine operation is established. The main types of timers are the repeat-cycle timer and the reset timer. The repeat-cycle timer is designed to cause the system to continue the sequential motion continuously until the timer is externally stopped, whereas the reset timer is designed to stop the machine operation after one complete cycle. The timer then has to be externally reset to start a new set of sequences.

4.5 Piezoelectric Actuators

Piezomotors move due to piezoelectricity, a property of certain materials to generate an electric charge when placed under compression or tension loads. An electric field placed over a piezocrystal changes the shape of the crystal. This ability to change shape is the basis for piezomotor technology. The motor shaft moves only nanometers for each step, but the motion can repeat thousands of times/second. At that rate, the armature can actually move at linear speeds up to 100 mm/sec.

Different models include designs for vacuum and nonmagnetic applications. Various sizes can handle pulling forces from one to several hundred Newtons. Moreover, the simple design supports mass production while still maintaining a high degree of precision.

Piezomotors are viable alternatives to standard DC motors, and in some cases, they may work better. Motion control in piezomotors can reach nanometer precision—a far greater resolution than available with DC motors. DC motors become more expensive as they get smaller, while piezomotors remain at a low cost in their size range. The direct linear drive offered by piezomotors removes the need for linear conversion of a DC motor's rotary motion.

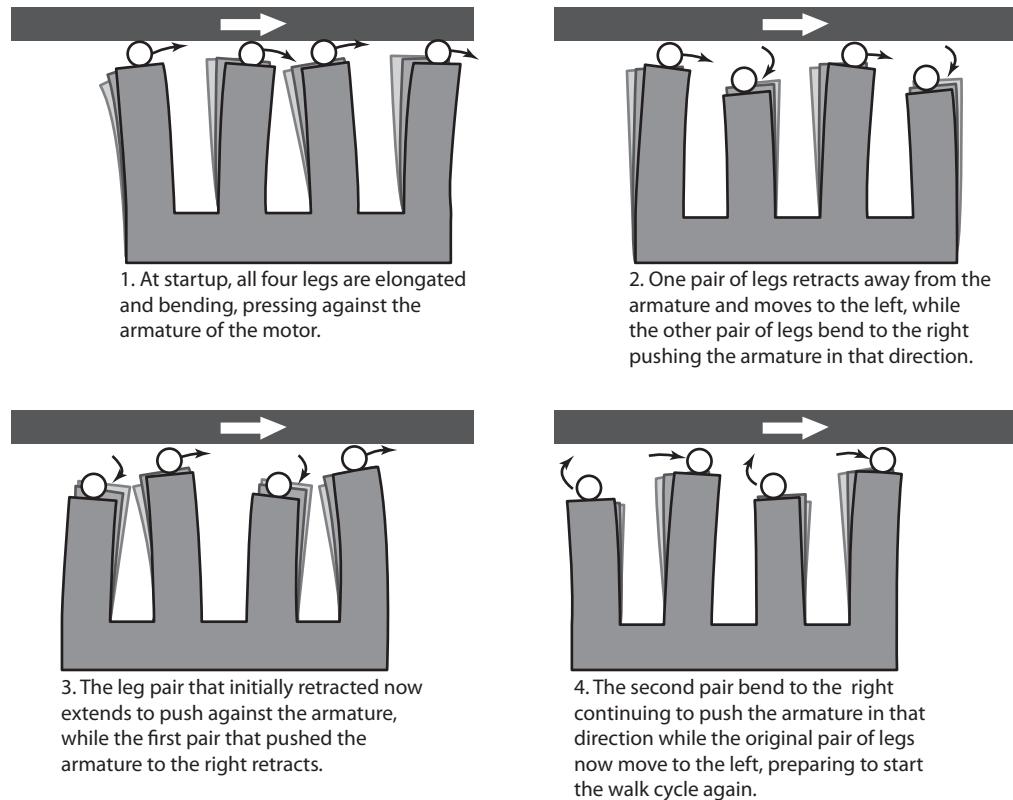
Piezoelectric motors can reduce product size. They also can be more precise, easier to control and adjust, lighter, and more reliable. For example, the PiezoWave™ motor was originally developed for mobile phones. It's now integrated into many applications, including other hand-held devices, medical technology equipment, electromechanical door locks, advanced toys, and cameras.

An ant-sized block of piezoceramic material produces linear motion in the Piezo LEGS® motors (Figure 4-26). Piezo LEGS is essentially a walking machine. It moves incrementally by synchronizing movements between each pair of its four legs. Though armature motion is limited to nanometers/step, thousands of steps/second can create linear-motion speeds up to 100 mm/sec. The PiezoWave motor has two piezoelements on opposite sides of the drive rail that vibrate at ultrasonic frequencies. Drive pads attached to the undulating elements push against either side of the drive rod to create linear movement.

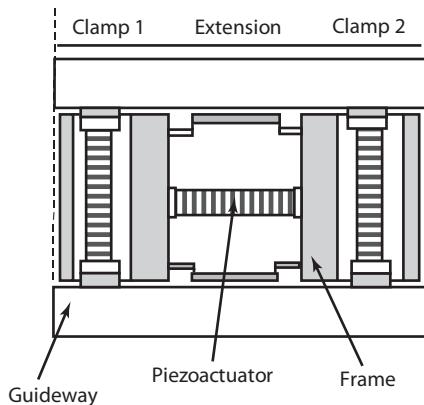
The concept of piezoelectricity, mechanics, and controls has been used for the development of piezoelectric actuators. The piezomotors, which use piezoelectric instead of electromagnetic operating principles, are able to provide high torque at low speeds and allow very precise positioning. Positioning techniques using linear piezomotors have been applied to achieve nanometer resolution over a long travel range for applications such as scanning tunneling microscopy. The positioning stages driven by a ball screw, a lead screw, or a friction drive have been used widely in industry to obtain submicron resolution. However, the problems due to Coulomb friction, stick-slip, elastic deformation, and backlash cause a reduction in resolution and accuracy. In addition, the feed drives used in manufacturing applications are required to have high positioning accuracy, stiffness, and output force over a long range of travel. Piezoelectric actuators are used to overcome these problems. As an example, a linear piezomotor can provide a positioning resolution of 5 nanometers, a stiffness of $90 \text{ N}/\mu\text{m}$, and an output force of 200 N.

The piezoelectric effect has been illustrated earlier in Section 3.2.4. Many different approaches have been used to convert the linear displacement of the piezoelectric material into rotational movement. Figure 4-27 shows the configuration of the linear piezomotor consisting of three piezoelectric actuators and a flexure frame.

The actuators are preloaded directly by the frame. The two side actuators are used to clamp on a guideway, and the central one is used to translate along the guideway. The piezomotor simulates the motions of an inchworm. During the motion, it is required that one of the side actuators should always clamp on the guideway. The linear piezomotors can be modeled as a multiple degree-of-freedom vibration system. The dynamic equation of the system is presented in matrix form as,

FIGURE 4-26 WALKING PRINCIPLE USING PIEZO LEGS

"Tiny motors make big moves," Machine Design, August 2008.

FIGURE 4-27 CONFIGURATION OF LINEAR PIEZOMOTOR

Courtesy of Z. Zhu.

$$M\ddot{x} + C\dot{x} + Kx = F \quad (4-29)$$

Here M , C , and K are 6×6 matrices representing masses, damping coefficients, and stiffness of the system, respectively. x is the displacement vector, and F is the force vector. The central actuator is considered as a mass–spring damper unit with a force input.

4.6 Summary

While selecting a drive method for a mechatronic application, positioning accuracy, speed, cost, and size are some of the considerations. Electric motors are capable of high positioning accuracy if used with a proper control system. The DC motors have the ability to generate the linear torque-to-power ratio and are capable of quick response due to low inductance in the armature. Stepper motors are used for light loads and for open-loop operation. Stepper motors accelerate and decelerate at each step. Fluid power systems generate greater power in a compact volume than do motors driven electrically. Fluid under pressure can be used to operate fluid motors at high torque. The power needed to control a fluid-power servo system is comparatively small. Piezo actuators, because of their ability to provide high torque and allow precise positioning, are useful for special-purpose mechatronic applications.

REFERENCES

- Fitzgerald, Charles Kingsley, Jr. and Stephen D. Umans, *Electric Machinery*. New York: McGraw-Hill, 1983, pp. 508–512.
- Clarence W. deSilva, *Control Sensors and Actuators*. New Jersey: Prentice-Hall, 1989, pp. 253–323.
- Acarnley, Paul P. *Stepping Motors: A Guide to Modern Theory and Practice*. New York: Peter Peregrinus Ltd., 1982, pp. 1–71.
- E. Snyder *Industrial Robots Computer Interfacing and Control*. New Jersey: Prentice Hall, 1985, pp. 67–85.
- Russ Henke. *Fluid Power Systems and Circuits*. Penton/IPC, 1983.
- Zhenqi Zhu and Bhi Zhang. “A microdynamic model for linear piezomotors.” *Proceedings International Manufacturing Engineering Conference*, Storrs, CT, 1996.
- Repas, Robert. “Tiny Motors Make By Moves.” August 21, 2008. <http://machinedesign.com/article/tiny-motors-make-big-moves-0821>

PROBLEMS

- 4.1. A machine table driven by a closed-loop positioning system consists of a servo motor, lead screw, and optical encoder. The lead screw has a pitch of 0.500 cm and is coupled to the motor shaft with a gear ratio of 4:1(4 turns of motor for 1 turn of lead screw). The optical encoder generates 150 pulses/rev of the lead screw. The table has been programmed to move a distance of 15 cm at a feed rate of 45 cm/min. Determine
- How many pulses are received by the control system to verify that table has moved exactly 15 cm?
 - What is the pulse rate?
 - What is the motor speed that corresponds to the specified feed rate?

(Note: The pitch is the axial distance traveled for one revolution of the screw.)

- 4.2. A CNC machine tool table is powered by a servo motor, lead screw, and optical encoder. The lead screw has a pitch of 5 mm and is connected to the motor shaft with a gear ratio of 16:1 (16 turns of the motor for one turn of the lead screw). The optical encoder is connected directly to the lead screw and generates 200 pulses per revolution of the lead screw. The table must move a distance of 100 mm at a feed rate of = 500 mm/min. Determine (a) pulse count received by the control system to verify that the table has moved exactly 100 mm, (b) pulse rate, and (c) motor speed that corresponds to the feed rate of 500 mm/min.
- If the range of the work table axis is 500 mm and there are 12 bits in the binary register used by the digital controller to store the position, determine the control resolution.
- 4.3. A 1.8° stepper motor is directly connected to a machine table driven by a lead screw with three threads per cm. (*Note:* The pitch is the axial distance traveled for one revolution of the screw.)
- Determine the axial distance traveled by the lead screw when an external input of 4355 pulses are sent to the motor.
 - A separate encoder is connected to the other end of the lead screw. The encoder generates 180 pulses/rev. What will be the number of pulses in the part(a)?
- 4.4. A computer-numerically-controlled PCB drilling machine uses a stepper motor for positioning purposes. The lead screw which drives the table of the machine tool has a pitch of 10 mm. The work table traverses a distance of 40 mm at a linear speed of 400 mm per minute. If the stepper motor has 180 step angles, calculate the speed of the stepper and the number of pulses needed to move the machine table to a desired location.
- 4.5. An arm of the cylindrical robot, which is driven by a direct-current motor, needs a torque of 12 N-m. The DC motor has a torque constant of 0.34 N-m per ampere. How much current is needed to drive the robot arm at maximum load?
- 4.6. A solar tracking system uses a stepper motor as an actuator. The stepper faces a constant load torque of 0.7 N-m. The step angle is 1.8° . The inertia of the solar collector is 0.14 N-m/s^2 . If the load needs to be accelerated to 150 steps per second in 1 s, find the minimum motor torque required to conduct this operation.
- 4.7. Prepare a table to compare and contrast the following actuators.
- DC motors
 - stepper motors
 - fluid power actuators
 - pneumatics

Include information on power, linearity, backlash, etc.

CHAPTER 5

SYSTEM CONTROL—LOGIC METHODS

- 
- 5.1 Number Systems in Mechatronics
 - 5.2 Binary Logic
 - 5.2.1 Proofs and Simplification of Several Logic Expressions
 - 5.2.2 Truth Tables
 - 5.3 Karnaugh Map Minimization
 - 5.3.1 Two-Variable Karnaugh Maps
 - 5.3.2 Three-Variable Karnaugh Maps
 - 5.3.3 Four-Variable Karnaugh Maps
 - 5.4 Programmable Logic Controllers
 - 5.5 Summary
 - References
 - Problems

Mechatronics integrates specialized areas including signal conditioning, hardware interfacing, control systems, and microprocessors. This chapter introduces the fundamental technologies responsible for the these areas: digital electronics, analog electronics, and programmable logic controllers. The digital electronic section discusses Boolean algebra and techniques for the optimization of digital circuits. Amplifier selection and analog-to-digital conversion techniques are the focus of the analog electronics section. The chapter ends with a discussion of programmable logic controllers.

5.1 Number Systems in Mechatronics

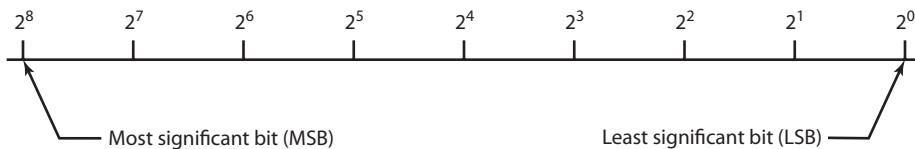
The interfacing of mechatronic systems relies heavily on digital electronics. The information flow in any mechatronic system must pass through digital electronic interface devices while moving from the real world to the computer. Once in the computer, control is often exercised using digital logic.

The concept of switching devices leads to the idea of two signal levels, ON-OFF or HIGH-LOW. Engineers use these to convey information about the operation of systems. From these signals, it is possible to make logical decisions about operating sequences. The information about logic states can be used to make decisions about the progress of a component in a production system. ON-OFF or HIGH-LOW situations are easier to classify than are quantitative situations. Table 5-1 presents the three basic numbering systems: binary, decimal, and hexadecimal.

The binary numbering system forms the basis of all digital computer operation. The electronic circuits in a digital system provide input and output signals that have only two distinct voltage levels. The two levels are referred to as 0 and 1. In addition, the logic circuits can be designed with

TABLE 5-1 THREE BASIC NUMBERING SYSTEMS

System	Base	Symbols
Binary	2	0,1
Decimal	10	0–9
Hexadecimal	16	0–9, A–F

FIGURE 5-1 BINARY CODE NUMBER SYSTEM WEIGHTING

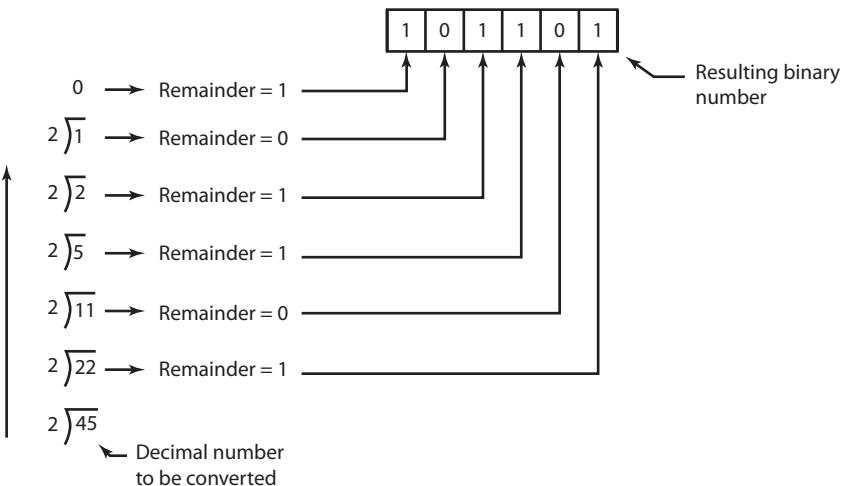
high reliability and are less sensitive to noise, temperature, and aging problems. For a binary code system, the weighting of each bit is presented in Figure 5-1. The most significant bit (MSB) is on the left and the least significant (LSB) one is on the right.

Table 5-2 shows the binary and hexadecimal numbers in the decimal range of integers between 0 and 20.

Decimal numbers are converted to binary form by using long division. The binary equivalent is formed from LSB to MSB as the remainder of successive divisions of the decimal number by the modulus 2. For example, the binary equivalent of 45_{10} is computed as shown in Figure 5-2.

TABLE 5-2 BINARY AND HEXADECIMAL EQUIVALENTS OF DECIMAL INTEGERS FROM 0 TO 20

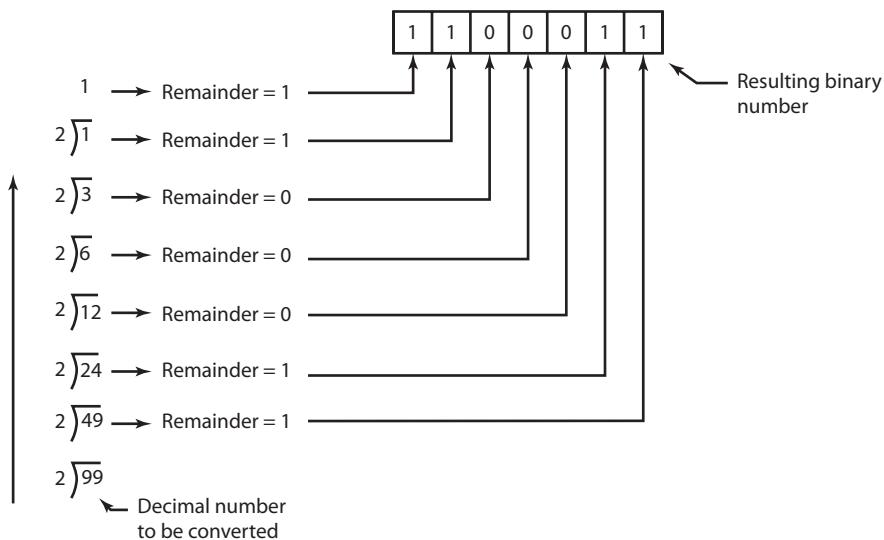
Decimal	Binary	Hex	Decimal	Binary	Hex
0	00000	0	11	01011	B
1	00001	1	12	01100	C
2	00010	2	13	01101	D
3	00011	3	14	01110	E
4	00100	4	15	01111	F
5	00101	5	16	10000	10
6	00110	6	17	10001	11
7	00111	7	18	10010	12
8	01000	8	19	10011	13
9	01001	9	20	10100	14
10	01010	A			

FIGURE 5-2 CONVERSION OF DECIMAL TO BINARY FORM**EXAMPLE 5.1 Computing the Decimal Equivalent of 45_{10} Using Long Division**

Conversion from base 2 back to its decimal equivalent is carried by an inverse operation. The modulus 2 is raised to a value equal to the placement of the bit in the binary number (0 for the LSB to n for the MSB), multiplied by the value of the bit (either 0 or 1), and accumulated to form the single decimal equivalent. Several solutions are presented to illustrate different techniques.

Solution

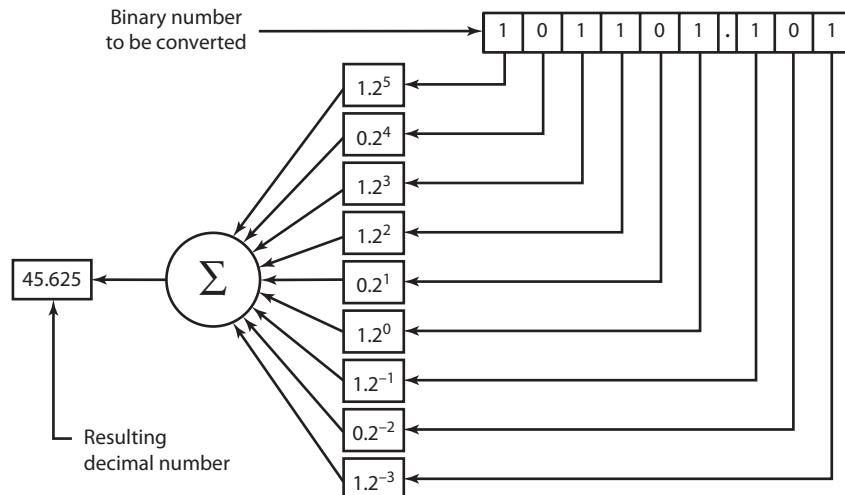
- (a) Conversion of 99_{10} to its binary equivalent is shown in Figure 5-3.

FIGURE 5-3

By computing the binary equivalent of 99_{10} using long division, the binary equivalent is formed from LSB to MSB using the remainder terms from the division.

- (b) Conversion of 101101.101_2 to its decimal equivalent is shown in Figure 5-4.

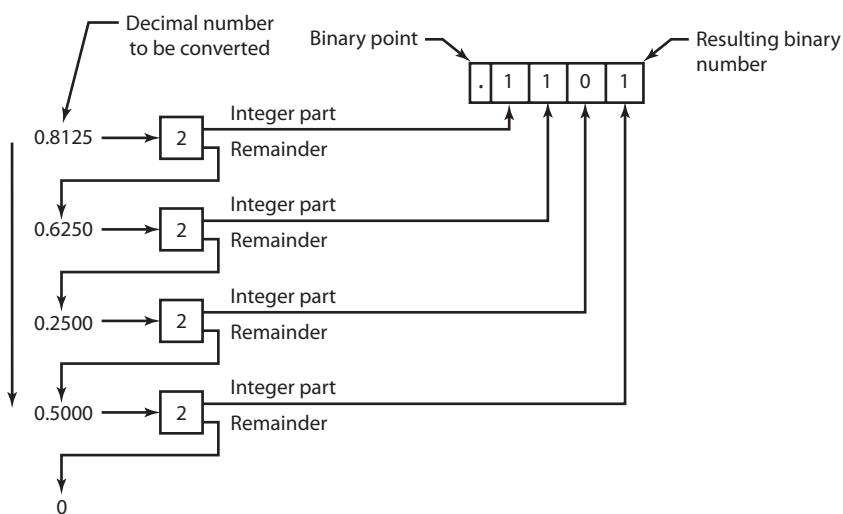
FIGURE 5-4



When a binary point is present, the bit to the left of the binary point is bit 0 and the bit to the right is bit -1 .

- (c) Conversion of 0.8125_{10} to its binary equivalent is shown in Figure 5-5.

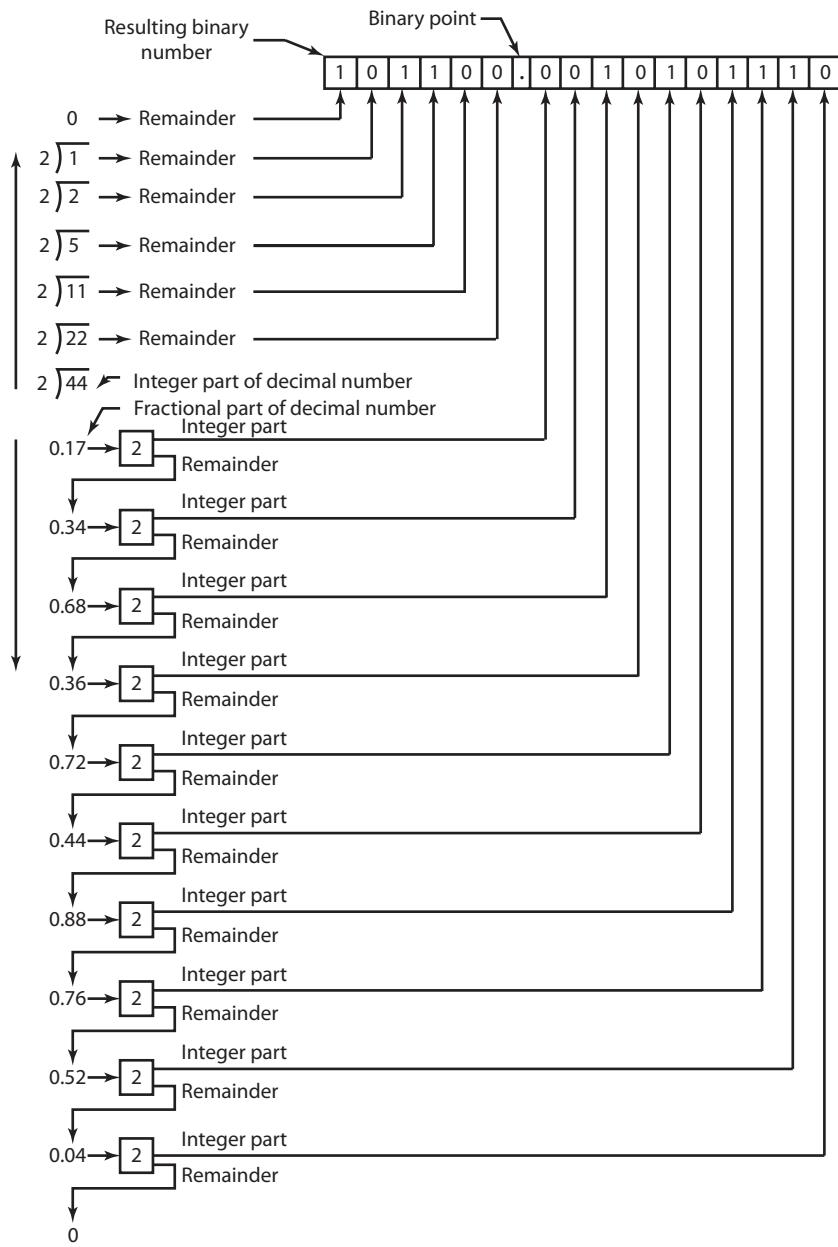
FIGURE 5-5



Computing the binary equivalent of the fractional part of a decimal number utilizes the inverse of long division, and successive multiplication of the fraction by the modulus (2) until the remainder term becomes 0. Binary bits are filled from bit -1 (just right of the binary point) downward.

- (d) Conversion of 44.17_{10} to its binary equivalent is shown in Figure 5-6.

FIGURE 5-6



Computing the binary equivalent of 44.17_{10} combines the integer part conversion (b) and the fractional part conversion (c). It is easy to see how quantization due to finite bit counts (wordlengths) affects precision of the resulting binary number.

In a typical binary system, several bits can change when you move from one state to another. When several bits change as a result of transitioning between two adjacent numbers, hardware problems associated with quantization may occur. For example, in a four-bit binary code, when a transition is made from 2_{10} to 3_{10} only one bit changes ($2_{10} = 0010_2$ and $3_{10} = 0011_2$), however, a change from 7_{10} to 8_{10} results in changes to all four bits, ($7_{10} = 0111_2$ and $8_{10} = 1111_2$).

Gray code is a reflective binary code. Only one bit is changed in Gray code when a change is made from one value to the next increment. In Gray code, $7_{10} = 0100_{\text{gray}}$ and $8_{10} = 1100_{\text{gray}}$, so transitioning between 7_{10} to 8_{10} results in only one bit changing. An error of only one bit in a large binary number can cause large errors in the decimal reconversion, depending on its location in the binary word. Gray codes reduce these type of errors, especially in the case of transducers, where an increment in the measured variable produces a change in only one digit. The Gray code representation of decimal numbers from 0 to 10 is presented in Table 5-3.

TABLE 5-3 GRAY AND BINARY EQUIVALENTS OF DECIMAL INTEGERS FROM 0 TO 10

Decimal	Binary	Gray
0	0000	0000
1	0001	0001
2	0010	0011
3	0011	0010
4	0100	0110
5	0101	0111
6	0110	0101
7	0111	0100
8	1000	1100
9	1001	1101
10	1010	1111

The hexadecimal system is used to represent binary numbers in a “shorthand” form. The conversion from binary to hexidecimal is accomplished by converting the binary information in groups of four bits using the following example. The information within any digital system must be represented by a binary code, since the circuitry allows only two voltage levels. The hexadecimal representation of binary numbers is illustrated in Example 5.2.

EXAMPLE 5.2 Binary Representation of the 9C.A Hex Number

$$9C.A_{16} = 10011100.1010_2$$

where

$$9 = 1001$$

$$C = 1100$$

$$A = 1010$$

Solution

Hexadecimal representation of the binary number 1111100110.011 as

$$1111100110.011_2 = 3E6.6_{16}$$

where

$$3 = 0011$$

$$E = 1110$$

$$6 = 0110$$

5.2 Binary Logic

The logic circuits can be described by the Boolean algebraic system in which the variables are limited to two values, usually denoted as 0 and 1. George Boole developed an algebra for the systematic treatment of logic. Boolean algebra deals with variables that take on two discrete values, 0 and 1, and with operations that assume logical meaning. Situations involving “yes-no”, “true-false”, “on-off”, etc. can be represented by Boolean logical expressions. The basic Boolean algebra laws are presented in Table 5-4.

TABLE 5-4 BASIC BOOLEAN ALGEBRAIC LAWS WHERE A, B, AND C ARE VARIABLES

1. $A + 1 = 1$	9. $A + B = B + A$
2. $A + 0 = A$	10. $AB + AC = A(B + C)$
3. $A \cdot 0 = 0$	11. $A + BC = (A + B)(A + C)$
4. $A \cdot 1 = A$	12. $\bar{A} + \bar{B} = \bar{A} \cdot \bar{B}$
5. $A + A = A$	13. $\bar{A} \cdot \bar{B} = \bar{A} + \bar{B}$
6. $A \cdot A = A$	14. $A \oplus B = A \cdot \bar{B} + \bar{A} \cdot B$
7. $A \cdot \bar{A} = 0$	15. $A + \bar{A}B = A + B$
8. $A + \bar{A} = 1$	

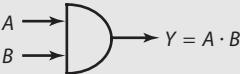
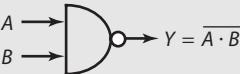
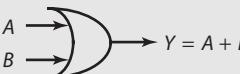
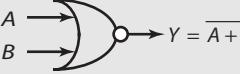
The laws presented in Table 5-4 are based on six axioms dealing with properties of Boolean algebra. The axioms; commutative, distributive, indempotency, absorption, complementation, and Demorgan’s laws are described in Table 5-5.

TABLE 5-5 FUNDAMENTAL BOOLEAN AXIOMS

Commutative Axiom:	Distributive Axiom:	Indempotency Axiom:
$A \cdot B = B \cdot A$	$A \cdot (B + C) = (A \cdot B) + (A \cdot C)$	$A \cdot A = A$
$A + B = B + A$	$A + (B \cdot C) = (A + B) \cdot (A + C)$	$A + A = A$
Absorption Axiom:	Complementation Axiom:	Demorgan’s Law:
$A \cdot (A + B) = A$	$A \cdot \bar{A} = 0$	$\bar{A} \cdot B = \bar{A} + \bar{B}$
$A + (A \cdot B) = A$	$A + \bar{A} = 1$	$\bar{A} + B = \bar{A} \cdot \bar{B}$

A summary of basic logic elements is presented in Table 5-6. These elements form the foundations of digital logic.

TABLE 5-6

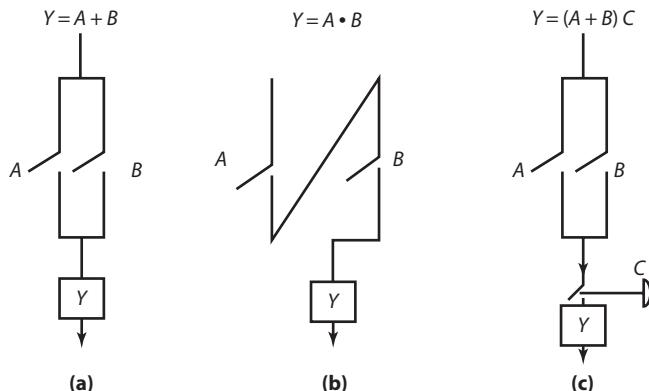
	Description	Truth Table	Logic Gate
AND Logic Element	The "AND" element has two or more inputs and one output. The output is true (1) when all the inputs are true. If one or more of the inputs are false the output will be false.	A B Y 0 0 0 0 1 0 1 0 0 1 1 1	 $Y = A \cdot B$
NAND Logic Element	The "NAND" element is identical to the "AND" element except it's output is negated.	A B Y 0 0 1 0 1 1 1 0 1 1 1 1	 $Y = \overline{A \cdot B}$
OR Logic Element	The "OR" element has two or more inputs and one output. The output is true if any of the inputs are true and false only when all inputs are false.	A B Y 0 0 0 0 1 1 1 0 1 1 1 1	 $Y = A + B$
NOR Logic Element	The "NOR" element is identical to the "OR" element except it's output is negated.	A B Y 0 0 1 0 1 0 1 0 0 1 1 0	 $Y = \overline{A + B}$
XOR Logic Element	Similar to the "OR" except the output is false when all inputs are true or false.	A B Y 0 0 0 0 1 1 1 0 1 1 1 0	 $Y = A \cdot \overline{B} + \overline{A} \cdot B$ <i>or</i> $Y = A \oplus B$

EXAMPLE 5.3

- (a) A machine can be operated by either of the two operators, A and B. The power that runs a machine can be connected from either of two locations.
- (b) Due to the safety requirements, the power must be channeled through both stations to operate the machine.
- (c) The final safety regulations allow either station to power the machine only if the operator is out of danger

Solution

The logic elements are given in Figure 5-7.

FIGURE 5-7 BASIC LOGIC ELEMENTS

5.2.1 Proofs and Simplification of Several Logic Expressions

Proof: $A \cdot B + A \cdot \bar{B} = A$

$$\begin{aligned}&= A \cdot (B + \bar{B}) \\&= A \text{ because } (B + \bar{B}) = 1\end{aligned}$$

Proof: $(A + B) \cdot (A + A \cdot \bar{B}) = A$

$$\begin{aligned}&= A \cdot A + A \cdot \bar{B} + A \cdot B + B \cdot \bar{B} \cdot A \\&= A + A \cdot \bar{B} + A \cdot B \cdot (B + \bar{B}) \\&= A + A \cdot \bar{B} + A \cdot B = A + A \cdot (B + \bar{B}) \\&= A\end{aligned}$$

Simplification Example: $\bar{A} \cdot B + A \cdot B + \bar{A} \cdot \bar{B} = B \cdot (A + \bar{A}) + \bar{A} \cdot \bar{B}$

$$\begin{aligned}&= B + \bar{A} \cdot \bar{B} = B + \bar{A} \cdot \bar{B} = B + \bar{B} \cdot \bar{A} \\&= B + \bar{A}\end{aligned}$$

Simplification Example: $W = X \cdot Y + X \cdot (Z + Y) + X \cdot Z$

$$\begin{aligned}&= X \cdot Y + X \cdot Z + X \cdot Z + X \cdot Y \\&= X \cdot Y + X \cdot Z + X \cdot (Y + Z) \\&= X \cdot (Y + Z) + X \cdot (Z + Y) = X \cdot (Y + Z)\end{aligned}$$

Simplification Example: $D = A \cdot B \cdot \bar{C} + \bar{A} \cdot B \cdot C + A \cdot B \cdot C + A \cdot B \cdot \bar{C} + A \cdot C$

$$\begin{aligned}&= B \cdot C \cdot (\bar{A} + A) + B \cdot \bar{C} \cdot (\bar{A} + A) + A \cdot C \\&= B \cdot C + B \cdot \bar{C} + A \cdot C = B \cdot (C + \bar{C}) + A \cdot C \\&= B + A \cdot C\end{aligned}$$

Simplification Example: $F = \bar{A} + A \cdot \bar{B} + \bar{A} \cdot B$

$$\begin{aligned}&= A \cdot (1 + \bar{B}) + \bar{A} \cdot B \\&= A + \bar{A} \cdot B = A + B\end{aligned}$$

Simplification Example: $F = \bar{A} \cdot B \cdot C + A \cdot \bar{B} \cdot C + A \cdot B \cdot C + A \cdot B \cdot \bar{C}$

$$\begin{aligned}&= \bar{A} \cdot B \cdot C + A \cdot (\bar{B} \cdot C + B \cdot C + B \cdot \bar{C}) \\&= \bar{A} \cdot B \cdot C + A \cdot (B \cdot C + B) = \bar{A} \cdot B \cdot C + A \cdot (C + B) = \\&\quad \bar{A} \cdot B \cdot C + A \cdot C + A \cdot B \\&= A \cdot B + C \cdot (A + \bar{A} \cdot B) = A \cdot B + C \cdot (A + B) \\&= A \cdot B + B \cdot C + C \cdot A\end{aligned}$$

Simplification Example,

Negate the expression: $F = \bar{X} \cdot \bar{Z} + \bar{Y} \cdot \bar{Z}$

$$\begin{aligned}\bar{F} &= \overline{\bar{X} \cdot \bar{Z} + \bar{Y} \cdot \bar{Z}} \\&= \overline{\bar{X} \cdot \bar{Z}} \cdot \overline{\bar{Y} \cdot \bar{Z}} \text{ (using DeMorgan's Theorem)} \\&= (X + Z) \cdot (Y + Z) = X \cdot Y + Y \cdot Z + X \cdot Z + Z \cdot Y \\&= X \cdot Y + Z \cdot (1 + X + Y) \\&= Z + X \cdot Y\end{aligned}$$

5.2.2 Truth Tables

A logical function $f(A_1, A_2, \dots)$ may be represented by a truth table. The truth table lists the dependent function evaluation for every possible combination of the independent variables. Table 5-7 presents an example of the truth table produced for DeMorgan's theorem. It can be seen from the truth table that column 4 and column 7 have similar logical states, which verifies the relationship $\bar{A} \cdot B = \bar{A} + \bar{B}$ and $\bar{A} + \bar{B} = \bar{A} \cdot \bar{B}$.

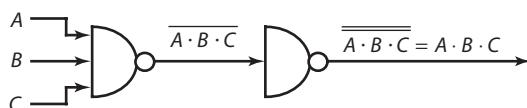
TABLE 5-7 TRUTH TABLE FOR DEMORGAN'S THEOREM: $A \cdot B \cdot \dots \cdot N = \bar{A} + \bar{B} + \dots + \bar{N}$ AND $A + B \cdot A + B + \dots + N = \bar{A} \cdot \bar{B} \cdot \dots \cdot \bar{N}$

A	B	$A \cdot B$	$\bar{A} \cdot \bar{B}$	\bar{A}	\bar{B}	$\bar{A} + \bar{B}$	$\bar{A} \cdot \bar{B} = \bar{A} + \bar{B}$
0	0	0	1	1	1	1	
0	1	0	1	1	0	1	
1	0	0	1	0	1	1	
1	1	1	0	0	0	0	
A	B	$A + B$	$\bar{A} + \bar{B}$	\bar{A}	\bar{B}	$\bar{A} \cdot \bar{B}$	$\bar{A} + \bar{B} = \bar{A} \cdot \bar{B}$
0	0	0	1	1	1	1	
0	1	1	0	1	0	0	
1	0	1	0	0	1	0	
1	1	1	0	0	0	0	

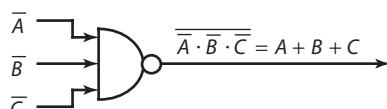
Logic diagrams provide another useful means of presenting the behavior of a logical function. Figure 5-8 illustrates how identical operations can be performed with different combinations of the logic elements.

Figure 5-9 illustrates the use of the logic elements.

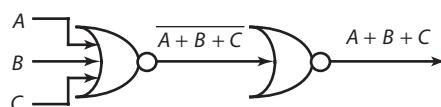
FIGURE 5-8 LOGIC DIAGRAMS



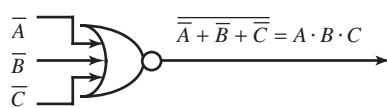
(a) "AND" operation using "NAND" elements



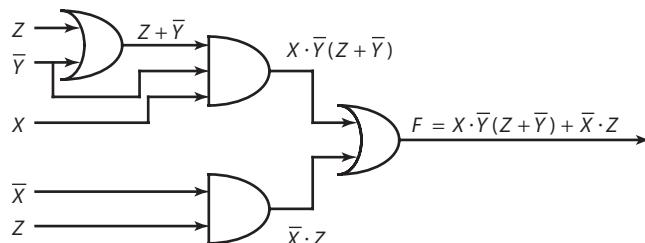
(b) "OR" operation using "NAND" elements



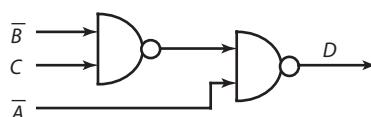
(c) "OR" operation using "NOR" elements



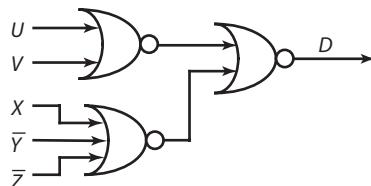
(d) "AND" operation using "NOR" elements

FIGURE 5-9 USES OF LOGIC ELEMENTS(a) Implementation of $F = X \cdot \bar{Y} (Z + \bar{Y}) + \bar{X} \cdot Z$

$$\begin{aligned} D &= A + \bar{B} \cdot C \\ &= \overline{\overline{A} + \bar{B} \cdot C} \\ &= \overline{\overline{A} \cdot \overline{\bar{B} \cdot C}} \end{aligned}$$

(b) Implementation of $D = A + \bar{B} \cdot C$ using "NAND" functions

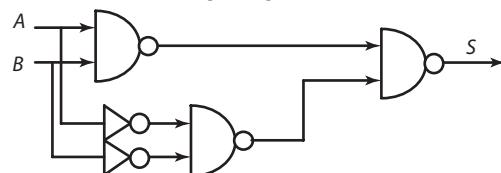
$$\begin{aligned} D &= (U + V) \cdot (X + \bar{Y} + \bar{Z}) \\ &= \overline{(U + V) \cdot (X + \bar{Y} + \bar{Z})} \\ &= \overline{\overline{(U + V)} + \overline{(X + \bar{Y} + \bar{Z})}} \end{aligned}$$

(c) Implementation of $D = (U + V) \cdot (X + \bar{Y} + \bar{Z})$ using "NOR" functions

Truth table

A	B	S
0	0	1
0	1	0
1	0	0
1	1	1

Logic diagram

(d) Construction of logic diagram from the truth table.
Note the use of the NOT operators to negate A and B

Truth table

A	B	C	F
0	0	1	0
0	0	0	0
0	1	1	1
0	1	0	0
1	0	1	1
1	0	0	0
1	1	1	1
1	1	0	1

Logic diagram

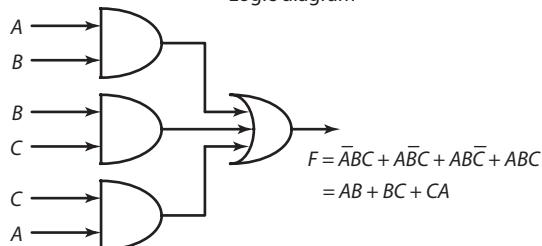
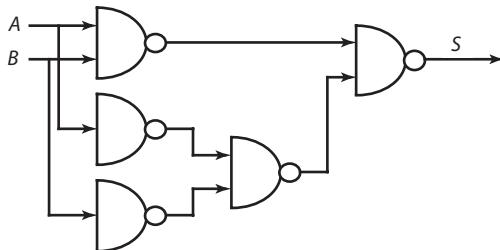
(e) Construction of three input logic diagram from truth table information.
An application of this circuit is presented later in this chapter**(Continued)**

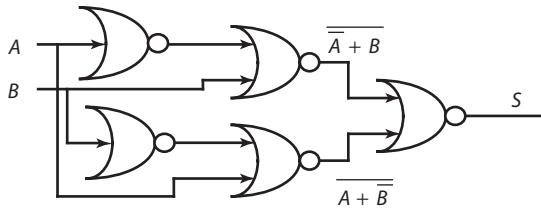
FIGURE 5.9 (CONTINUED)

$$\begin{aligned} S &= \overline{\overline{A} \cdot \overline{B} + AB} \\ &= \overline{\overline{A} \cdot \overline{B}} \cdot \overline{AB} \end{aligned}$$



(f) Design the logic circuit; $S = \overline{\overline{A} \cdot \overline{B}} + AB$ using "NAND" elements.

$$\begin{aligned} S &= \overline{\overline{A} \cdot \overline{B}} + AB \\ &= \overline{\overline{A} \cdot \overline{B} + A \cdot \overline{A} + A \cdot B + B \cdot \overline{B}} \\ &= \overline{\overline{A} \cdot (\overline{B} + A)} + B \cdot (A + \overline{B}) \\ &= (\overline{A} + B) \cdot (A + \overline{B}) \\ &= \overline{\overline{A} + B} \cdot (A + \overline{B}) \\ &= \overline{\overline{A} + B} + (A + \overline{B}) \end{aligned}$$



(g) Implementation can also be made using "NOR" elements

Truth table

<i>I</i>	<i>A</i>	<i>B</i>	<i>Y</i>
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

Automated test system example with three inputs; *A*, *B*, and *I* (an instruction bit) and one output, *y*. The output is determined through the following logic. If *I* = 0, then *Y* = *A* + *B* else, *Y* = *AB*

5.3 Karnaugh Map Minimization

Generally the expression(s) for the output of a digital system are not available in minimum form. Minimizing these expressions using Boolean theorems is a tedious and inefficient process. An equivalent but simpler graphical approach called the *Karnaugh map method* is usually employed. This method is based on the distributive, complementation, idempotency, and “0” and “1” laws.

A Karnaugh map (K-map) is a visual representation of a logic expression which contains all the information in the truth table for that expression presented as a group of boxes or areas labelled in a particular way. It is an orderly arrangement of squares with assignments so that there is only a one-variable change for any adjacent squares. A Karnaugh map contains 2^n squares where n is the number of inputs influencing the logical function. Every square represents an input combination

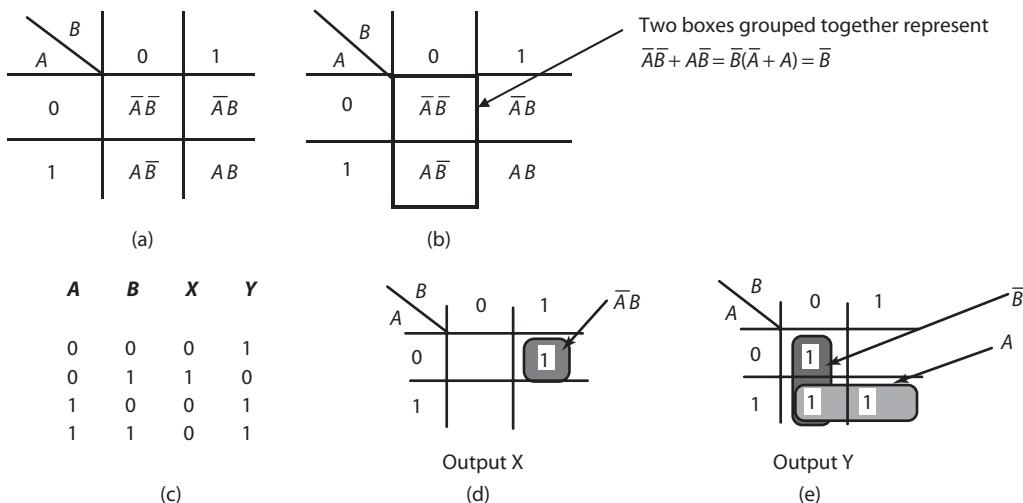
and results in a component of the sum of the product term. A value of “0” or “1” inside the square represents the output of the logical function for that input combination.

5.3.1 Two-Variable Karnaugh Maps

EXAMPLE 5.4

Consider the truth table in Figure 5-10(c) for a two-input, two-output digital system.

FIGURE 5-10 KARNAUGH MAPS



Solution

Output $X = \bar{A} \cdot B$. The output X is already minimized, since no terms can be combined.

$$\text{Output } Y = \bar{A} \cdot \bar{B} + A \cdot \bar{B} + A \cdot B$$

For output Y , minimization using Boolean algebra would result in

$$\text{Output } Y = \bar{A} \cdot \bar{B} + A \cdot \bar{B} + A \cdot B$$

$$(\text{Formula } A + A = A)$$

Using the Karnaugh map in Figure 5-10(e), if the adjacent boxes are combined, the function can be read as $A + B$. The K-map is configured so that there is a difference in only one variable between any two adjacent squares. This setup makes it easy for minimizing Boolean functions without using Boolean algebra manipulations. Therefore, for each grouping of two adjacent 1's (or minterms) in the K-map, a corresponding combination and reduction occurs. To get the minimized boolean sum-of-product (sp) expression from the K-map:

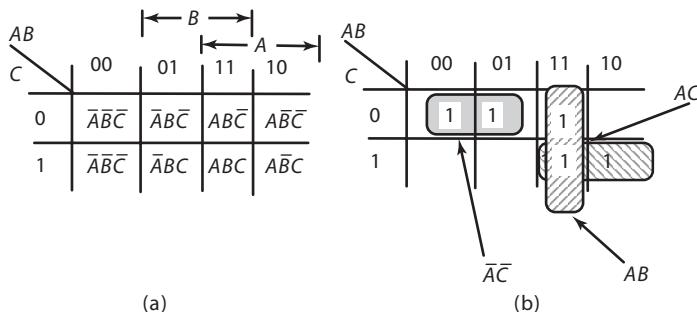
Every “1” in the map must be circled at least once to account for each minterm. Each circled term is a product term in the minimization. To obtain it, first drop the variables that change within the encirclement. The resulting minimized product term is developed by **ANDing** the remaining variables together where the value ‘0’ ‘1’ of each remaining variable indicates complementation (uncomplementation) for that variable. Finally, all the reduced product terms are together to form the minimized sum of products for the Boolean expression.

In this example, the K-map has two product terms. The vertical and horizontal encirclements give the reduced product terms A and B , respectively. The resulting output expression for Y , as shown in Figure 5-10(e), is the **OR** of these terms. This is the same result obtained previously using the cumbersome Boolean algebra theorems directly.

5.3.2 Three-Variable Karnaugh Maps

In a three-variable Karnaugh map, there are 2^3 combinations. Typical examples of combining neighbouring cells is shown in Figure 5-11.

FIGURE 5-11



EXAMPLE 5.5

Consider the states of input variables (A, B, C) shown in the truth table. Output (1) occurs at 010, 011, 110, 111. Simplify the output expression.

A	B	C	F(A, B, C) (Output)
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	0
1	1	0	1
1	1	1	1

Solution

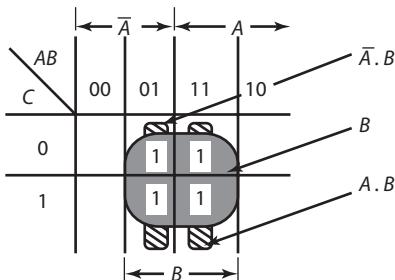
The K-map is shown in Figure 5-12.

Considering the two vertical groupings get reduced to $\bar{A}B$, $\bar{A}B\bar{C}$ and $\bar{A}BC$ get reduced to AB .

$$F = \bar{A} \cdot B + A \cdot B = (\bar{A} + A) \cdot B$$

$$F = 1 \cdot B$$

$$F = B$$

FIGURE 5-12 THREE-VARIABLE KARNAUGH MAP FOR EXAMPLE 5.5

However, by simply considering the grouping of four 1's in the K-map and applying the previously specified rules, the same result is obtained, because grouping A and C change and $B = 1$ leads to the conclusion that $F = B$.

EXAMPLE 5.6

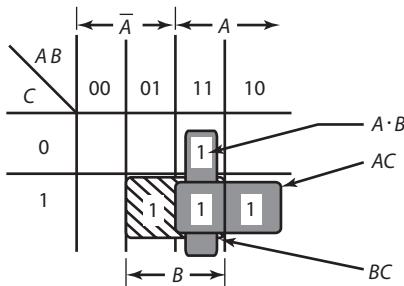
Design a start circuit for a semi-automated punching machine with three variables as control parameters. The variables are the protective guard control signal (A), remote start signal (B), and normal start signal (C). The truth table for implementation is

<i>A</i>	<i>B</i>	<i>C</i>	Start
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

Solution

The K-map is shown in Figure 5-13. The algebraic procedure for simplification is also shown. It is obvious that the K-map method provides the same results in a simpler fashion.

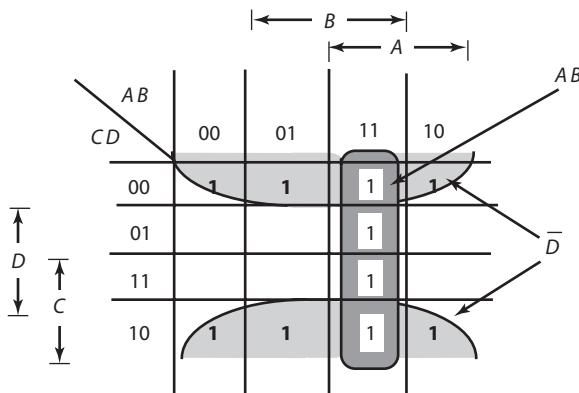
$$\begin{aligned}
 \text{Output} &= \bar{A}\bar{B}C + A\bar{B}C + AB\bar{C} + ABC \\
 &= \bar{A}\bar{B}C + A[\bar{B}C + B(\bar{C} + C)] \\
 &= \bar{A}\bar{B}C + A[\bar{B}C + B] \\
 &= \bar{A}\bar{B}C + A[\bar{B}C + B] \\
 &= \bar{A}\bar{B}C + A[C + B] \\
 &= AB + C(A + \bar{A}B) \\
 &= AB + C(A + B) = AB + BC + CA
 \end{aligned}$$

FIGURE 5-13 KARNAUGH MINIMIZATION FOR EXAMPLE 5.6

5.3.3 Four-Variable Karnaugh Maps

In the four variable K-map, there are 2^4 combinations (Figure 5-14), which shows the minimized boolean expression from the two groupings of eight and four 1's respectively as

$$F = \bar{D} + AB$$

FIGURE 5-14 FOUR-VARIABLE KARNAUGH MAP

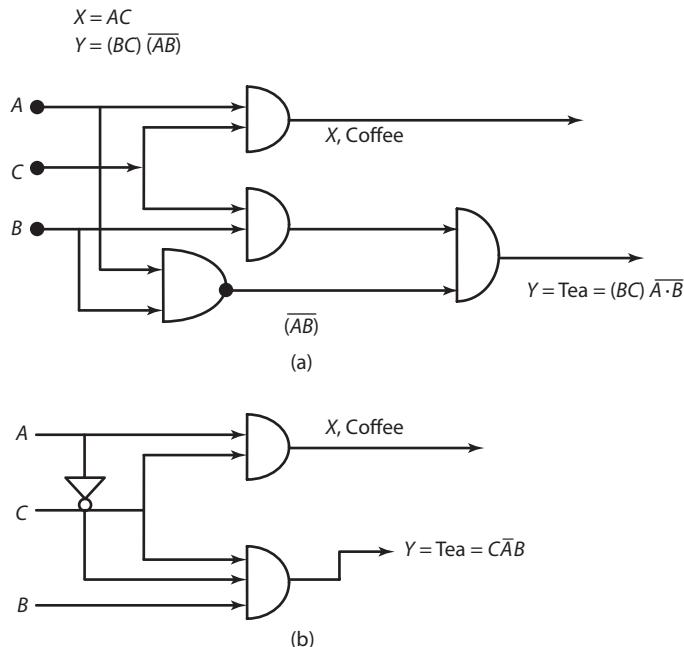
In some logic systems, there are some input combinations that are not defined or indicate inputs for which outputs are not specified. They are known as “Don’t care states.” While examining the K-map, the cells that correspond to don’t care states can be set to either “0” or “1” in such a way that the output equations can be simplified.

EXAMPLE 5.7

Design a combinational logic system for a vending machine that dispenses either coffee or tea when coins are inserted. Let A, B, and C represent coffee, tea, and coin inputs, respectively. The condition for output is such that either coffee or tea will be dispensed when someone inserts the coin and presses the appropriate button. If, on the other hand, you press both the coffee and the tea buttons after inserting the coin, the machine should dispense *coffee* only.

Solution

The logic diagram (Figure 5-15) and truth table (Table 5-8) for the vending machine are shown. Figure 5-15(a) shows using AND/NAND elements. Figure 5-15(b) shows an alternate arrangement.

FIGURE 5-15 LOGIC DIAGRAM OF THE VENDING MACHINE FOR EXAMPLE 5.7**TABLE 5-8 TRUTH TABLE FOR EXAMPLE 5.7**

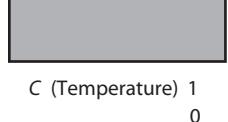
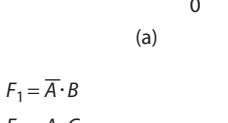
A (Coffee)	B (Tea)	C (Coin)	X (Coffee Output)	Y (Tea Output)
0	0	0	0	0
0	0	1	0	0
0	1	0	0	0
0	1	1	0	1
1	0	0	0	0
1	0	1	1	0
1	1	0	0	0
1	1	1	1	0

EXAMPLE 5.8

Consider a chemical tank for which there are three variables to be monitored. These variables are level, pressure, temperature. The circuit has to be designed such that an alarm is sounded when certain combinations of conditions of the variables occur. The alarm will sound if

1. The liquid level is low and the pressure is high.
2. The liquid level is high and the temperature is high.
3. High liquid level with low temperature and high pressure.

FIGURE 5-16 LOGIC DIAGRAM OF THE TANK FOR EXAMPLE 5.8

	A (Level)	1
		0
	B (Pressure)	1
		0
	C (Temperature)	1
		0

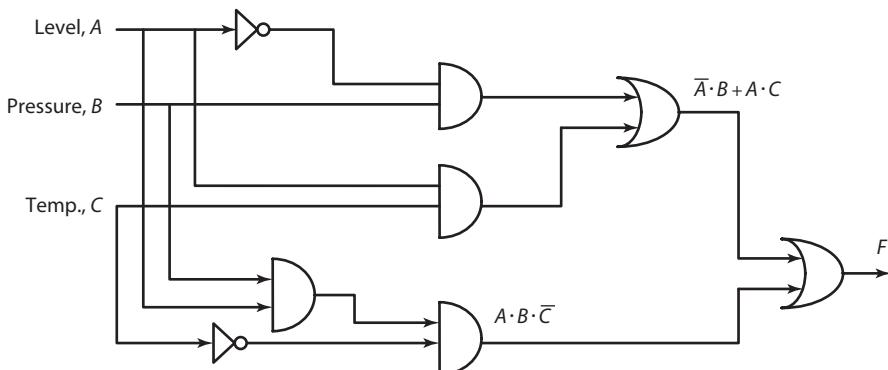
(a)

$$F_1 = \overline{A} \cdot B$$

$$F_2 = A \cdot C$$

$$F_3 = A \cdot \overline{C} \cdot B$$

$$F = \overline{A} \cdot B + A \cdot C + A \cdot B + A \cdot \overline{C} \cdot B$$



(b)

EXAMPLE 5.9

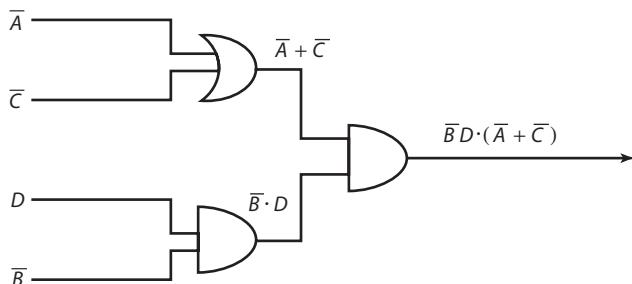
A metal-punching press with logic control should operate when the four combinations defined in Table 5-9 exist, and it should not operate if any other combination exists. Design a logic system for starting. The signal from the sensor operated by the guard is A , the signal from the operator is B , the signal from the workpiece is C . D is the signal from the remote sensor. (*Note: x* represents don't care in the truth table)

Solution

The conditions for the *start* are identified from the Table 5.9. The logic expression is derived by combining various start conditions. Using the Karnaugh map, the logic expression is minimized. Figure 5-17 shows the logic diagram for implementation.

TABLE 5-9 TRUTH TABLE FOR EXAMPLE 5.8

A	B	C	D	Start	A	B	C	D	Start
0	0	0	0	0	1	0	0	0	0
0	0	0	1	1	1	0	0	1	x
0	0	1	0	0	1	0	1	0	0
0	0	1	1	1	1	0	1	1	0
0	1	0	1	0	1	1	0	0	x
0	1	1	0	0	1	1	0	1	0
0	1	1	1	0	1	1	1	0	0
0	1	1	1	0	1	1	1	1	x

FIGURE 5-17 LOGIC DIAGRAM OF THE PUNCHING PRESS FOR EXAMPLE 5.8

5.4 Programmable Logic Controllers

The *programmable logic controller* (PLC) is an extremely durable and reliable modular commercial-off-the-shelf computer system used primarily in the automation industry for controlling machines, assembly lines, processes (including chemical, nuclear, pharmaceutical, paper, beer, wastewater, and others), material handling systems, and even amusement park rides. In today's market, there are many suppliers of PLC systems. Some of the most popular suppliers include Allen Bradley, Schnieder (formerly Modicon), Omron, GE, Mitsubishi, and Siemens.

Most PLC suppliers offer a broad selection of add-on modules to their PLC base module, ranging from input and output modules (capable of interfacing directly with various types of sensors and motors with little or no intermediate hardware necessary), displays, and various types of network connectivity (MODBUS, DeviceNet, Ethernet, RS232, and others). PLCs are generally preferred over custom designed embedded solutions when changes to the control system logic over its lifetime are expected. They generally are applied to systems that are significantly much more expensive than the first cost of the PLC system.

PLCs were introduced in the late 1960s as a software programmable alternative to the current state of the art *hard wired* relay controller. The hard wired relay controller used electrical circuits to implement control logic. Changes to the logic were risky, costly, and extremely labor intensive. In response to a 1968 request from the General Motor Hydramatic Division for an industrial rated programmable factory controller, Bedford Associates developed the first PLC named modular digital controller (MODICON). As part of MODICON, a programming language, similar to the hard wired relay control diagrams, was introduced. This language, called ladder logic, was easily understood by existing engineers and streamlined the transition from hard wired relay controllers to PLCs. Ladder logic remains

a standard to this day, however, in recent years the ability to program PLCs in additional languages, such as C and BASIC, has also become popular and is supported by many PLC suppliers.

PLC systems are normally configured in a chassis. The chassis is a mountable rack with slots for the modules to plug into. Typical chassis sizes range from four slots to as many as 16. Larger systems may require several chassis to achieve the desired number of inputs and outputs. These chassis are connected using interface modules and cabling. An example is shown in the following figure of the Allen Bradley SLC 500 system chassis with seven modules.

The large module to the left is the power module providing power for the modules in the chassis. Moving to the right, the next module is the base module (the PLC CPU) which contains the control program. The remaining modules to the right are a combination of input and output modules. The rightmost module is an interface module (called a scanner module in the Allen Bradley product line) used to interface with other chassis.

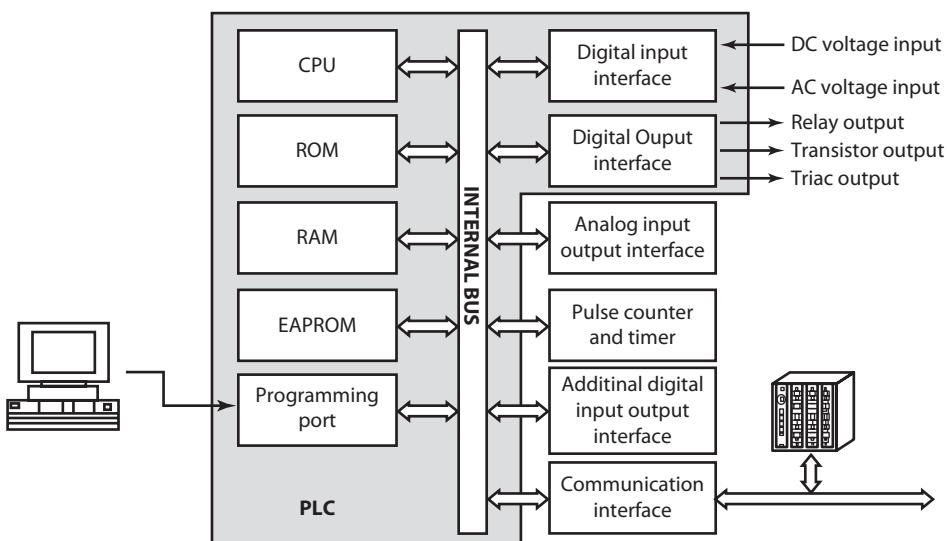
FIGURE 5-18 ALLEN BRADLEY SLC 500 SYSTEM



Courtesy of Rockwell Automation, Inc.

PLC Architecture From a hardware perspective, the PLC consists of a *central processing unit* (CPU), various types of memory, a programming port, I/O modules, and communication interfaces. A typical hardware configuration is presented in the following figure.

FIGURE 5-19 PLC HARDWARE CONFIGURATION



The CPU reads the input data from various sensing devices (digital inputs, analog inputs, timers, and communication interface), executes the stored user program from memory, and sends output commands to control devices (digital outputs, counter timers, communication interface).

The PLC memory consists of ROM, RAM and EEPROM (electronically erasable programmable read only memory, also known as flash memory). The ROM contains the operating system, the RAM contains system data and memory mapped input/output, and the EEPROM contains the control program. The system data section of RAM is used by the operating system to store its data. The memory mapped data section contains a copy of the input values that are used by the control program in the EEPROM and also a copy of the output values calculated by the control program.

The process of reading inputs, executing the control program, and controlling outputs is done sequentially and is called *scanning*. During the first part of the scan, all inputs are read and their values copied as states to an input table located in RAM. During the second part of the scan, the control program (ladder logic), located in the EEPROM, executes using the state values from the input table and, in turn, calculates and writes the output values into an output table, located in RAM. During the third and final part of the scan, the output table values are copied to the physical output channels. The scan time is a function of the I/O count and the complexity of the control program. For very simple systems with fewer than 10 I/O points, scan times of a millisecond or less can be achieved. For larger applications with a thousand points or more, scan times of 20 milliseconds or longer are common.

Connections to input and output devices are made through terminal strips. These devices cover the full range of AC and DC voltages for inputs and up to 10 amps per point for output devices. A PLC does not require a monitor and keyboard to be permanently attached. It can be programmed by several types of peripheral devices including PCs, programming consoles, and hand held programming devices. Once the PLC has been programmed, the programming device can be removed.

The operating system of the PLC operates in one of two modes: the programming mode and the run mode. In the programming mode, the PLC communicates with a programming device, PC, console, or hand held, connected to the programming port enabling a control program to be downloaded into the EEPROM memory. In the run mode, the PLC executes the instructions in the control program. For life-critical applications, most PLC suppliers support redundant operation where two separate identical PLC systems are used. Two modes that are typical are the hot backup and cold backup modes. In the hot backup mode, one PLC system, called the primary controller, runs in the foreground and the second PLC system, called a backup controller, runs in the background. If a failure should occur, the primary controller is automatically taken off line and is replaced by the background controller, all within one scan time of the control algorithm. In the cold backup mode, the operation is performed manually with a switch. When using a hot backup redundant system, the controller scan time can be significantly increased, in some situations by up to a factor of two. The increased time is a function of how the PLC supplier supports redundancy and data sharing. Scan time is always an important consideration when applying a PLC to a system with critical timing requirements.

PLCs are often networked when used in large applications. Although many network configurations are possible, one of the most common uses an ethernet backbone to interface the PLCs with a database server and a human machine interface (HMI) server. In addition, local device networks (such as ControlNet, shown in the figure below) may be included to reduce the level of communication traffic on the Ethernet network. An example of this network configuration is shown in the following figure.

This type of network configuration is common to many applications, in particular, supervisory control and data acquisition (SCADA) systems. SCADA systems are used in most industrial process industries including steel, power, chemical, pulp/paper, wastewater, and pharmaceutical, as well as material handling application. It is not abnormal to have systems with dozens of PLC racks and tens of thousands of I/O points in a SCADA application. Baggage handling systems in airports use the SCADA architecture exclusively. A single airline terminal alone may require 10 PLCs, 2000 I/O

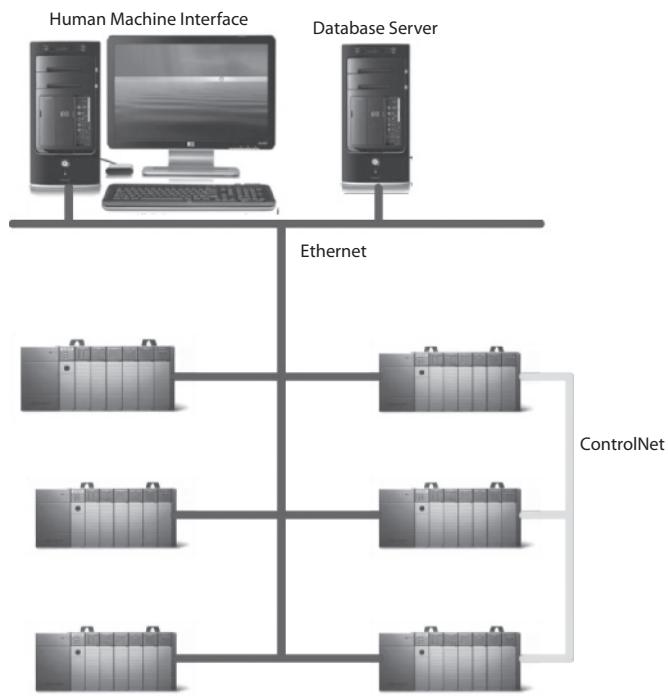
points, and redundant operation with hot backup. Interfacing with third party equipment, such as x-ray machines, is essential as is positive tracking of luggage after it leaves the x-ray machine.

Most SCADA systems support a standard communication mechanism called OPC (OLE for process control). An OPC interface allows third party OPC compliant software to interface with the SCADA system through either the database server or the HMI. This interface is particularly valuable when the mechatronics model based design approach is employed. For example, third party software could be used to create a dynamical real-time model of the industrial process that is to be controlled. The I/O of the model could be communicated to the SCADA system database from which the individual PLC controllers could process the data and provide the feedback control signals back to the model. Systems designed to work in this manner must incorporate a provision for the PLCs to either read and write to the physical I/O or to read and write from the SCADA database internal I/O.

Basics of PLC Programming The PLC utilizes a unique form of programming referred to as *ladder programming*. The ladder diagram provides a method of displaying the logic, timing, and sequencing of the system. The ladder program contains instructions (Figure 5-20) which represent external input and output devices and several other instructions to be used in the user program. The user program is scanned during normal operation of the PLC controller and the state of inputs and outputs are examined to update the programmed ladder logic.

A ladder program consists of two vertical *rails* connected by *rungs*. The program execution begins at the top left and travels across the first rung from left (the input side of the rung) to right (the output side of the rung). Program execution then moves down to the next rung and again executes from left to right, and so on until all rungs have been executed. Each instruction has a related address which identifies it as a physical input, physical output or an internal point. Physical inputs and outputs have actual

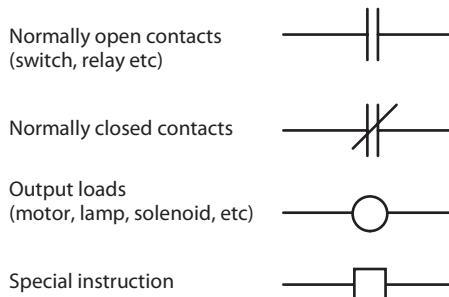
FIGURE 5-20



real world devices hardwired to them (contacts, timers, counters, and others). Internal inputs and outputs are not connected to any real world device but through programming are used to control outputs.

During PLC controller operation, the processor determines the ON/OFF state of the bits in the input state array that was copied to RAM. Once the processor determines the state of the bits in the data file, it then evaluates the rung logic and calculates the state of the outputs according to the logical continuity of the rungs in the user program. The output values are then written to the output state array (also located in RAM).

FIGURE 5-21 SYMBOL FORMAT IN LADDER LOGIC



Features of Programmable Controller Programming a PLC is supported either with the aid of a circuit diagram, a ladder diagram, or logic equations in a textual form. The programming system consists of a keyboard device to enter the control logic and other data or the video display and it permits the programmer to use either a relay ladder diagram or other programming language to input the control logic into memory. The power supply drives the PLC and serves as a source of power for the output signals. It is also used to protect the PLC against noise in the electrical power lines. The operating cycle consists of a series of operations performed sequentially. They are input scan, program scan, output scan, and service communications.

The main elements in a ladder logic are

- Rails
- Rungs
- Branches
- Inputs
- Outputs
- Timer
- Counter

Rails are vertical lines and provide the source of energy to relays and logic system. Rungs are horizontal and contain the branches, inputs, and outputs. As an example of the input, *Examine On* is present when the input is *ON*. *Examine Off* is active when the input is *OFF*. The output is referred to as a coil and it is on the right side of the rung.

A ladder program consists of a set of instructions used to control a machine or a process. Logic sequences entered into the microcontroller makes up a ladder program. Ladder logic is a graphical

programming language based on electrical relay diagrams. Instead of having electrical rung continuity, ladder logic looks for logical rung continuity. A ladder diagram identifies each of the elements in an electromechanical circuit and represents them graphically. This allows you to see how your control circuit operates before you actually start the physical operation of your system. In a ladder diagram, each of the input devices are represented in series or parallel combinations across the rungs of the ladder. The last element on the rung is the output that receives the action as a result of the conditional state of the inputs on the rung.

Instruction Set Overview PLCs are *reduced instruction set computers* (RISC) specifically designed for industrial control applications. The following overview of the instruction set is intended to provide a listing of the instructions in the set with a brief description of each instruction.

The instruction set can be divided into the following subsets:

- Bit instructions
- Timer and counter instructions
- Communications instructions
- ASCII instructions, input/output (I/O), and interrupt instructions
- Comparison instructions
- Math instructions: move and logical instructions
- Copy file and fill file instructions
- Bit shift, FIFO, and LIFO instructions
- Sequence instructions, control instructions, and proportional integral derivative instructions

Bit Instructions The first subset of instructions are bit instructions, which are conditional instructions which can refer to input or output either. They are the most widely used instructions in the programming of PLCs. The first of these instructions is the *Examine if Closed* (XIC) instruction. This instruction is a conditional input instruction which examines the state of a memory location or I/O address bit in the PLC and becomes true when the bit is on or (1).

The next instruction is the *Examine if Open* (XIO) instruction. This instruction is a conditional input instruction which examines the state of a memory location or I/O address bit and is true when the bit is off or (0).

The final bit instruction is the *Output Energize* (OTE) instruction. This is an output instruction which becomes true or (1) when the conditions of the bits preceding it are true. The output becomes false or (0) when one condition of the bits in the logical sequence preceding the output is false.

Timer and Counter Instruction Timers and counters are output instructions which have common instruction parameters used to set up the timing accuracy, timebases, accumulated value (ACC), and preset value (PRE). Timers and counters also have status bits depending on the type of timer or counter instruction. The first instruction in this subset is the *Timer On Delay* (TON) instruction. This output instruction counts time intervals when the conditions of the bits preceding it in the rung are true. The output of the timer is true when the ACC of the timer is equal to or greater than the PRE. The status bits for this instruction are the *Timer Done Bit* (DN) which is set when the output of the timer is true, the *Timer Enable Bit* (EN) which is set when the rung conditions are true and

is reset when the rung conditions are false, and the *Timer Timing Bit* (TT) which is set when the rung conditions are true and the ACC is less than the PRE and is reset when the DN is true or the rung conditions are false. The next instruction is the *Timer Off Delay* (TOF) instruction. This output instruction counts time intervals when the rung conditions preceding it are false. The output of the timer is true when the timer is initially enabled by the rung conditions becoming true and the output remains true when the rung conditions of the timer become false and remain false until the PRE of the timer reaches the ACC.

Communications Instructions The communications instructions are output instructions which are used to communicate between PLCs connected to different nodes on a PLC network.

The first instruction is the *Message Read/Write* (MSG) instruction. This instruction transfers data from one node to another on a communications network. When enabled, the message transfer is pending until the actual transfer takes place at the end of the program scan. The second instruction is the *Service Communications* (SVC) instruction. When the conditions of the rung preceding this instruction are true, the instruction interrupts the scan of the program to execute the service communications portion of the operating cycle.

Sequence Instructions Sequence instructions are output instructions which are used in sequential machine control applications. Several parameters for sequences must be established for proper operation.

Control Instructions Control instruction are conditional or output instructions which allow the user to change the order in which the processor scans the program. The purpose of these instructions are to minimize scan time, create a more efficient program, and provide diagnostic programming tools to facilitate troubleshooting.

Input and Output Devices The two types of I/O devices available to the systems integrator are discrete and analog. Analog input devices have a continuous range associated with their output state. Examples of analog input devices are transducers that output a 4-20 mA or 0-10 Vdc signal based upon input conditions (such as a change in temperature, pressure, stress and strain, or weight). Other types of analog input devices include potentiometers, which output a continuously varying resistance in Ω .

Discrete output devices are those which, when actuated, have only an ON or OFF state. Examples of discrete output devices are pilot lights, electro-mechanical relays and counters, pneumatic and hydraulic solenoid valves, and a variety of horns, buzzers, or other similar devices. Another discrete input device is an optical encoder, which generates a pulse train of ON and OFF signals based upon the relative position of an input shaft. This type of device typically has 1024 pulses per revolution of the input shaft. High-speed counters are required when encoders are employed as input devices in process solutions.

Ladder Logic Diagram provides a method of displaying the logic, the timing and sequencing of the system. Based on Boolean logic, the ladder diagram shows the steps of a process that is controlled by a sequence of discrete events.

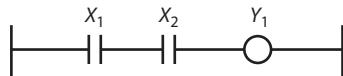
The first type of logic is series logic (AND) which will energize the output when all input conditions are true in a series path preceding an output (Figure 5-22(a)). The next type of logical continuity is parallel (OR) logic. In this case, when one or another path of logic are true, the output is energized (Figure 5-22(b)).

The typical PLC instructions used depends upon the manufacturer. Table 5-10 shows the PLC instruction code used by Mitsubishi.

FIGURE 5-22 SERIES AND PARALLEL INPUT LADDER DIAGRAMS

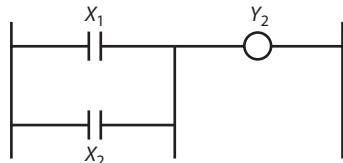
- (a) Output Y_1 occurs when input X_1 AND X_2 occur

$$Y_1 = X_1X_2$$



- (b) Output Y_2 occurs when input X_1 OR X_2 occur

$$Y_2 = X_1 + X_2$$

**TABLE 5-10 PLC INSTRUCTION CODE**

Instruction Code	Description
LD	Start a rung with an open contact
LDI	Start a rung with a closed contact
AND	A series element with an open contact
ANI	A series element with a closed contact
ANB	Branch two blocks in series
OR	A parallel element with an open contact
ORI	A parallel element with a closed contact
ORB	Branch two blocks in parallel
OUT	An output

Typical AND program for Figure 5.22(a) is,

```
LD   X1
AND X2
OUT Y1
```

Typical OR program for Figure 5.22(b) is,

```
LD   X1
OR   X2
OUT Y1
```

The designer can use an *input branch* in the application program to allow more than one combination of input conditions to form parallel branches (OR-logic conditions). Figure 5-23b uses an input branch to allow more than one combination of input conditions to form parallel branches. If either of these OR branches forms a true logic path, the output will be energized. If neither of the parallel branches forms a true logic path, the output will not energize.

This concept of branching also can be utilized for output portions of a rung. The user can program parallel outputs on a rung to allow a true logic path to control multiple outputs. When there is a true logic path, all parallel outputs become true.

Input and output branches can be nested to provide a more efficient form of PLC program. The need for redundant contacts is eliminated, and consequently, the scan time for the processor is reduced. A nested branch is one in which logical functions start and end within a branch.

Figure 5.23(a) shows an example of linking two parallel networks in series to one output using “ANB.” Figure 5.23(b) shows an example of linking two ladder-rung series in parallel to one output using “ORB.”

FIGURE 5-23 (A) PARALLEL INPUTS (B) SERIES INPUTS

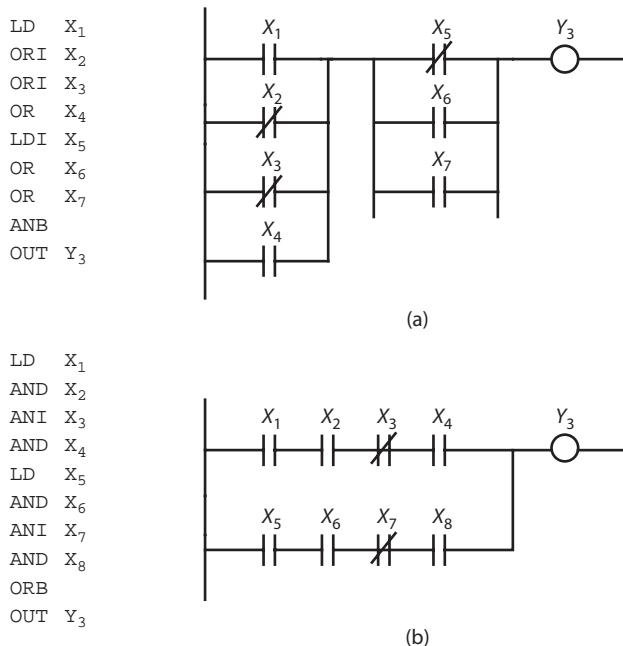
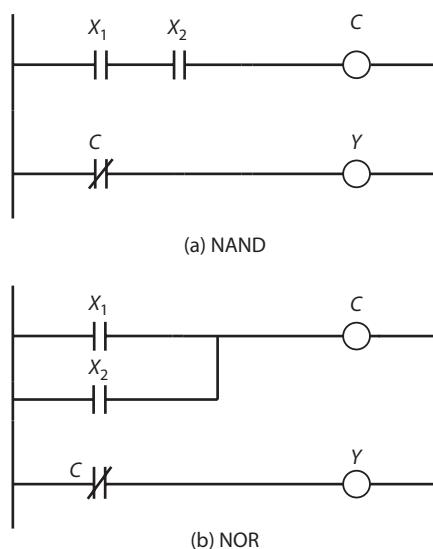


FIGURE 5-24 (A) NAND (B) NOR



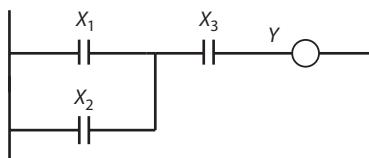
EXAMPLE 5.9

Construct the ladder logic diagrams for the following Boolean logic equations: (a) $Y = (X_1 + X_2) X_3$, (b) $Y = (X_1 + X_2) (X_3 + X_4)$, (c) $Y = (X_1 X_2) + X_3$,

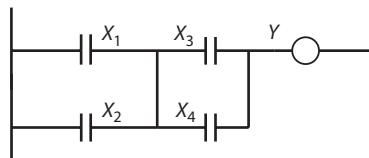
Solution

Ladder logic diagrams, as shown in Figure 5-25.

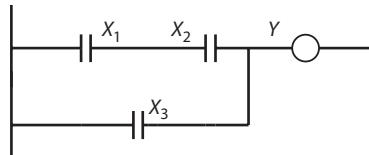
FIGURE 5-25 (A) $Y = (X_1 + X_2) X_3$ (B) $Y = (X_1 + X_2) (X_3 + X_4)$ (C) $Y = (X_1 X_2) + X_3$



(a)



(b)

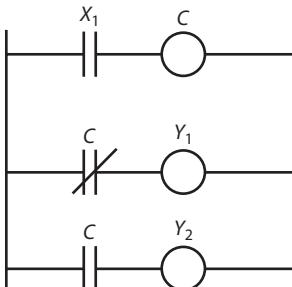
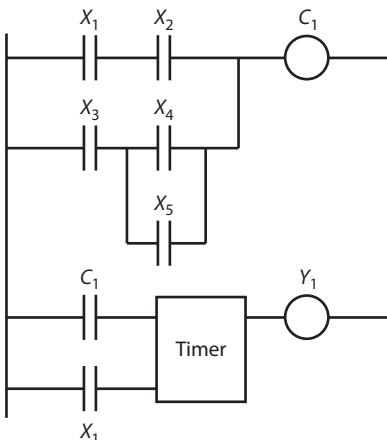


(c)

Relays Relays are the most popular components of the PLC hardware. Relays are used as outputs in the ladder diagram. They can be also used to control ON/OFF actuation of a powered device. A relay can be *latching* or *non-latching*. A latching relay needs an electrical impulse to close the power circuit. Another impulse is needed to release the latch. Non-latching relays hold only while the switching relay is energized and require continuous electrical signal. Relays (Figure 5-26) are useful in triggering next steps in the development of an automatic process after verifying the completion of the previous step. It is analogous to the closed-loop control approach.

In Figure 5-27, the control relay is shown by load C, which controls the on/off operation of two output loads (such as motors) Y_1 and Y_2 . When the control switch is closed, the relay becomes energized.

During normal controller operation, the processor checks the state of the input data file bits, then executes the program instructions individually—rung by rung—from the beginning to the end of the program. As it does, it updates the data file bits and energizes the appropriate output-data file bits accordingly. Data associated with external output is transferred from the output-data file to the output terminals, which are hardwired to the actual output devices. Also, during the I/O scan, the inputs are scanned to determine their state, and the associated ON/OFF state of the bits in the input data file are changed accordingly. During the program scan, the updated status of the external input

FIGURE 5-26 USE OF RELAYS**FIGURE 5-27 USE OF TIMER AND INTERNAL CONTROL RELAY**

devices are applied to the user program. The processor processes all the instruction in ascending rung order. Bits are updated according to logical Boolean continuity rules as the program scan moves from instruction to instruction through all rungs in the program.

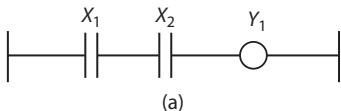
EXAMPLE 5.10

An industrial furnace is to be controlled as follows. The contacts of a bimetallic strip inside the furnace close if the temperature falls below the set point and open when the temperature is above the set point. The contacts regulate a control relay, which turns on and off the heating elements of the furnace. If the door to the furnace is opened, the heating elements are temporarily turned off until the door is closed.

- (a) Specify the input/output variables for this system operation and define symbols for them (e.g., X_1 , X_2 , C_1 , Y_1 , etc.). (b) Construct the ladder logic diagram for the system. (c) Write the low-level language statements for the system.

Solution

- (a) Let X_1 = temperature below set point, X_2 = door closed, and Y_1 = furnace on. Refer to Figure 5-28.

FIGURE 5-28 (A) LADDER LOGIC DIAGRAM (B) LOW-LEVEL LANGUAGE

LD	X1
AND	X2
OUT	Y2

(b)

EXAMPLE 5.11

In the manual operation of a sheet-metal stamping press, a two-button safety interlock system is often used to prevent the operator from inadvertently actuating the press while his hand is in the die. Both buttons must be depressed to actuate the stamping cycle. In this system, one press button is located on one side of the press while the other button is located on the opposite side. During the work cycle, the operator inserts the part into the die and depresses both pushbuttons, using both hands.

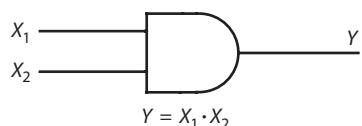
- Write the truth table for this interlock system.
- Write the Boolean logic expression for the system.
- Construct the logic network diagram for the system.
- Construct the ladder logic diagram for the system.

Solution

Let X_1 = button one, X_2 = button 2, and Y = safety interlock. Refer to Figure 5-29.

FIGURE 5-29 (A) TRUTH TABLE (B) BOOLEAN LOGIC EXPRESSION (C) LADDER DIAGRAM

X_1	X_2	Y
0	0	0
0	1	0
1	0	0
1	1	1



5.5 Summary

Mechatronics brings together areas of specialization that involves sensors, signal conditioning, hardware interface, control systems, actuation systems, and the technology of microprocessors. The signal conditioning by analog and digital electronics are the primary components of mechatronic system. In a general sense, signal-conditioning devices consist of elements that start with sensor output signal and provide a suitable signal for further control or display. They normally include electronic devices that perform the functions of amplification, impedance matching, filtering, modulating, comparing, and converting the data. In this chapter, digital electronics is introduced initially through Boolean algebra, including implementation of optimal design using minimization techniques. In the analog electronics section, amplifier selection is addressed through a discussion of various types of operational amplifiers and followed by a discussion of analog-to-digital conversion techniques. The chapter ends with a section on Programmable logic controllers that use programmable memory to store instructions, to implement logical and timing sequences, and to perform control actions.

REFERENCES

- Smaili, A., Mirad, F., *Applied Mechatronics*, Oxford University Press, NY, 2008.
- Bolton, W., *Programmable Logic Controllers*, Second Edition, Newnes, Woburn, MA, 2000.
- Barney, G.C., *Intelligent Instrumentation, Microprocessor Applications in Measurement and Control*, Second Edition, Prentice Hall, Englewood Cliffs, NJ, 1988. 532 pp.
- Bollinger, J.G., Duffie, N.A., *Computer Control of Machines and Processes*. Addison-Wesley Publishing Company, 1988.
- Garrett, P.H., *Advanced Instrumentation and Computer I/O Design*. IEEE. Wiley-Press, 1994.
- Johnson, C., *Process Control Instrumentation Technology*. John Wiley & Sons, 1982.
- Pallas-Aveny, R. and Webster, J., *Sensor and Signal Conditioning*. John Wiley & Sons, 1991.
- Rembold, U., *Computer-Integrated Manufacturing Technology and Systems*. Marcel Dekker, Inc., 1985.
- Advanced Programming Software Reference Manual*, 1747-PA2E Publication 1747- 6.11, August 1994.

PROBLEMS

- 5.1. The manufacturing cell will operate only if certain conditions are met. For the cell to start, one of two start buttons (X and Y) must be pressed, and the guard (G) must be in position. The cell is designed to stop if the safety guard is disturbed or if either of two stop buttons (S_1 and S_2) is pressed. The sensor monitoring the guard sends a 1 whenever the guard is in its right position. Otherwise, the sensor transmits a 0. The start and stop buttons are activated by relay sensors, which in turn will send 1's when pressed.

Design a logic circuit to monitor the cell.

- 5.2. In a machining operation using a horizontal boring machine, assume that sensors have been installed to measure cutter vibration (v), product surface roughness (s), product dimensional accuracy (a), and cutter temperature (t). Assume that the sensors send the following digital signals:

$$v = 1 \text{ for excessive vibration}$$

$$t = 1 \text{ for high temperature}$$

$$s = 1 \text{ for poor product surface}$$

$$a = 1 \text{ for poor quality}$$

otherwise these signals are zeros.

Design a logic circuit which has two outputs codes: yellow (Y) and red (R). Code yellow is a 1 if any one of the sensor signals is a 1. Code red is a 1 if more than one of the sensor signals is 1, otherwise both outputs are zeros.

- 5.3. Consider a chemical tank for which there are three variables to be monitored are, (a) level (b) pressure (c) temperature. The alarm will sound if the liquid level is high and the temperature is high. Another condition for alarm is a combination of high liquid level with low temperature and high pressure.

Design the circuit such that an alarm is sounded when certain combinations of conditions occur between the variables.

TABLE P5-4

A	B	C	D	Start	A	B	C	D	Start
0	0	0	0	0	1	0	0	0	0
0	0	0	1	1	1	0	0	1	x
0	0	1	0	0	1	0	1	0	0
0	0	1	1	1	1	0	1	1	0
0	1	0	1	0	1	1	0	0	x
0	1	1	0	0	1	1	0	1	0
0	1	1	1	0	1	1	1	0	0
0	1	1	1	0	1	1	1	1	x

- 5.4. A metal punching press with pneumatic logic shall operate when the four combinations defined in Table 5-4 exist and should not operate if any other combination exists.

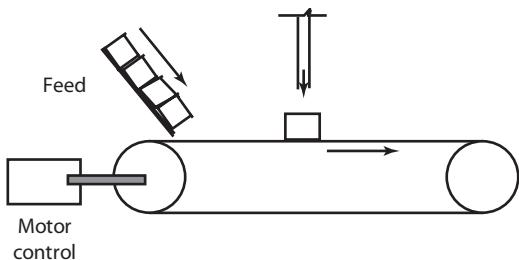
Design a logic system for starting. The signal from the sensor operated by the guard is *A*, the signal from the operator is *B*, and the signal from workpiece is *C*. *D* is the signal from the remote sensor. (*x* represents don't care in truth table).

- 5.5. An on-line manufacturing work cell performs a series of four quality control tests on a manufactured product. *A,B,C* and *D* are identified as four tests or inputs to the logic system. Bins #1, #2, and #3 are classified as outputs. If the product passes two OR three tests, bin #1 will receive the part. If it passes one of the tests, bin #2 will be open. Bin #3 accepts perfect units only.

Design a logic system that will simultaneously examine the results of all four tests and decide into which of the three output containers the piece will drop.

- 5.6. A bottling plant uses an automated mechanism for filling the container and transporting them from one point to other as shown in Figure P5-6. The sensors monitor the amount of solid or liquid filled. A conveyor mechanism transports the containers.

Design a mechatronic system for the case described. Identify the types of sensors you used, describe how they work, and explain how you are going to interface and control them. Make suitable sketches if needed.

FIGURE P5-6

- 5.7. A transducer used for temperature measurement in a chamber provides an output of $5\text{mV}/^\circ\text{C}$. The range of temperature measurement is from 0 to 100°C . A sixbit A/D converter is used. Reference voltage is 12 V.

Find the input voltage. Design a A/D converter to provide the required temperature resolution.

5.8. Write

- The binary equivalent of A90E; 44.17_{10} , $9CA_{16}$, $.6875_{10}$
- Hexadecimal equivalent of 101100000101110
- Decimal equivalent of the binary 1111.1010_2

5.9. Simplify the following.

(a) $C = (A + \bar{A} \cdot B) \cdot (A + \bar{B})$

(b) $X = U \cdot V + V \cdot W + U \cdot W + V \cdot \bar{W}$

(c) $D = \bar{A} \cdot B \cdot C + A \cdot B \cdot \bar{C} + A \cdot B \cdot C + \bar{A} \cdot B$

(d) $C = (A \cdot \bar{B} + A \cdot B) \cdot (A \cdot B)$

(e) Negate $\bar{A} \cdot B + A \cdot \bar{B}$

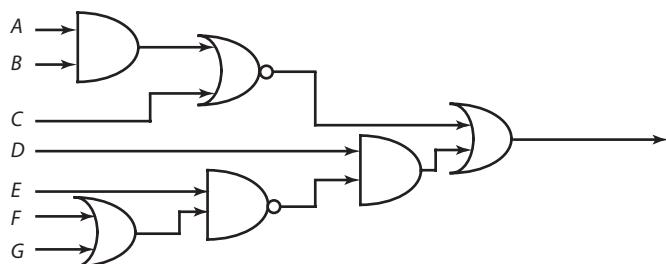
5.10. An absolute encoder grating consists of three bits. The white region represents transparency (1) and the black region represents opaqueness (0). The grating rotates clockwise. The first three sequences are 000, 001, 010. What are the remaining sequences for one full revolution?

5.11. (a) Is the following equation true or false?

$$(X \text{ NAND } Y) \text{ NAND } Z = X \text{ NAND } (Y \text{ NAND } Z)$$

(b) Transform the circuit in Figure P5-11, into an equivalent one that use only NAND gates.

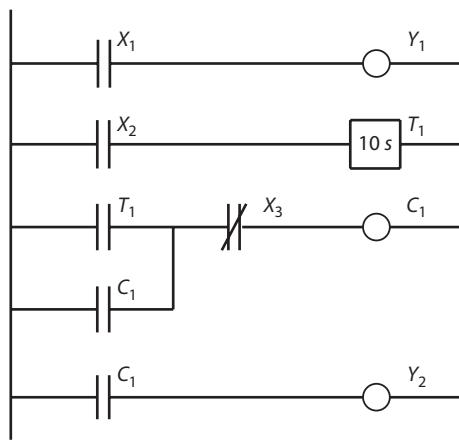
FIGURE P5-11



- 5.12. A level control system operates with two float sensors S_1 and S_2 , which are set at the minimum and the maximum levels, respectively. These produce signals 0 and 1 depending on whether they are tripped or not. The level in the tank is to be kept within the minimum and maximum values while some fluid is drawn off. A pump P is used to supply the fluid and excess fluid gets drained off by a solenoid operating output valve V . Both the pump and the solenoid operating valve are switched ON by logic-level control signals, where level 1 switches the device ON and level 0 switches the device OFF.

The answer to this problem can be presented either as a Boolean expression with logic circuits or as a relay logic diagram using PLC.

FIGURE P5-12 INDUSTRIAL ROBOT EXAMPLE



- 5.13. An industrial robot performs a machine loading and unloading operation. A PLC is used as the robot cell controller. The cell operates as follows: (1) a human worker places a workpart into a nest, (2) the robot reaches over and picks up the part and places it into an induction heating coil, (3) a time of 10 seconds is allowed for the heating operation, and (4) the robot reaches in and retrieves the part and places it on an outgoing conveyor. A limit switch X_1 (normally open) will be used in the nest to indicate part presence in step (1). Output contact Y_1 will be used to signal the robot to execute step (2) of the work cycle. This is an output contact for the PLC, but an input interlock for the robot controller. Timer T_1 will be used to provide the 10 second delay in step (3). Output contact Y_2 will be used to signal the robot to execute step (4). Construct the ladder logic diagram and write the low level language statements for the system.

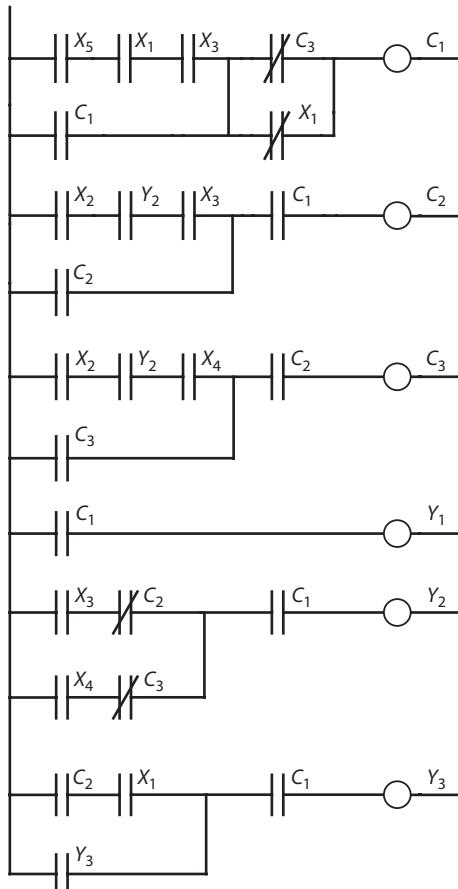
Suggested solution: Ladder logic diagram

- 5.14. A PLC is used to control the sequence in an automatic drilling operation. A human operator loads and clamps a raw workpart into a fixture on the drill press table and presses a start button to initiate the automatic cycle. The drill spindle turns on, feeds down into the part to a certain depth (the depth is determined by limit switch), and then retracts. The fixture then indexes to a second drilling position, and the drill feed-and-retract is repeated. After the second drilling operation, the spindle turns off, and the fixture moves back to the first position. The worker then unloads the finished part and loads another raw part. Let the input/output variables for this system operation be (X_1, X_2, C_1, Y_1 etc.). As a first step, construct the ladder logic diagram and write the low level language statements for the system using the PLC instruction.

Suggested Solution: Let X_1 = spindle up
 X_2 = spindle at desired depth
 X_3 = fixture at position 1
 X_4 = fixture at position 2
 X_5 = start button
 Y_1 = spindle on
 Y_2 = spindle down
 Y_3 = fixture to position 2
 C_1 = drill cycle permit
 C_2 = hole 1 drilled

Ladder logic diagram:

FIGURE P5-14 AUTOMATED DRILLING EXAMPLE



CHAPTER 6

SIGNALS, SYSTEMS, AND CONTROLS

- 
- 6.1 Introduction to Signals, Systems, and Controls
 - 6.2 Laplace Transform Solution of Ordinary Differential Equations
 - 6.3 System Representation
 - 6.3.1 Transfer Function Form
 - 6.3.2 Basic Feedback System and G-Equivalent Form
 - 6.4 Linearization of Nonlinear Systems
 - 6.5 Time Delays
 - 6.6 Measures of System Performance
 - 6.6.1 Stability
 - 6.6.2 Accuracy
 - 6.6.3 Transient Response
 - 6.6.4 Sensitivity
 - 6.7 Root Locus
 - 6.7.1 Fundamentals
 - 6.7.2 Sketching Rules
 - 6.7.3 Sketching Examples
 - 6.7.4 Controls
 - 6.8 Bode Plots
 - 6.8.1 Controls
 - 6.9 Controller Design Using Pole Placement Method
 - 6.10 Summary
 - References
 - Problems

This chapter provides the student with the basic tools and experience necessary to design and analyze basic single-input/single-output control systems. Following some essential introductory material which includes definitions and terminology, we discuss techniques used for system and performance representation based on transfer functions and block diagrams. A review of Chapters 1 and 2 may be necessary for those somewhat unfamiliar with either of these topics. Linearization, time delays, and the Laplace transform are then introduced. Analysis techniques using root locus and Bode plots are discussed and followed by a description of standard control structures and their application. Design steps and examples using the standard control structures (which include lead, lag, rate feedback, PI, PID, and gain) are presented in the final sections.

6.1 Introduction to Signals, Systems, and Controls

A system (or plant) is a naturally occurring or man-made entity which transforms causes (or inputs) into effects (or outputs). System behavior can be modified by interactions with other systems. Modification of the behavior of a system such that a desired behavior is achieved is called *control*. Controls are implemented by attaching a controller or compensator to the plant. The resulting combined system is called a control system. Control systems incorporate either human or machine controllers. When the controller is machine based, it is called automatic control.

Within any control system are variables and functions. Variables can be either constant or may vary with respect to some independent variable. Constant variables are called *parameters*, and

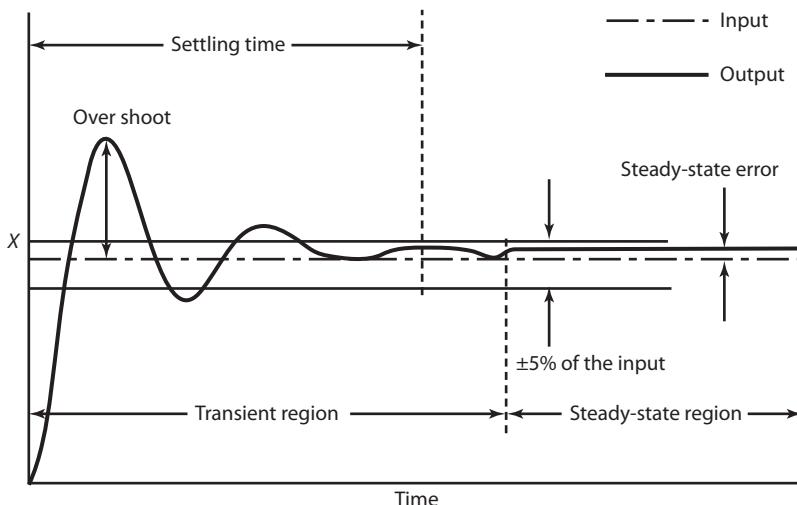
varying variables are called *signals*. Signals evolve (or change) with respect to an independent variable, usually time. The behavior of a signal is often considered in two regions:

Transient Region: In this region, the signal derivatives dominate the shape of the signal.

Steady-State Region: In this region, all signal derivatives die out, leaving only the offset or DC value.

Examples of the transient and steady-state regions for an arbitrary signal, $x(t)$, are shown in Figure 6-1. Any response can be quantified by measuring defined waveform characteristics, such as those given here.

FIGURE 6-1 TRANSIENT AND STEADY-STATE REGIONS OF A SIGNAL



Rise Time: This is the amount of time the system takes to go from 10 to 90% of the steady-state (or final) value.

Percent Overshoot (P.O.): This is the amount that the process variable overshoots the final value—expressed as a percentage of the final value. The expression for percent overshoot for a unit step response of a second-order system is

$$\text{P.O.} = 100e^{-\zeta\pi/\sqrt{1-\zeta^2}}$$

where ζ is the damping ratio.

Steady-State Error (e_{ss}): This is the final difference between the process variable and set point.

Settling time (T_s): The time required for the process variable to settle to within a certain percentage (commonly 5%) of the final value. The expression for settling time for a unit step response of a second-order system is

$$T_s = \frac{\log(e_{ss})}{\zeta\omega_n}$$

where ω_n is the natural frequency.

There are four categories of signals (Table 6-1) in any control system.

TABLE 6-1 FOUR BASIC CONTROL SYSTEM SIGNALS

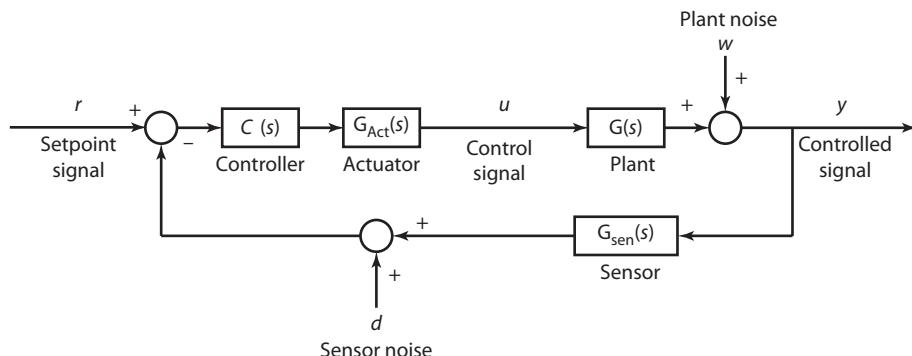
Signal Name	Function	Typical Variable
Reference, command, or setpoint signals	These are external (exogenous) commands signals provided to the controller.	$r(t), y^*(t)$
Control signals	These are input signals created by the controller and provided to the plant.	$u(t)$
Controlled signals	These are output signals created by the plant which are to be controlled.	$y(t)$
Disturbance signals	Noise or other disturbances reflecting sensor noise, variations in plant parameters (due to linearization), and changes in the environment of operation.	$d(t), w(t)$

In addition to the four categories of signals, there are four basic functions also found in any control system, as shown in Table 6-2.

TABLE 6-2 FOUR BASIC CONTROL SYSTEM FUNCTIONS

Signal Name	Function	Typical Variable
Compensator or controller	The controller system which is attached to the plant through sensors and actuators to modify its overall performance.	$C(s), G_c(s)$
Process, plant, or uncontrolled system	The system or process which is to be controlled.	$G(s), G_p(s), T(s)$
Sensors	A device which converts a physical quantity (temperature, pressure, etc.) into a low-power electrical signal capable of being read by a computer.	$G_{Sen}(s)$
Actuators/drive	A device which converts a low-power command signal (from a computer) to a high-power signal, which creates motion, heat, pressure, etc.	$G_{Act}(s)$

A diagram of a general control system with all signals and functions is presented in Figure 6-2, which presents a specific and fundamental configuration in which the controller is in cascade (series)

FIGURE 6-2 GENERAL CONTROL SYSTEM DIAGRAM

with the plant. There are other configurations for placing the controller and (in addition to providing the basic control function) each serves its own purpose. For example, a controller placed in the feedback path with derivative action might be used if we were concerned about amplifying reference derivatives.

6.2 Laplace Transform Solution of Ordinary Differential Equations

The Laplace transform is used extensively in control system analysis and design. This section summarizes the procedure for using the Laplace transform to solve ordinary differential equations (ODEs) or transfer functions which can be converted to ODE form.

In Chapter 2, we introduced the D operator and the Laplace s operator. For our modeling purposes, these operators were used interchangably to represent time differentiation; however, as we introduce the Laplace transform, we need to be a little more careful when applying these operators. The D operator is a time-domain operator. It is used as a notational convenience to write an ODE or transfer function in the time domain. The s operator is used to represent the ODE in a different domain called the Laplace domain (complex-variable domain). In this domain not only are the derivative of a signal represented by multiplications by the s operator, additional signal information (including initial conditions) is also included, which changes the form of the original ODE.

For example, consider the ODE given by

$$\dot{x}(t) = -3x(t) + r(t)$$

with $x(0) = -2$. Using the D operator, we can write the transfer function for this equation as

$$Dx(t) = -3x(t) + r(t) \rightarrow x(t) = \frac{r(t)}{D + 3}$$

Aside from being able to write the transfer function, the D operator form of the ODE does not provide any analysis tools, enabling us to analytically compute its solution.

On the other hand, if we take the Laplace transform of the equation, we write

$$sx(s) - x(0) = -3x(s) + r(s) \rightarrow x(s) = \frac{x(0) + r(s)}{s + 3}$$

First notice that the equation is no longer in the time domain, it is in the s domain (Laplace domain). In the s domain, we have access to tools which enable us to analytically solve the ODE. In going from the time domain to the Laplace domain, we use a Laplace transform table (Table 6-3) and properties of the Laplace transform which convert initial conditions. In the Laplace domain, the ODE is represented as an algebraic equation which can be solved and transformed back to the time domain using an inverse Laplace transform technique.

The Laplace transform is the preferred method for analytical solution of the response of continuous dynamic systems represented either as an ODE or as a transfer function. It is used to compute the total response (zero state plus zero input) of the system. Since any linear SISO system can be represented as a transfer function and subsequently as an ordinary differential equation (Chapter 2), a general procedure for the Laplace transform solution of an ODE is presented next.

TABLE 6-3 LAPLACE TRANSFORM TABLE (ONE SIDED, $t \geq 0$)

$F(s)$	$f(t)$
1	$\delta(t)$
$\frac{1}{s}$	$u(t)$
$\frac{1}{s+a}$	e^{-at}
$\frac{1}{(s+a)^n}$	$\frac{1}{(n-1)!} \cdot t^{n-1} \cdot e^{-at}$
$\frac{s+a}{(s+a)^2 + b^2}$	$e^{-at} \cos(bt)$
$\frac{b}{(s+a)^2 + b^2}$	$e^{-at} \sin(bt)$

Given: ODE (or transfer function converted to an ODE), initial conditions, output signal, $y(t)$, and input signal, $r(t)$.

Solution:

Step 1. Take the Laplace transform of the ODE on a term-by-term basis. Initial conditions are included using the differentiation property of the Laplace transform, (Table 6-4), and the Laplace transform of the input is included using table entries (Table 6-3). By taking the Laplace transform of the ordinary differential equation, it has been transformed from the time domain to the s -domain.

TABLE 6-4 LAPLACE TRANSFORM PROPERTIES

Property	t Domain	s Domain
1. Time Delay	$f(t-\tau)$	$F(s) \cdot e^{-s\tau}$
2. Time Scaling	$f(at)$	$\frac{1}{ a } \cdot F(s/a)$
3. Differentiation	$f^{(n)}(t)$	$L\{\dot{y}(t)\} = sY(s) - y(0)$ $L\{\ddot{y}(t)\} = s^2Y(s) - \dot{y}(0) - sy(0)$ $L\{\ddot{\ddot{y}}(t)\} = s^3Y(s) - \ddot{y}(0) - s^2y(0) - s\dot{y}(0)$ \vdots

Step 2. Solve the s -domain algebraic equation (from step 1) for the desired output variable using

$$Y(s) = \frac{N(s)}{D(s)}$$

Step 3. Perform a partial fraction expansion on the step 2 function:

$$Y(s) = \frac{A}{s+a} + \frac{B}{s+b} + \dots$$

where A, B, a, b may be complex

Step 4. Take the inverse Laplace transform of $Y(s)$ to compute $y(t)$ using the Laplace transform table (see Table 6-3).

In addition to the Laplace transform table, the following properties (Table 6-4) of the Laplace Transform are used extensively.

The third property, differentiation, is used to capture initial conditions associated with derivative terms when taking the Laplace transform. For example, consider the following second-order ODE and initial conditions.

$$\ddot{x}(t) = -\dot{x}(t) + 3x(t) + r(t); \dot{x}(0) = -1, x(0) = 2$$

Using this property, the Laplace Transform is computed as

$$\begin{aligned}s^2X(s) - sX(0) - \dot{X}(0) &= -(sX(s) - X(0)) + 3X(s) + R(s) \\ s^2X(s) - 2s + 1 &= -(sX(s) - 2) + 3X(s) + R(s) \\ (s^2 + s - 3)X(s) &= R(s) + 2s + 1\end{aligned}$$

Partial fraction expansion provides a method for decomposing a strictly proper transfer function (strictly proper means that the order of the numerator polynomial is less than the order of the denominator polynomial) into a sum of first- and second-order terms which can be found in the inverse Laplace transform technique.

In situations where the transfer function is not strictly proper, the numerator first must be divided by the denominator to produce a constant, s -terms, plus a strictly proper term which can be expanded using partial fractions.

To illustrate the overall Laplace transform procedure, we'll consider several example problems, each focusing on different situations encountered in practice.

EXAMPLE 6.1 Conventional Partial Fractioning and Inverse Laplace Transform

$$Y(s) = \frac{s + 4}{s^3 + 6s^2 + 11s + 6} = \frac{s + 4}{(s + 1)(s + 2)(s + 3)}$$

Partial fraction form permits us to write the transfer function as a sum of the factored terms each multiplied by an unknown coefficient A , B , and C (called the partial fraction coefficients or residuals).

$$Y(s) = \frac{s + 4}{(s + 1)(s + 2)(s + 3)} = \frac{A}{s + 1} + \frac{B}{s + 2} + \frac{C}{s + 3}$$

In addition, the s value which makes the denominator of each term zero is called a singularity. The singularity of the first term is -1 , the second term is -2 , and the third term is -3 .

Solution

The residuals are solved as follows. This will be referred to as the conventional form, and we will illustrate the process to compute the A residual.

First, multiply through by the denominator associated with A to isolate the numerator of the A term.

$$Y(s) \cdot (s + 1) = \frac{s + 4}{(s + 2)(s + 3)} = A + \frac{B \cdot (s + 1)}{s + 2} + \frac{C \cdot (s + 1)}{s + 3}$$

In this equation, we can select any value of s , and the equality will still hold. If we select s such that all right-hand terms disappear except the A term, we will be able to complete the solution for A . To accomplish this, s is selected to equal the singularity associated with the A term, ($s = -1$), the numerators of the B and C terms become zero, and A is evaluated as

$$Y(s) \cdot (s + 1)|_{s=-1} = \frac{s + 4}{(s + 2)(s + 3)} \Big|_{s=-1} = A = \frac{3}{2}$$

The procedure is repeated to solve for B and C :

$$\begin{aligned} Y(s) \cdot (s + 2)|_{s=-2} &= \frac{s + 4}{(s + 1)(s + 3)} \Big|_{s=-2} = B = -2 \\ Y(s) \cdot (s + 3)|_{s=-3} &= \frac{s + 4}{(s + 1)(s + 2)} \Big|_{s=-3} = C = \frac{1}{2} \end{aligned}$$

The final solution is obtained by taking the inverse Laplace transform using Table 6-3.

$$Y(s) = \frac{3/2}{s + 1} - \frac{2}{s + 2} + \frac{1/2}{s + 3} \rightarrow y(t) = \frac{3}{2}e^{-t} - 2e^{-2t} + \frac{1}{2}e^{-3t}$$

EXAMPLE 6.2 Deflation and Partial Fraction Expansion

B and C also could be computed by a technique known as deflation which is described as follows. Imagine A has been found. The A term then can be subtracted from both right and left sides of the equation to yield a new equation:

$$Y_1(s) = Y(s) - \frac{A}{s + 1} = \frac{B}{s + 2} + \frac{C}{s + 3}$$

Solution

The right-hand side of this new equation is of second order, but the left-hand term, $Y_1(s) = Y(s) - \frac{A}{s + 1}$, appears to be third order (the same as $Y(s)$). Lets look at this more closely.

$$\begin{aligned} Y_1(s) = Y(s) - \frac{A}{s + 1} &= \frac{s + 4}{(s + 1)(s + 2)(s + 3)} - \frac{3/2}{s + 1} \\ &= \frac{(s + 4) - 3/2(s + 2)(s + 3)}{(s + 1)(s + 2)(s + 3)} \\ &= \frac{-(3/2)s^2 - (13/2)s - (10/2)}{(s + 1)(s + 2)(s + 3)} = \frac{-1}{2} \cdot \frac{3s^2 + 13s + 10}{(s + 1)(s + 2)(s + 3)} \end{aligned}$$

For the left and right sides to be equal, there MUST be a common factor of $(s + 1)$ in the $Y_1(s)$ equation. This has to be, because $(s + 1)$ is the singularity that was subtracted from $Y(s)$ to create the right side of the equation. Lets check using long division:

$$\begin{array}{r} 3s + 10 \\ s + 1 \sqrt{3s^2 + 13s + 10} \\ \underline{3s^2 + 3s} \\ 10s + 10 \end{array}$$

As expected, the $(s + 1)$ term was common to both the numerator and denominator, leaving a second-order left-hand side.

$$Y_1(s) = \frac{-(1/2)(s + 1)(3s + 10)}{(s + 1)(s + 2)(s + 3)} = \frac{-(1/2)(3s + 10)}{(s + 2)(s + 3)} = \frac{B}{s + 2} + \frac{C}{s + 3}$$

Therefore, when a polynomial is deflated by subtracting one the residual associated with one of its poles, its order is reduced by one. Continuing with the example, B can be found conventionally as

$$B = (s + 2) \cdot Y_1(s)|_{s=-2} = -2$$

The remaining residual, C , can be solved by deflating $Y_1(s)$:

$$\begin{aligned} Y_2(s) &= Y_1(s) - \frac{B}{s + 2} = \frac{C}{s + 3} \\ &= \frac{-(1/2)(3s + 10) + 2(s + 3)}{(s + 2)(s + 3)} \rightarrow \text{Know } (s + 2) \text{ must be a factor of the numerator} \end{aligned}$$

The numerator is simplified to $(1/2)(s + 2)$, clearly showing the term. As expected, the common term $(s + 2)$ cancels, and the transfer function order is again reduced by one, yielding

$$Y_2(s) = \frac{(1/2)}{(s + 3)} = \frac{C}{s + 3} \text{ and } C = 1/2$$

Deflation is obviously a longer procedure than the conventional method of finding the residual values; however, it will tend to reduce the possibility of making an error by during hand calculations. This is due to the built-in feedback mechanism which requires a common term in the numerator and denominator which must cancel for the deflation to progress. If this common term does not appear, you know you have made an error during that part of the deflation process. Deflation also eliminates the need for using complex arithmetic when repeated roots are present. Deflation can be used to handle two shortcomings of the conventional partial fraction approach, repeated roots, and complex roots. These situations are presented in the following two examples.

EXAMPLE 6.3 Repeated Roots, Deflation, and Inverse Laplace Transform

Solve $Y(s) = \frac{1}{(s + 2)^3(s + 3)}$ for $y(t)$.

Solution

Form the partial fraction expansion

$$Y(s) = \frac{A_{13}}{(s + 2)^3} + \frac{A_{12}}{(s + 2)^2} + \frac{A_{11}}{(s + 2)} + \frac{B}{(s + 3)}$$

Evaluate the highest power residual conventionally as

$$A_{13} = (s + 2)^3 Y(s)|_{s=-2} = 1$$

Then, deflate the transfer function by subtracting off one of the three singularities at -2 :

$$\begin{aligned} Y_1(s) &= Y(s) - \frac{A_{13}}{(s+2)^3} = \frac{1}{(s+2)^3(s+3)} - \frac{1}{(s+2)^3} = \frac{-(s+2)}{(s+2)^3(s+3)} \\ &= \frac{-1}{(s+2)^2(s+3)} = \frac{A_{12}}{(s+2)^2} + \frac{A_{11}}{(s+2)} + \frac{B}{(s+3)} \end{aligned}$$

Compute the highest power residual conventionally as

$$A_{12} = (s+2)^2 Y_1(s)|_{s=-2} = -1$$

Deflate the transfer function by subtracting off one of the two remaining singularities at -2 :

$$\begin{aligned} Y_2(s) &= Y_1(s) - \frac{A_{12}}{(s+2)^2} = \frac{-1}{(s+2)^2(s+3)} - \frac{-1}{(s+2)^2} = \frac{(s+2)}{(s+2)^2(s+3)} \\ &= \frac{1}{(s+2)(s+3)} = \frac{A_{11}}{(s+2)} + \frac{B}{(s+3)} \end{aligned}$$

There are no more repeated roots and the remaining residuals are computed conventionally.

$$\begin{aligned} A_{11} &= (s+2)Y_2(s)|_{s=-2} = 1 \\ B &= (s+3)Y_2(s)|_{s=-3} = -1 \end{aligned}$$

The final solution is obtained by taking the inverse Laplace transform using the Laplace transform Table 6-3.

$$\begin{aligned} Y(s) &= \frac{1}{(s+2)^3} - \frac{1}{(s+2)^2} + \frac{1}{(s+2)} - \frac{1}{(s+3)} \\ Y(s) &= \frac{t^2}{2} e^{-2t} - t e^{-2t} + e^{-2t} - e^{-3t} \end{aligned}$$

EXAMPLE 6.4 Complex Roots

$$\text{Solve } Y(s) = \frac{10}{s^2 + 8s + 41}.$$

Solution

Any quadratic term (having complex roots), can be expressed in the factored form; $(s+a)^2 + b^2$. Applying this to the example transfer function, we obtain,

$$\begin{aligned} Y(s) &= \frac{10}{s^2 + 8s + 41} \rightarrow s^2 + 8s + 41 = (s+a)^2 + b^2 \\ &= s^2 + 2as + a^2 + b^2 \end{aligned}$$

We solve for the coefficients a and b by equating coefficients in like powers of s .

$$s^1 \text{ term: } 8 = 2a \rightarrow a = 4$$

$$s^0 \text{ term: } 41 = a^2 + b^2 \Big|_{a=4} \rightarrow 41 - 16 = b^2 \rightarrow b = 5$$

And the transfer function becomes

$$Y(s) = \frac{10}{s^2 + 8s + 41} = \frac{10}{(s + 4)^2 + 5^2}$$

Now we'll use the form found in last two rows of the Laplace Transform in Table 6-3 for sin and cosine:

$$Y(s) = \frac{10}{s^2 + 8s + 41} = \frac{10}{(s + 4)^2 + 5^2} = K_1 \frac{(s + 4)}{(s + 4)^2 + 5^2} + K_2 \frac{5}{(s + 4)^2 + 5^2}$$

The coefficients K_1 and K_2 are then found by equating coefficients in the numerator for like powers of s:

$$10 = K_1(s + 4) + 5K_2 \rightarrow 0s = K_1s \rightarrow K_1 = 0$$

and

$$10 = 4K_1 + 5K_2|_{K_1=0} \rightarrow K_2 = 2$$

The resulting form and solutions becomes

$$Y(s) = 2 \cdot \frac{5}{(s + 4)^2 + 5^2} \rightarrow y(t) = 2e^{-4t} \sin(5t)$$

6.3 System Representation

Systems are commonly represented in any of three forms: transfer function form, state-space form, and block diagram form. State-space form relies heavily on matrix-based calculations and is a necessary format for multi-variable, and multi input/output applications. It is represented by two vector equations: the state equation and the output equation. Some state space fundamentals were covered in Chapter 2. In this section, we'll concentrate on two forms: the transfer function and the block diagram. In particular, we will use the basic feedback system, introduced in Chapter 2, to develop an additional form called the G-equivalent form. We will not cover state-space form in this text.

6.3.1 Transfer Function Form

This form applies to linear systems which have a single input and a single output, often referred to as *SISO* systems.

The transfer function, introduced in Chapter 2 and repeated here, is a ratio of the input/output signals represented as polynomials in operator notation. The transfer function provides a concise means of representing an ordinary differential equation. Provided the equation has a single input signal and a single output signal, it is linear, proper, and has initial conditions all set to zero. The term “proper” means that the order of the numerator polynomial is less than or equal to the order of the denominator polynomial. A more thorough description of this term is given in Chapter 1.

To illustrate, consider a differential equation with input $R(t)$ and output $Y(t)$ represented in operator notation as

$$Y(t) = T(D) \cdot R(t) \text{ where } T(D) \equiv \frac{N(D)}{D(D)}$$

If the equation is linear, it can be rewritten by factoring $Y(t)$ and $R(t)$ out of each side as

$$Y(t) \cdot D(D) = R(t) \cdot N(D)$$

Here $D(D)$ and $N(D)$ are polynomials represented in the operator notation. The transfer function representation of the differential equation is presented in Equation 6-1.

$$\frac{\text{Output}}{\text{Input}} = \frac{Y(t)}{R(t)} = T(D) = \frac{N(D)}{D(D)} \quad (6-1)$$

where

$$\begin{aligned} N(D) &= a_m D^m + a_{m-1} D^{m-1} + \dots + a_1 D + a_0 \\ D(D) &= b_n D^n + b_{n-1} D^{n-1} + \dots + b_1 D + b_0 \end{aligned}$$

The transfer function is proper provided that the order or degree of the $D(D)$ polynomial is greater than or equal to that of the $N(D)$ polynomial. To be proper, $m \leq n$.

The leading coefficients of the $N(D)$ and $D(D)$ polynomials are (in general) not equal to 1. When they are equal to 1, the polynomial is said to be in *monic form*. The $N(D)$ polynomial could be made monic by dividing through by a_m and the $D(D)$ by dividing through by b_n .

Equation 6-2 presents the monic form of the transfer function.

$$\frac{\text{Output}}{\text{Input}} = \frac{Y(t)}{R(t)} = \frac{a_m}{b_n} \frac{(N(D)/a_m)}{(D(D)/b_n)} = k \cdot \frac{(N(D)/a_m)}{(D(D)/b_n)} \quad (6-2)$$

K is the scaling gain required to make the numerator and denominator polynomials of the transfer function monic.

The roots of the numerator of the transfer function $N(D) = 0$ are called zeros, and the roots of the denominator $D(D) = 0$ are called poles. The transfer-function denominator equation, $D(D) = 0$, is an important equation called the characteristic equation. The *characteristic equation* is universally written using the lower case Greek letter rho and defined as $\rho(D) \equiv D(D) = 0$.

Three examples are provided to demonstrate the use of transfer function form. The first example illustrates how a differential equation is converted into a transfer function. The second applies the transfer function to the design of a low-pass filter. The third utilizes the transfer function to approximate time differentiation.

EXAMPLE 6.5 Differential Equation to Transfer Function

Consider the differential equation presented in Equation 6-3 with input $R(t)$, output $Y(t)$, and zero initial conditions.

$$3\ddot{Y}(t) + 2\dot{Y}(t) + Y(t) = 7\ddot{R}(t) - R(t) \quad (6-3)$$

Rewriting Equation 6-3 in operator form yields

$$Y(t) \cdot (3D^3 + 2D + 1) = R(t) \cdot (7D^3 - 1) \quad (6-4)$$

The monic form of the transfer function is then written in Equation 6-5.

$$\frac{Y(t)}{R(t)} = \frac{7}{3} \cdot \frac{D^3 - 1/7}{D^3 + 2/3D + 1/3} \quad (6-5)$$

EXAMPLE 6.6 Low-Pass Filter Transfer Function

A noisy signal often can be smoothed by passing it through a low-pass filter. The premise here is that the noisy signal contains a message component which carries all of its useful information and a noise component. The message component is assumed to occur at low frequencies, and the noise component occurs at higher frequencies. The low-pass filter attenuates the high-frequency components (noise), but leaves the low-frequency components (message) unaltered.

Solution

Assume that the noisy signal consists of a sine wave which takes on frequencies below 10 Hz and high-frequency noise. The low-pass filter should pass the components under 10 Hz unaltered (with a gain of one) and attenuate higher-frequency components (with a gain less than one). A possible transfer function for this filter is

$$T(D) = \frac{1}{\tau} \cdot \frac{1}{D + 1/\tau} \quad (6-6)$$

where $\tau = 2\pi \cdot 10$.

The units of $1/\tau$ are sec⁻¹ which is frequency in units of rad/sec. Taking the Laplace transform of Equation 6-6 simply replaces all occurrences of the D operator with the Laplace s operator, provided all initial conditions are zero. The units of the Laplace operator become $s = j\omega$ = frequency, rad/sec, as desired. At frequencies much less than τ , the $1/\tau \ggg s$ and the transfer function gain is approximately equal to one. At frequencies much greater than τ , the $1/\tau \lll s$ and the transfer function gain approaches $1/\tau s \Rightarrow 0$ as $s \rightarrow \infty$.

The degree of the polynomial used in the denominator of the filter determines how rapidly in frequency the transfer gain approaches zero. Generally, the higher the degree the more rapid the gain is attenuated.

EXAMPLE 6.7 Approximating Time Differentiation Using an Integrator

Often a differentiator is needed. Rather than use a pure differentiator, one with low-pass filtering is often desirable to reduce noise associated with the differentiating operation.

Consider the problem of differentiating a signal $R(t)$ to get $\dot{R}(t)$. Let us denote $\dot{R}(t)$ as the output, $Y(t)$. The transfer function for differentiation is

$$Y(t) = \frac{dR(t)}{dt} = D \cdot R(t) \quad (6-7)$$

The transfer function becomes $\frac{Y(t)}{R(t)} = D$.

Solution

This transfer function is not proper and therefore cannot be solved using integration; however, it can be approximated as

$$\begin{aligned} \frac{Y(t)}{R(t)} &= D \approx \lim_{\varepsilon \rightarrow 0} \frac{D}{\varepsilon D + 1} \\ \frac{Y(t)}{R(t)} &= \frac{1}{\varepsilon} \cdot \frac{D}{D + 1/\varepsilon} \text{ (in moni form)} \end{aligned} \quad (6-8)$$

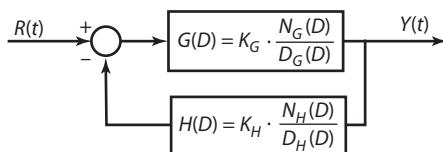
The transfer function contains two terms connected in series and a low-pass filter with transfer function $\frac{1}{\varepsilon} \cdot \frac{1}{D + 1/\varepsilon}$ followed by a differentiation. The value of ε determines how much frequency of the incoming signal can be differentiated.

Smaller values of ε will broaden the frequency range, making the differentiation operation more accurate, but will also make it more susceptible to noise. Larger values of ε do just the opposite. When using this transfer function for differentiation in a simulation, it is common to select $\varepsilon \geq 2 \cdot \Delta T$, where ΔT is the simulation stepsize.

6.3.2 Basic Feedback System and G-Equivalent Form

The *basic feedback system* (BFS) shown in Figure 6-3, is one of the most fundamental forms of block diagrams used for control applications. With some manipulation, any SISO system can be represented in this form.

FIGURE 6-3 BASIC FEEDBACK SYSTEM FORM



The BFS consists of two transfer functions: a forward-loop transfer function, $G(D)$ and a feedback transfer function, $H(D)$. In Figure 6-3, each of these transfer functions are represented in general monic form. $N(D)$ and $D(D)$ are the numerator and denominator polynomials, respectively, and K is the ratio of the gains necessary to create the monic form. The *loop transfer function* (LTF) and the *closed-loop transfer function* (CLTF) of the BFS are two commonly required transfer functions and are computed as

$$\text{LTF: } G(D)H(D) = K_G \cdot K_H \frac{N_G(D) \cdot N_H(D)}{D_G(D) \cdot D_H(D)} \quad (6-9)$$

$$\text{CLTF: } \frac{Y(t)}{R(t)} = \frac{G(D)}{1 + G(D)H(D)} = \frac{\text{forward loop transfer}}{1 + \text{loop transfer}} \quad (6-10)$$

$$\begin{aligned} &= \frac{K_G \frac{N_G(D)}{D_G(D)}}{1 + K_G \cdot K_H \frac{N_G(D) \cdot N_H(D)}{D_G(D) \cdot D_H(D)}} \\ &= \frac{K_G \cdot N_G(D) \cdot D_H(D)}{D_G(D) \cdot D_H(D) + K_G \cdot K_H \cdot N_G(D) \cdot N_H(D)} \end{aligned}$$

It is often necessary to transform a block diagram from the BFS form to a form which has unity feedback. This form is called the G-equivalent form and provides a measure of how close the input signal is to the output signal. This measure is simply the output of the summing junction present in the BFS form. To convert a BFS form to a G-equivalent form, it is necessary to know only the CLTF

of the BFS, which we'll refer to as $T(D)$. The conversion requires solving Equation 6-11 for the G-equivalent transfer function, $G_{\text{eq}}(D)$, in terms of $T(D)$.

$$T(D) = \frac{G_{\text{eq}}(D)}{1 + G_{\text{eq}}(D)} \quad (6-11)$$

The solution (or conversion to G-equivalent form) is presented as

$$G_{\text{eq}}(D) = \frac{T(D)}{1 - T(D)} \quad (6-12)$$

The following examples illustrates the G-equivalent conversion process.

EXAMPLE 6.8 G-Equivalent Conversion Process for a Feedback System

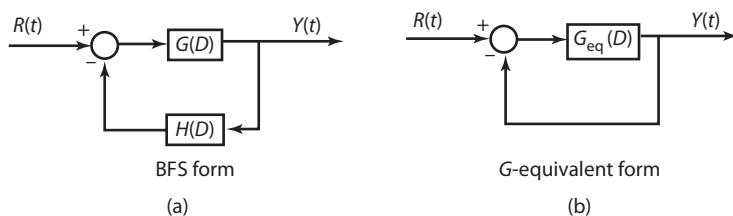
In this example, we will convert a BFS to G-equivalent form. The BFS transfer functions are given as

$$G(D) = \frac{1}{D^2 + 1} \text{ and } H(D) = \frac{D}{D + 1}$$

Solution

The G-equivalent conversion process converts any feedback system into and equivalent feedback system with unity feedback. The forms are shown in Figure 6-4.

FIGURE 6-4 EXAMPLE G-EQUIVALENT CONVERSION



The CLTF is computed in Equation 6-13 for the BFS form.

$$T(D) = \frac{G(D)}{1 + G(D) \cdot H(D)} = \frac{D + 1}{(D^2 + 1) \cdot (D + 1) + D} = \frac{D + 1}{D^3 + D^2 + 2D + 1} \quad (6-13)$$

The G-equivalent transfer function is computed by applying Equation 6-12. The result is

$$G_{\text{eq}}(D) = \frac{T(D)}{1 - T(D)} = \frac{D + 1}{D^3 + D^2 + D} \quad (6-14)$$

The G-equivalent transfer function can and should always be checked to make sure it produces the same CLTF as the BFS system. This check is performed as

$$T(D) = \frac{G_{\text{eq}}(D)}{1 + G_{\text{eq}}(D)} = \frac{D + 1}{D^3 + D^2 + 2D + 1} \quad (6-15)$$

Since the $T(D)$ computed in Equation 6-15 agrees with the original CLTF, we have confidence that our $G_{\text{eq}}(D)$ transfer function is correct.

EXAMPLE 6.9 G-equivalent Conversion Process for a Non-Feedback System

In this example, we will convert a transfer function without explicit feedback to G-equivalent form. The transfer function is given as

$$G(D) = \frac{1}{D^2 + 1}$$

Solution

The CLTF is simply $G(D)$. The G-equivalent transfer function is computed by applying Equation 6-12. The result is

$$G_{\text{eq}}(D) = \frac{G(D)}{1 - G(D)} = \frac{1}{D^2 + 1 - 1} = \frac{1}{D^2} \quad (6-16)$$

The example is completed by checking the closed-loop transfer function for the G-equivalent form and verifying that it agrees with the original CLTF. This check is performed in Equation 6-17.

$$T(D) = \frac{G_{\text{eq}}(D)}{1 + G_{\text{eq}}(D)} = \frac{1}{D^2 + 1} \quad (6-17)$$

As expected, the results agree. The G-equivalent transformation is important because it is used to determine the accuracy metric of a control system. This topic will be discussed in detail later in this chapter.

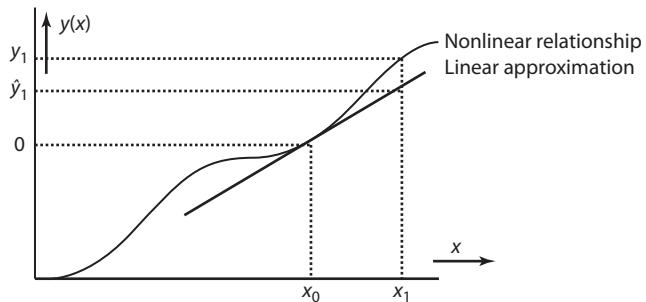
6.4 Linearization of Nonlinear Systems

In order to express a system as a transfer function, the system must be linear. Most systems are nonlinear but can usually be linearized. One commonly employed linearization technique, which can be applied to many nonlinear systems, establishes a linear approximation using the linear terms of a Taylor series expansion, which is computed at a specified operating condition.

To illustrate the technique, consider the nonlinear function of one variable, $y(x)$, shown in Figure 6-5. The purpose of the linearization is to approximate the behavior of the function for small variations in the independent variable, x , near an operating condition or point (x_0, y_0) , where $y_0 \equiv y(x_0)$.

The linear approximation is the line which passes through the operating point and tangent to the nonlinear relationship at that point. To see how well the linearization works, pick an arbitrary point, x_1 (near x_0). The approximate function value at x_1 , designated as \hat{y}_1 , is

$$\hat{y}_1 = y_0 + [\text{slope of tangent at } (x_0, y_0)] \cdot (x_1 - x_0) \quad (6-18)$$

FIGURE 6-5 NONLINEAR FUNCTION OF ONE VARIABLE AND ITS LINEAR APPROXIMATION

The slope of the tangent at (x_0, y_0) is the partial of the nonlinear function taken with respect to the independent variable, x , evaluated at the operating condition. This partial is defined as

$$[\text{Slope of tangent at } (x_0, y_0)] \equiv \left. \frac{\partial y}{\partial x} \right|_{\substack{x=x_0 \\ y=y_0}} \quad (6-19)$$

After substitution, Equation 6-20 reveals the general form of the linearization.

$$\hat{y} \approx y_0 + \left. \frac{\partial y}{\partial x} \right|_{\substack{x=x_0 \\ y=y_0}} \cdot (x - x_0) \quad (6-20)$$

Clearly, if x is chosen too *far* from x_0 in Equation 6-20, the linear relationship may not hold very well, thus rendering a large error between the actual nonlinear function value, $y(x)$, and the linearized approximation value, $\hat{y}(x)$.

Equation 6-20 is a linear Taylor series expansion of the function $y(x)$. It has *one degree of freedom* because y is a function of *one* variable, and it is a linear series because all second and higher partial terms have been omitted.

To compute a linearized approximation for a system, an operating condition and the partials of the output at the operating condition are needed. To illustrate, consider a function, z , of two variables, x and y ; $z = z(x, y)$. The linearized approximation of the system at the operating condition, (x_0, y_0) , is

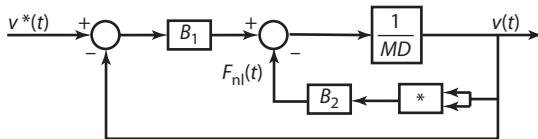
$$z = z_0 + \left. \frac{\partial z}{\partial x} \right|_{(x_0, y_0)} \cdot (x - x_0) + \left. \frac{\partial z}{\partial y} \right|_{(x_0, y_0)} \cdot (y - y_0) \quad (6-21)$$

Here $z_0 = z(x_0, y_0)$ and both partials are evaluated at the operating condition, $\hat{y}(x)$. Equation 6-21 is a linear Taylor series expansion of the function $z(x, y)$ with *two degrees of freedom*.

The linearization process is applied to a nonlinear block diagram model in the next example. The resulting linearized block diagram could be reduced to a transfer function for analysis and would have satisfactory performance around the point of linearization.

EXAMPLE 6.10 Linearization of a Nonlinear Function in a Block Diagram

The block diagram representation of a mechanical system consisting of a mass and friction is presented in Figure 6-6, where

FIGURE 6-6 EXAMPLE 6.10—NONLINEAR BLOCK DIAGRAM


$v^*(t)$ = the reference speed

$v(t)$ = the speed of the mass

M = the mass

B_1 = linear viscous friction coefficient

B_2 = nonlinear viscous friction coefficient

The object of this example is to linearize the nonlinear friction term, F_{nl} . We will perform the linearization at the operation condition $v(t) = v_0 = 50$. At this condition, the nominal value of $F_{nl}(t) = F_{nl0}$ is computed as

$$F_{nl0} = B_2 \cdot v_0^2 = 2500B_2$$

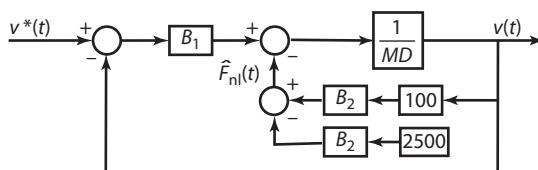
An actual value for B_2 would normally be provided, but we will keep it general in this example. Next, the partial of F_{nl} with respect to variations in $v(t)$ is computed and evaluated at the operation condition.

$$\frac{\partial F_{nl}(t)}{\partial v(t)} = 2 \cdot B_2 \cdot v(t) \Big|_{v(t)=v_0=50} = 100 \cdot B_2$$

The final linearization becomes

$$\hat{F}_{nl}(t) = 2500 \cdot B_2 + (100 \cdot B_2) \cdot (v(t) - 50) = -2500 \cdot B_2 + 100 \cdot B_2 \cdot v(t)$$

Here we have *hatted* the $F_{nl}(t)$ signal to distinguish it from the true nonlinear signal. The linearized block diagram is shown in Figure 6-7.

FIGURE 6-7 EXAMPLE 6.10—LINEARIZED BLOCK DIAGRAM


Our linearized block diagram will behave very similarly to the nonlinear block diagram—provided the variations in $v(t)$ are small. That is, $v(t)$ must be near $v_0 = 50$, which is the operating condition.

Frequently, the nonlinear function is not conveniently available—such is the case with complex nonlinear systems, either physical or simulated. In these situations, the partials must be approximated by applying external *probing* signals to the system. For example, reconsider the $y(x)$ system presented in Figure 6-8 operating at the point (x_0, y_0) where $y_0 \equiv y(x_0)$. The partial needed in Equation 6-20 could be computed by perturbing the independent variable, x , by a small value or perturbation, $\Delta x \equiv x - x_0$.

$$\frac{y(x) - y_0}{x - x_0} = \frac{\Delta y}{\Delta x} \approx \left. \frac{\partial y}{\partial x} \right|_{x=x_0, y=y_0} \quad (6-22)$$

The resulting perturbation in the output, $\Delta y \equiv y(x) - y_0$, divided by the input perturbation produces a linear approximation for the partial in Equation 6-22.

6.5 Time Delays

Time delays are encountered so frequently in systems that they deserve special attention. Using Table 6-1, the Laplace transform of a time delay of T seconds is $e^{-T \cdot s}$. This is a nonrational function and, as such, has no exact transfer function equivalent; however, one can be approximated using an approximation called a *Pade approximation*.

A Pade approximation allows an infinite series to be approximated as a ratio of two polynomials. The approximation has an interesting property called *telescoping*, allowing X terms of the infinite series to be represented by two Pade polynomials each of order $X/2$.

For example, a time delay of T seconds, represented by the Laplace term $e^{-T \cdot s}$, can be expressed by the following infinite exponential series:

$$e^{-T \cdot s} = 1 - Ts + \frac{(Ts)^2}{2} - \frac{(Ts)^3}{3!} + \frac{(Ts)^4}{4!} + \dots$$

Our objective is to write this as an approximate transfer function so it can be used for analysis involving poles and zeros. We'll assume the transfer function we're going to use is first order of the form

$$T(s) = \frac{1 + Bs}{1 + As}$$

Here A and B are unknown and will be determined to best match the exponential series. Proceeding, we first long divide $T(s)$ to create another series in the s operator.

$$\begin{array}{r} 1 + (B - A)s + A(B - A)s^2 \\ 1 + As \overline{) 1 + Bs + 0s^2 + 0s^3 + \dots} \\ 1 + As \\ (B - A)s + 0s^2 \end{array}$$

$$\begin{aligned}
 & (B - A)s + A(B - A)s^2 \\
 & + A(B - A)s^2 + 0s^2 \\
 & + A(B - A)s^2 + A(B - A)s^3
 \end{aligned}$$

Next, we equate terms in like powers of s between the exponential expansion and the long division result to obtain the following two equations.

$$\begin{aligned}
 s^1 \text{ term: } -T &= B - A \\
 s^2 \text{ term: } \frac{T^2}{2} &= A(A - B)
 \end{aligned}$$

Solving for A and B , we obtain

$$\begin{aligned}
 A &= \frac{T}{2} \\
 B &= -\frac{T}{2}
 \end{aligned}$$

The resulting approximate transfer function becomes

$$e^{-T \cdot s} \approx \frac{1 - \frac{sT}{2}}{1 + \frac{sT}{2}}$$

This approximation is called a Pade approximation. Pade approximations can be of any order, we have developed the approximation for the first-order transfer function case here. The first-order approximation is used extensively in control system analysis to represent time delays.

Table 6-5 summarizes Pade approximation for second- and third-order approximate transfer functions. These transfer functions were derived in the same manner as the first-order approximation.

TABLE 6-5 PADE APPROXIMATIONS OF ORDERS 1, 2, AND 3 FOR A PURE TIME DELAY, $e^{-T \cdot s}$

Order	Pade Approximation
1	$\frac{1 - (s \cdot T)/2}{1 + (s \cdot T)/2}$
2	$\frac{1 - (s \cdot T)/2 + (s \cdot T)^2/12}{1 + (s \cdot T)/2 + (s \cdot T)^2/12}$
3	$\frac{1 - (s \cdot T)/2 + (s \cdot T)^2/10 - (s \cdot T)^3/120}{1 + (s \cdot T)/2 + (s \cdot T)^2/10 + (s \cdot T)^3/120}$

The following example illustrates how the Pade approximation is used to convert a nonrational time delay present in a system transfer function into a rational transfer function approximation.

EXAMPLE 6.11 Heat Exchanger Time Delay Transfer Function

The transfer function (in the s domain) for a heat exchanger is given as

$$G(s) = \frac{.001 \cdot e^{-10s}}{(s + .1)(s + .01)} \quad (6-23)$$

A transfer function suitable for analysis can be derived by replacing the 10-second time delay with a first-order Pade approximation. Using the appropriate entry from Table 6-2, the time delay approximation is

$$e^{-10s} \approx \frac{1 - 10 \cdot s/2}{1 + 10 \cdot s/2} = \frac{s - .2}{s + .2} \quad (6-24)$$

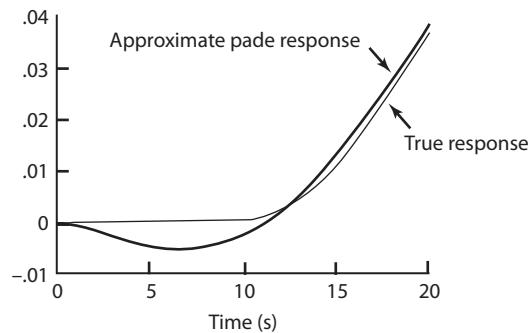
Substituting the approximation back into the original transfer function produces the desired rational approximation

$$\hat{G}(s) \approx \frac{.001 \cdot (s - .2)}{(s + .1)(s + .01)(s + .2)} \quad (6-25)$$

Solution

To illustrate the accuracy of the Pade approximation, the unit step response of the actual and approximate transfer functions, Equations 6-23 and 6-25, is computed and summarized in Figure 6-8.

FIGURE 6-8 PERFORMANCE OF A TRUE TIME DELAY AND ITS FIRST-ORDER PADE APPROXIMATION



Note the characteristic “wrong initial direction” of the first-order Pade approximation response. This behavior is part of the price we pay for the approximation; it cannot be eliminated, but it can be reduced by using a higher-order Pade approximation. This, however, is usually not necessary when the Pade approximation is used for control-system design purposes. A first-order Pade approximation normally supplies enough information about the time delay for control-system design purposes.

6.6 Measures of System Performance

System performance is based on four metrics: stability, accuracy, transient response, and sensitivity. Most systems will lack one or more of these measures. In these situations, the system must be compensated. A description of each is presented in the next few sections.

6.6.1 Stability

A stable system is one which produces a bounded, or finite, response when subjected to a bounded input. The conditions for stability are established by inspecting the general form of a system response, calculated using the Laplace Transform, and expressed in Equation 6-26.

$$y(t) = A \cdot e^{a \cdot t} + B \cdot e^{b \cdot t} + C \cdot e^{c \cdot t} + \dots \quad (6-26)$$

where

a, b, c = poles of the system transfer function (roots of its denominator)

A, B, C = residuals which are a function of the zeros of the transfer function

Equation 6-26 relates a system's output to its poles and zeros. The stability of a system depends entirely on its pole locations. The conditions for stability are summarized as

A system is stable if the real part of all poles are < 0 .

A system is marginally stable if the real part of all poles are ≤ 0 .

A system is unstable if the real part of any pole is positive.

The poles of a transfer function are the roots of the characteristic equation. There are analytical methods for computing the pole locations, but today it is far easier to employ a computer-based factoring program.

6.6.2 Accuracy

Accuracy (or steady-state tracking error) is the error between input and output signals in the steady state for a system which is in G-equivalent form. In this form, the input and output signals are compared directly (apples and apples) at the summing junction (see Figure 6-6) and because of this, the input signal can be viewed as a desired output signal, suggesting it is how we want the actual output to behave. The difference between the two is the *steady-state error*.

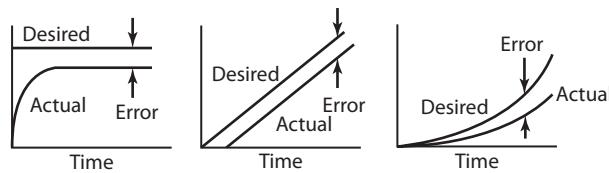
Three classes of desired output signals are used to determine a system's accuracy:

- Step
- Ramp
- Parabola

Figure 6-9 illustrates how the steady-state error is computed for each of the three signal classes.

Once the desired output signal has been defined by a step, ramp, parabola, or a combination of the three, the accuracy or steady-state error of the system can be computed. Three steps are necessary and summarized here.

Step 1. Transform the system into a G-equivalent form in the s domain.

FIGURE 6-9 THREE MEASURES OF ACCURACY

Step 2. Compute the error coefficient, K_p , K_v , or K_a , for the desired output-signal class:

$$\text{Unit step error coefficient: } K_p \equiv \lim_{s \rightarrow 0} \{G(s)\}$$

$$\text{Unit ramp error coefficient: } K_v \equiv \lim_{s \rightarrow 0} \{s \cdot G(s)\}$$

$$\text{Unit parabolic error coefficient: } K_a \equiv \lim_{s \rightarrow 0} \{s^2 \cdot G(s)\}$$

Step 3. Compute the steady-state error as a function of the error coefficient:

$$\text{Unit step error: } e_{ss}(\text{step}) = \frac{1}{1 + K_p}$$

$$\text{Unit ramp error: } e_{ss}(\text{ramp}) = \frac{1}{K_v}$$

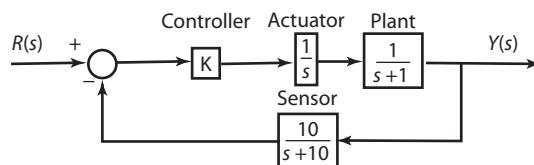
$$\text{Unit parabolic error: } e_{ss}(\text{para}) = \frac{1}{K_a}$$

In this process, all signals are of unit value—meaning the step has an amplitude of one, the ramp has a slope of one, and the parabola has a curvature of one. When the actual input values differ, the scaling must be carried through the calculation to compute the steady-state errors.

An example is presented to illustrate the accuracy procedure.

EXAMPLE 6.12 Accuracy Calculation for an Electromechanical System

To illustrate the calculation of accuracy in terms of steady-state error, consider the electromechanical system in the s domain shown in Figure 6-10.

FIGURE 6-10 SYSTEM USED TO COMPUTE ACCURACY FOR EXAMPLE 6.12

We can compute how closely this system will track step, ramp, and parabolic inputs by applying the previous procedure.

Solution

First the system must be converted to G-equivalent form. The closed-loop transfer function and G-equivalent forms are presented in Equations 6- 27 and 6-28.

$$T(s) = \frac{K \cdot (s + 10)}{s^3 + 11s^2 + 10s + 10K} \quad (6-27)$$

and

$$G_{\text{eq}}(s) = \frac{K \cdot (s + 10)}{s^3 + 11s^2 + (10 - K)s} \quad (6-28)$$

The error coefficients and resulting steady-state errors are computed using the G-equivalent transfer function and summarized in Table 6-6.

TABLE 6-6 STEADY-STATE ERRORS FOR EXAMPLE 6.12 SYSTEM

Desired Output	Error Coefficient	Steady-State Error
Step	$K_p = \lim_{s \rightarrow 0} \{G_{\text{eq}}(s)\} = \infty$	$e_{ss}(\text{step}) = \frac{1}{1 + K_p} = 0$
Ramp	$K_v = \lim_{s \rightarrow 0} \{s \cdot G_{\text{eq}}(s)\} = \frac{10K}{10 - K}$	$e_{ss}(\text{ramp}) = \frac{1}{K_v} = \frac{10 - K}{10K}$
Parabola	$K_a = \lim_{s \rightarrow 0} \{s^2 \cdot G_{\text{eq}}(s)\} = 0$	$e_{ss}(\text{para}) = \frac{1}{K_a} = \infty$

It is clear from Table 6-6 by selecting $K = 10$ the system is capable of tracking both step and ramp input signals with zero steady-state error. Regardless of the choice of K , the system will never track a parabolic input with anything but a constantly increasing error, eventually going to infinity.

EXAMPLE 6.13 Accuracy Design for an Electromechanical System

To further illustrate the concept of system accuracy, reconsider this example with a design value of $K = 5$. The signal we wish the system to track is a combination of the step and ramp signals given as

$$r(t) = 1 + 5 \cdot t$$

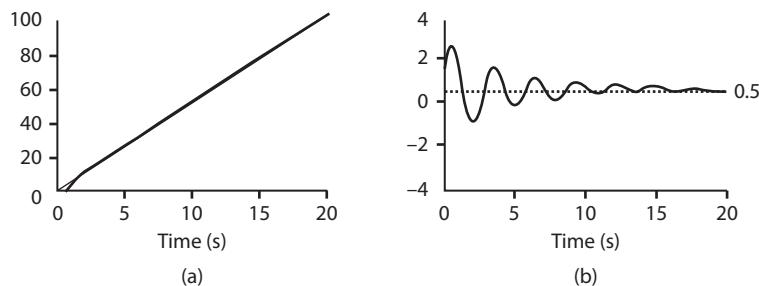
Solution

From the past example, the response of the system due to the step part of $r(t)$ will have zero error; however, the ramp part will not. Using Table 6-3 and the design value of $K = 5$, the unit ramp error is 0.1. Our input signal, $r(t)$, does not contain a unit slope ramp but rather a ramp with slope equal to 5. This is five times the unit slope, so we would expect a factor of five in the error (that is, an error of .5 should result).

The simulated response of the system shown in Figure 6-10 (subject to this input) is presented in Figure 6-11.

As expected, the response of the system due to the step part of the input has zero error, and the response due to the ramp part has an error of .5. This is the total error and agrees with the predicted results.

FIGURE 6-11 ACCURACY OF SYSTEM WITH INPUT $r(t) = +5t$ (A) DESIRED AND ACTUAL RESPONSES (B) ERROR BETWEEN THE TWO



When investigating the accuracy metric of a system represented in G-equivalent form, the notion of system type is useful. Consider a transfer function, $G(s)$, consisting of a numerator polynomial, $N(s)$, and a denominator polynomial, $D(s)$. After making any cancellations that may exist, the system type is defined as follows.

System type is the number of free s terms in $D(s)$. A free s term appears as an $(s + 0)$ factor.

Several examples are presented to familiarize the reader with system type.

EXAMPLE 6-14 System Type of Several Transfer Functions

The system type is computed for several $G_{eq}(s)$ transfer functions in Table 6-7. In the G-equivalent form, each system has unity feedback.

TABLE 6-7

$G_{eq}(s)$	System Type	Explanation
$G_{eq}(s) = \frac{(s + 2)}{s^2 + 1}$	0	No free s terms in the denominator of $G_{eq}(s)$.
$G_{eq}(s) = \frac{(s + 2)}{s^2}$	2	Two free s terms in the denominator of $G_{eq}(s)$.
$G_{eq}(s) = \frac{s(s + 2)}{s^2}$	1	After a pole-zero cancellation there is one free s term.

Solution

Knowing the system type tells us a lot about the system accuracy (that is, how well the system can track various classes of input signal). The relationship between system type, error coefficient, and steady-state error is presented in Table 6-8.

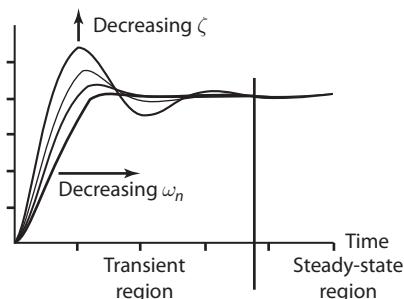
TABLE 6-8 RELATIONSHIP BETWEEN SYSTEM TYPE AND STEADY-STATE ERROR

System Type	Error Coefficient	Steady State Error
0	$K_p = \text{finite}$ $K_v = 0$ $K_a = 0$	$e_{ss}(\text{step}) = \text{finite}$ $e_{ss}(\text{ramp}) = \infty$ $e_{ss}(\text{parabola}) = \infty$
1	$K_p = \infty$ $K_v = \text{finite}$ $K_a = 0$	$e_{ss}(\text{step}) = 0$ $e_{ss}(\text{ramp}) = \text{finite}$ $e_{ss}(\text{parabola}) = \infty$
2	$K_p = \infty$ $K_v = \infty$ $K_a = \text{finite}$	$e_{ss}(\text{step}) = 0$ $e_{ss}(\text{ramp}) = 0$ $e_{ss}(\text{parabola}) = \text{finite}$

Knowing the system type, Table 6-8 allows you to predict what classes of signals your system can track. For example, if you had a type 1 system and you needed zero steady-state ramp, tracking the table indicates you would need to increase the system type to 2.

6.6.3 Transient Response

Transient response is the shape of a signal as it moves between two steady-state points. It is quantified in terms of two parameters: the damping ratio, ζ , pronounced zeta, and the natural undamped frequency, ω_n . The effect of these two parameters on the transient response of a signal is presented in Figure 6-12.

FIGURE 6-12 SIGNAL TRANSIENT-RESPONSE REGION AND CHARACTERISTIC PARAMETERS

The signal in Figure 6-12 begins with a value of 0 and ends with a value of 1. These values are arbitrary. What is important is that they both represent steady-state points, which are the same points introduced in the previous section on accuracy.

The time constant of a signal is another measure of its performance and is computed knowing the damping ratio and natural, undamped frequency as

$$\tau \equiv \frac{1}{\zeta \cdot \omega_n}$$

Since damping ratio is (as implied) the ratio of two values, it has no units. The damping ratio is analogous to normalized friction in mechanical systems and normally ranges between 0 and 1.

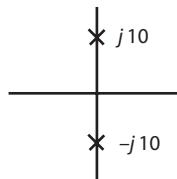
Negative damping is possible, but this implies an unstable system and is not normally used. Because the natural undamped frequency has units of rad/sec, the units for the time constant are seconds.

Since system poles and zeros are generally complex in nature, they can be displayed in a Cartesian coordinate system in the complex s -plane or Laplace domain where $s = \sigma + j\omega$. In the s -plane, the x axis represents the real component, σ , and the y axis represents the imaginary component, $j\omega$. A plot of the system poles and zeros in the s -plane is called a *pole-zero* (PZ) plot. To distinguish between poles and zeros in a PZ plot, the poles are plotted as x 's and the zeros as o 's. An example is presented to illustrate the procedure for constructing a PZ plot.

EXAMPLE 6.15 PZ Plot for a Sinusoid

Consider the signal $y(t) = \sin 10t$. This time-domain signal can be converted to its s -domain equivalent by using Table 6-1 to take its Laplace transform. The resulting s -domain signal becomes $Y(s) = \frac{10}{s^2 + 100}$. In the s -domain, the signal has two poles: $s_{1,2} = \pm j10$ and no zeros. A PZ plot for $Y(s)$ is shown in Figure 6-13.

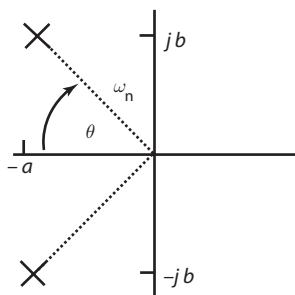
FIGURE 6-13 PZ PLOT FOR $Y(t) = \sin 10t$



Solution

The PZ plot is important because from it we can extract the transient metric information (ζ, ω_n, τ) for the signal or system. The process is simply a rectangular to polar conversion and is illustrated by considering the PZ plot for a second-order signal or system with complex poles. Such a plot is presented in Figure 6-14.

FIGURE 6-14 COMPLEX TWO-POLE SYSTEM OR SIGNAL PZ PLOT



The poles for this system are located at $s_1 = a + jb$ and $s_2 = -a - jb$. Since complex roots occur so often in systems, we will adopt the shorthand notation of $s_{1,2} = -a \pm jb$ to express them as one term.

The relationship between the s -plane real and imaginary values and the polar values (ζ and ω_n) is summarized as

$$\omega_n = \sqrt{a^2 + b^2}$$

$$\zeta = \cos^{-1}\theta$$

where θ is the angle of the pole measured from 180° and counterclockwise is positive.

Table 6-9 presents the relationship between signal or system pole locations in the s -plane as displayed on a PZ plot and the transient response of the signal or system in the time domain. With no loss in generality, all calculations are based on a two-pole transfer function.

TABLE 6-9 EFFECTS OF POLE LOCATION ON TRANSIENT RESPONSE

Pole Location in s -plane	Time Response	Comments
	A plot of a signal over time showing exponential growth towards infinity. The x-axis is labeled "Time".	$\zeta < 0$, unstable
	A plot of a signal over time showing sinusoidal oscillations with constant amplitude. The x-axis is labeled "Time".	$\zeta = 0$, marginally stable
	A plot of a signal over time showing sinusoidal oscillations that dampen over time towards zero. The x-axis is labeled "Time".	$0 < \zeta < 1$, stable, underdamped
	A plot of a signal over time showing an exponential decay towards zero. The x-axis is labeled "Time".	$\zeta = 1$, stable, critically damped, nondistinct poles
	A plot of a signal over time showing an exponential decay towards zero. The decay is slower than the critically damped case. The x-axis is labeled "Time".	$\zeta > 1$, stable, overdamped, distinct poles

Although unstable systems are difficult to control, they are not bad. Intelligent use of instabilities can reduce actuator size and power requirements dramatically; however, precise control must be employed to keep the response in a range which can be stabilized.

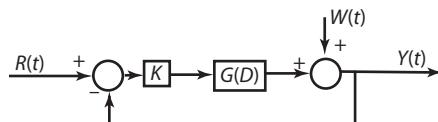
6.6.4 Sensitivity

Sensitivity is the measure by which controlled signals are influenced by disturbances which include parameter variations within the plant and external signals such as noise. This topic was discussed in Chapter 1 as being one of the important features of feedback control.

A well designed control system has low sensitivity to disturbances. If we restrict external signals to sensor noise, low sensitivity can be achieved with a high LTF gain.

Consider the BFS shown in Figure 6-15 with plant $G(D)$, controller K , and sensor noise $W(t)$. Sensor noise can be caused by electrical interference or by the operation of the sensor itself. It always results in unpredictable random errors in the measurand. In general, the controller could be any transfer function, but for the purpose of illustrating sensitivity, it is simpler to consider it as a pure gain.

FIGURE 6-15 PLANT VARIATIONS AND SENSOR NOISE DISTURBANCE INPUTS TO THE BFS



Imagine that parameters within $G(D)$ vary randomly in such a way that the plant may be represented as a nominal transfer function, $G^\circ(D)$, plus a random variation, $\Delta G(D)$. The new varying plant model becomes $G(D) = G^\circ(D) + \Delta G(D)$. The sensor noise signal is also considered a random variation and applied to the controlled variables additively. Other configurations are possible, but this is suitable for our illustrative purposes.

The transfer functions from command input, $R(t)$, to output, $Y(t)$, and from disturbance input, $W(t)$ to $Y(t)$, are presented in Equations 6-29 and 6-30, respectively.

$$\frac{Y(t)}{R(t)} = \frac{K \cdot (G^\circ(D) + \Delta G(D))}{1 + K \cdot (G^\circ(D) + \Delta G(D))} \quad (6-29)$$

$$\frac{Y(t)}{W(t)} = \frac{1}{1 + K \cdot (G^\circ(D) + \Delta G(D))} \quad (6-30)$$

Since the control, K , is something that we design and select large so the loop gain $K \cdot (G^\circ(D) + \Delta G(D)) \gg 1$ results in Equation 6-29 approaching one and Equation 6-30 approaching zero. The large loop gain provides exactly what is needed, the part of the response affected by the command input, $R(t)$, approaches a gain of one, and the part due to the sensor disturbance input, $W(t)$, approaches a gain of zero.

In summary, for low sensitivity to plant parameter variations as well as sensor noise, the loop gain should be selected as large as possible.

This section has discussed the four metrics used to judge system performance, stability, accuracy, transient response, and sensitivity. When confronted with a control-system design task, the following design sequence will prove helpful.

Step 1. Assess the accuracy requirements in terms of the system type. This step will help define the type of control structure needed. For example, you can determine whether the controller must be a gain or PI.

Step 2. Map the transient and stability specifications, usually a damping and a time constant or natural undamped frequency, onto the s -plane. The specification region is usually a cone-shaped region. This will be illustrated in the root locus design examples presented later in this chapter.

Step 3. Design the controller such that it drags the plant poles into the design cone mentioned in step 2. This step includes selecting a controller structure as well as determining the parameters in that structure.

Step 4. Assess the design for sensitivity via simulation. Since the control design is performed on a linearized version of the plant, this step allows you to tune the controller to operate correctly on the nonlinear plant.

The following sections introduce two methods for analyzing the plant and designing the controller: the root locus method and the Bode plot method.

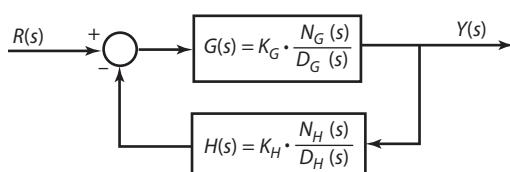
6.7 Root Locus

The root locus is a plot of the closed-loop poles of a transfer function as one gain in the transfer function is varied. Normally, this gain is a control gain, but it also could be a parameter variation in the plant. The root locus provides stability, accuracy, sensitivity, and transient information and is useful for system analysis and design. This section describes the fundamentals of the root locus, sketching rules, and its interpretation. Two examples regarding its use in determining system stability and sensitivity are also presented.

6.7.1 Fundamentals

Our discussion of root locus is based the basic feedback system (BFS) which is represented in Figure 6-16.

FIGURE 6-16 BASIC FEEDBACK SYSTEM BLOCK DIAGRAM



The BFS is represented in the Laplace domain; however, it could also be represented in the time domain by redefining s as the derivative operator, D .

The BFS consists of two monic transfer functions; a forward-loop transfer function, $G(s)$; and a feedback transfer function, $H(s)$. $N_G(s)$, $N_H(s)$, $D_G(s)$, and $D_H(s)$ are the numerator and denominator

polynomials, respectively. K_G and K_H are the ratio of the gains necessary to create the monic form. The loop transfer function (*LTF*) and the closed-loop transfer function (*CLTF*) of the *BFS* are presented in Equations 6-31 and 6-32.

$$\text{LTF: } G(s)H(s) = K_G \cdot K_H \frac{N_G(s) \cdot N_H(s)}{D_G(s) \cdot D_H(s)} \quad (6-31)$$

$$\begin{aligned} \text{CLTF: } \frac{Y(s)}{R(s)} &= \frac{G(s)}{1 + G(s)H(s)} = \frac{K_G \frac{N_G(s)}{D_G(s)}}{1 + K_G \cdot K_H \frac{N_G(s) \cdot N_H(s)}{D_G(s) \cdot D_H(s)}} \\ &= \frac{K_G \cdot N_G(s) \cdot D_H(s)}{D_G(s) \cdot D_H(s) + K_G \cdot K_H \cdot N_G(s) \cdot N_H(s)} \end{aligned} \quad (6-32)$$

The poles and zeros of the CLTF, as well as those of the LTF, are four key system performance indicators. Provided $N_G(s)$, $D_G(s)$, $N_H(s)$, and $D_H(s)$ are in factored form. Three of the indicators are available by inspection.

LTF zeros = zeros of $G(s)H(s)$ = roots of $N_G(s) \cdot N_H(s)$

LTF poles = poles of $G(s)H(s)$ = roots of $D_G(s) \cdot D_H(s)$

CLTF zeros = zeros of $G(s)$ and poles of $H(s)$ = roots of $N_G \cdot D_H(s)$

The remaining indicator, the CLTF poles, which cannot be calculated by inspection, are the roots of the characteristic equation

$$\rho(s) = D_G(s) \cdot D_H(s) + K \cdot N_G(s) \cdot N_H(s) = 0$$

Here the gain K , called the *loop sensitivity gain*, is defined as $K \equiv K_G \cdot K_H$.

The CLTF pole calculation is not trivial, it requires factoring the sum of two polynomials; however, the behavior at extreme values of K can be easily ascertained without factoring.

At one extreme, for small values of K , the characteristic equation becomes; $\rho(s) \approx D_G(s) \cdot D_H(s) = 0$ with roots at the *poles* of $G(s)$ and $H(s)$.

At the other extreme, for large values of K , the characteristic equation becomes; $\rho(s) \approx K \cdot N_G(s) \cdot N_H(s) = 0$ with roots at the *zeros* of $G(s)$ and $H(s)$.

Notice that at the high extreme, K can be large negatively or positively with no effect on the roots. The behavior of the *CLTF* poles between the two extremes is still not known at this point. Enter the root locus, whose sole purpose is to display the trajectories of these roots, as K is varied from one extreme to the other.

The root locus is a plot of the *CLTF* poles as a function of K in the s -plane. The root locus method of analysis is typically applied to two types of problems; compensator design problems where K is varied over one range, either $\infty > K > 0$ or $-\infty < K < 0$ or sensitivity analysis problems where K is varied over both ranges, $\infty > K > 0$ and $-\infty < K < 0$.

The root locus is based on two conditions (1) *magnitude condition* and (2) *phase condition*. These conditions are obtained from the characteristic equation. The characteristic equation of the *BFS* is the denominator polynomial set to zero; $\rho(K, s) = 1 + G(K, s) \cdot H(K, s) = 0$. Subtracting 1 from both sides results in

$$G(K, s) \cdot H(K, s) = -1 \quad (6-33)$$

The loop sensitivity gain, K , has been included as an argument in Equation 6-33, as it may represent either a gain or a parameter in either (but not both) of the transfer functions $G(s)$ or $H(s)$.

Equation 6-33 must be satisfied for all values of K . Since $G(s)$ and $H(s)$ are transfer functions which are linear, K enters them linearly and can therefore be factored out as

$$-\frac{1}{K} = F(G(s) \cdot H(s)) \equiv F(s) \quad (6-34)$$

Equation 6-34 is called the *root locus equation*.

The right side of the equation, $F(s)$, is a complex term ($s = \sigma + j\omega$) and can be represented as having a magnitude and phase as

$$F(s) = |F(s)| \cdot e^{j(\tan^{-1}(F(s)))}$$

The left side of Equation 6-34 consists of the K term, which is a real number; however, it can be positive or negative. Since both sides *must be equal*, we represent the left side in terms of its magnitude and phase and then equate the terms with those on the right side. At a given K value, the left side is represented by

$$-\frac{1}{K} = \begin{cases} \left| \frac{1}{K} \right| \cdot e^{j(2h+1)\pi} & \text{if } K > 0 \\ \left| \frac{1}{K} \right| \cdot e^{j(2h)\pi} & \text{if } K < 0 \end{cases} \quad (6-35)$$

In Equation 6-35, h is any integer sequence; however, it is usually defined to start at 0 and increase ($h = 0, 1, 2, \dots$). Equating the magnitude and phases of the left and right sides of Equations 6-34 and 6-35 results in the magnitude and phase conditions for the root locus.

$$\text{Magnitude condition: } \left| \frac{1}{K} \right| = |F(s)| \quad (6-36)$$

$$\begin{aligned} \text{Phase condition: } (2h+1)\pi &= \tan^{-1}(F(s)) \text{ for } K > 0 \\ (2h)\pi &= \tan^{-1}(F(s)) \text{ for } K < 0 \end{aligned} \quad (6-37)$$

where $h = 0, 1, 2, 3, \dots$ until repeat occurs

6.7.2 Sketching Rules

When a hand sketch of a root locus is required (usually for compensator design), a short procedure consisting of six steps (Table 6-10) may be applied. The root locus sketching procedure assumes the following are given: $G(K,s)$, $H(K,s)$, and a K range (either positive, $\infty > K > 0$, or negative, $-\infty < K < 0$).

Two extremely common pole-zero configurations deserve specific attention. In the first configuration, one real zero is present to the left of two real poles. As the gain is increased positively ($K > 0$), the two poles come together on the real axis and are then drawn leftward *around the zero*

TABLE 6-10 ROOT LOCUS SKETCHING STEPS

Step	Description
1	Solve $G(K,s) \cdot H(K,s) = -1$ for $-\frac{1}{K} = F(s)$
2	On the s -plane, plot the poles (x) and zeros (o) of $F(s)$.
3	Compute the pole-zero excess (PZE) defined as $PZE \equiv \# \text{Poles in } F(s) - \# \text{Zeros in } F(s)$
4	If $(PZE < 2)$ omit step 4 and go to step 5. Plot the straight line asymptotes on the s -plane plot. The asymptotes are identified by their angle with respect to the real axis, θ , and by their intersection with the real axis, σ_c , called the centroid. θ is defined as 0° along the real axis going in the positive direction (rightward) and takes on increasing positive value for counterclockwise rotation. The asymptote angle and centroid are computed, depending on the K range, as $\sigma_c = \frac{\sum \text{Poles} - \sum \text{Zeros}}{PZE}$ if $K > 0$; $\theta = \frac{(2h + 1) \cdot 180}{PZE}$; $h = 0, 1, 2, 3, \dots$ until θ begins to repeat if $K < 0$; $\theta = \frac{(2h) \cdot 180}{PZE}$; $h = 0, 1, 2, 3, \dots$ until θ begins to repeat
5	For a $K > 0$ range, a loci may lie on the real axis if the total number of poles plus zeros to the right of the section is odd (0 is even). For a $K < 0$ range, a loci may lie on the real axis if the total number of poles plus zeros to the right of the section is even.
6	Sketch the root locus from the LTF poles (here $K = 0$) to the LTF zeros (where $K = \pm\infty$). Keep in mind that the root locus MUST be symmetric with respect to the real axis, because every complex root has a conjugate.

through the complex plane in a circular path centered at the zero. The poles continue their leftward circular motion until they intersect the real axis to the left of the zero. Once on the real axis, one pole moves leftward towards $-\infty$, and the other moves rightward towards the zero. In the second configuration, two real zeros are to the left of two real poles. As the gain is increased positively ($K > 0$), the two poles come together on the real axis and are then drawn leftward in a circular path which intersects the real axis between the two zeros. One pole then goes to one zero and the other to the remaining zero.

6.7.3 Sketching Examples

In this section, the sketching procedure is applied to several examples to sketch the root locus by hand and extract information from it for the eventual use in the design of the control system.

EXAMPLE 6.16 Calculation of the Gain Range for Stability

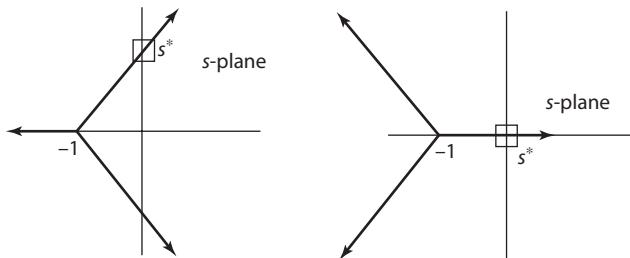
Consider a BFS system with forward and feedback transfer functions given as

$$G(s) = \frac{K}{(s + 1)^3} \quad \text{and} \quad H(s) = 1$$

It is desired to compute the range of K values for which this system is stable. Following the Table 6-10 sketching rules, two root locus plots are drawn—one for the $K > 0$ range and the other for the $K < 0$ range. The root locus sketches are shown in Figure 6-17.

Solution

The point s^* is the point where the root locus intersects the right-half plane (RHP), which is the unstable region of the s -plane. We can compute the gain at the s^* point from Equation 6-36 as

FIGURE 6-17 **$K > 0$ ROOT LOCUS (LEFT), $K < 0$ ROOT LOCUS (RIGHT)**

For $K > 0$ range: the gain at $s^* = j1.86$ is

$$\left| \frac{1}{K} \right| = |F(s^*)| = \frac{1}{|j1.86 + 1|^3} = \frac{1}{9.4} \Rightarrow K = 9.4$$

For $K < 0$ range: the gain at $s^* = 0$ is

$$\left| \frac{1}{K} \right| = |F(s^*)| = \frac{1}{|0 + 1|^3} = \frac{1}{1} \Rightarrow K = -1$$

The gain range for stability becomes

$$9.4 > K > -1$$

EXAMPLE 6.17 Stability Sensitivity of a System Due to Parameter Variation

Consider a BFS system with forward and feedback transfer functions given as

$$G(s) = \frac{(s + 2)}{(s + 3) \cdot (s + P)} \text{ and } H(s) = 1$$

where P is a parameter known to vary randomly around a mean value of 5. This variation can be represented as an additive perturbation to the mean as

$$P = P^o + \Delta P \text{ with } P^o = 5 \text{ and } K = \Delta P$$

Solution

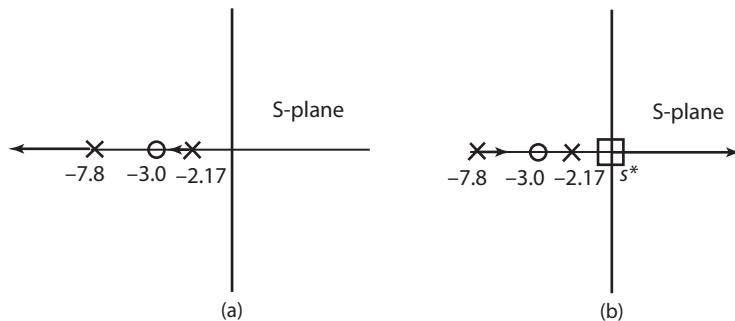
Substitution into $G(s)$ and solution for $F(s)$ using Equation 6-36 yields;

$$-\frac{1}{K} = F(s) = \frac{(s + 3)}{s^2 + 10s + 17} = \frac{(s + 3)}{(s + 7.8) \cdot (s + 2.17)}$$

Since the parameter P can vary positively as well as negatively, both the $K > 0$ and $K < 0$ root loci must be computed. These loci are presented in Figure 6-18.

From the $K > 0$ root locus, a positive perturbation of ΔP (or K) does not cause any stability problem; however, a negative perturbation eventually results in instability as shown by the $K < 0$ root locus. The instability point is $s^* = 0$.

In a manner similar to Example 6.16, the K range for stability is computed as $\infty > K > -\frac{17}{3}$, which means that the parameter P must remain larger than $5 - \frac{17}{3} = -\frac{2}{3}$ for the system to remain stable.

FIGURE 6-18 (a) $K > 0$ root locus (b) $K < 0$ root locus

6.7.4 Controls

This section introduces control design using several common controls. The simplest form of feedback compensation is gain compensation. This special form is implemented using either a feedback gain or a cascade gain with unity feedback.

Although gain compensation doesn't always work, it should always be investigated as a means to achieve desired performance before any other type of compensation due to its simplicity.

Table 6-11 presents a summary of the basic compensator structures and their effect on the Root Locus. It is assumed that gain or compensation has already been tried when applying this table.

TABLE 6-11 BASIC COMPENSATOR STRUCTURES

System Performance Deficiency	Required Compensation	Effect on Root Locus	Compensator
Unstable or marginally stable	Add LHP zero	Draws RL left	Lead, rate Feedback
Transient too slow or lightly damped	Add LHP zero	Draws RL left	Lead, rate Feedback
Poor accuracy—error coefficient too small	Add LHP pole	Draws RL right	Lag, PI
Too sensitive at some frequency	Add a pole or zero		

In the remainder of this section, four common controls are introduced and a design procedure is outlined for each. The plant or uncompensated system is designated as $G_x(s)$ in each procedure.

Lag Compensator Design A lag compensator is used to increase low frequency gain to provide better accuracy. The lag compensator transfer function is presented as

$$G_{\text{Lag}}(s) = K \cdot \frac{s + A}{s + A/\alpha} \Big|_{\alpha > 1} \quad (6-38)$$

Design Steps:

- Step 1.** Pick the lag zero ($-A$) to lie at $\frac{1}{10}$ times the distance of the nearest pole or zero to the origin.

Step 2. Select $\alpha = \frac{K_p^{\text{Desired}}}{G_x(0)}$; typically, $10 < \alpha < 50$.

Step 3. Lag pole = $\frac{1}{\alpha} \cdot$ Lag zero

Step 4. Implement in cascade with unity feedback; select K for acceptable transient response. Initial value of $K = 1$. If response is too slow, move zero leftward and repeat design steps 2, 3, and 4.

An example illustrating the design steps for the lag compensator is presented here.

EXAMPLE 6.18 Lag Compensator

A lag compensator is designed to modify the behavior of the plant:

$$G_x(s) = \frac{1}{(s + 1)(s^2 + 2s + 2)}$$

such that the following performance specifications are met.

- $e_{ss}(\text{step}) \leq 0.05$.
- Determine that the system is stable.

Solution

Step 1. Pick the lag zero ($-A$) to lie at $\frac{1}{10}$, which is the distance of the nearest pole or zero to the origin. The plant has no zeros and poles at $s_{1,2,3} = -1, -1 \pm j 1$. The nearest pole is 1 unit from the origin,

so the lag zero is chosen as $A = -\frac{1}{10} \cdot 1 = -0.1$.

Step 2. Select $\alpha = \frac{K_p^{\text{Desired}}}{G_x(0)}$. $G_x(0)$ is easily computed by evaluating the plant transfer function at $s = 0$

to produce $G_x(0) = 0.5$. The K_p^{Desired} is computed from the $e_{ss}(\text{step}) = 0.05$ specification. Note that we are designing to the boundary (or worst case) of this specification. The calculation is presented here.

$$e_{ss}(\text{step}) = 0.05 = \frac{1}{1 + K_p^{\text{Desired}}} \Rightarrow K_p^{\text{Desired}} = 19$$

Next α is calculated according to the formula:

$$\alpha = \frac{K_p^{\text{Desired}}}{G_x(0)} = \frac{19}{.5} = 38$$

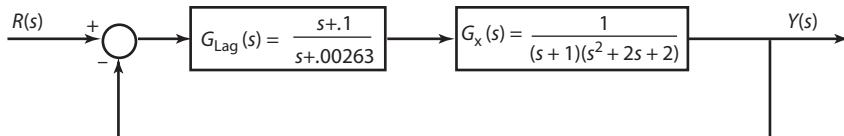
Step 3. Select the lag pole = $\frac{1}{\alpha}$, the lag zero = $\frac{-0.1}{38} = -0.00263$.

Step 4. The lag compensator transfer function is

$$G_{\text{Lag}}(s) = \frac{s + .1}{s + .00263}$$

This compensator is implemented in the cascade configuration shown in Figure 6-19.

FIGURE 6-19



PI Compensator Design A PI compensator serves the same purpose as the lag compensator. It is used to increase low frequency gain to provide better accuracy. The PI compensator transfer function is presented as

$$G_{\text{PI}}(s) = K \cdot \frac{s + A}{s} \quad (6-39)$$

Design Steps

Step 1. Pick the lag zero ($-A$) to cancel the slowest pole of $G_x(s)$.

Step 2. Implement the PI compensator in cascade with unity feedback, and select K for the acceptable transient response.

An example illustrating the design steps for the PI controller is presented here.

EXAMPLE 6.19 PI Compensator

A lag compensator is designed to modify the behavior of the plant:

$$G_x(s) = \frac{1}{(s^2 + 9s + 18)}$$

such that the following performance specifications are met.

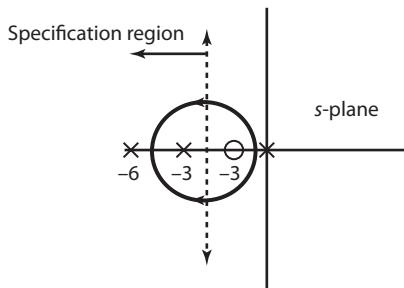
- $e_{\text{ss}}(\text{step}) = 0$.
- Determine that the system is stable.
- $\tau \leq 0.5 \text{ sec}$.

Solution

Step 1. Pick the PI zero ($-A$) to lie at $\frac{1}{10}$, which is the distance of the nearest pole or zero to the origin. The plant has no zeros and poles at $s_{1,2} = -3, -6$. The nearest pole is 3 units from the origin, so the PI zero is chosen as $A = -\frac{1}{10} \cdot 3 = -0.3$.

Step 2. Select K for the desired closed-loop pole locations. For this step, a root locus sketch is necessary. Figure 6-20 presents the PZ plot for the control system ($G_{PI}(s)$ and $G_x(s)$) and the performance specification, which is just a time constant. This time constant value of .5 identifies a pole location at $\frac{-1}{.5} = -2$.

FIGURE 6-20

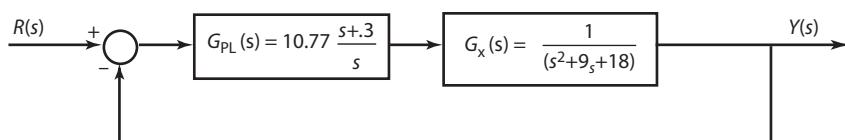


The diameter of the circular path is approximately $4.5 - .15 = 4.35$ with the center at $\frac{-4.35}{2} = -2.15$ ≈ 2.35 . The design point, s^* , is $s^* \approx -2 \pm j 2$. Because of other inaccuracies in the sketch, it is permissible to round (in this case to integers) to simplify the arithmetic. To find the gain value, we simply solve the root locus magnitude condition at s^* . This calculation is

$$\left| \frac{1}{K} \right| = |F(s)| = \frac{|s + .3|}{|s + .00263| \cdot |s^2 + 9s + 18|} \Big|_{s=s^*=-2+j2} = 0.0929 \Rightarrow K = 10.77$$

This compensator is implemented in the cascade configuration shown in Figure 6-21.

FIGURE 6-21



Lead Compensator Design A lead compensator is used to increase system stability and speed its response. The lead compensator transfer function is

$$G_{\text{Lead}}(s) = K \cdot \frac{s + A}{s + A/\alpha} \Big|_{\alpha < 1} \quad (6-40)$$

Design Steps:

Step 1. Given the specifications on the desired closed-loop performance in terms of ζ^{dom} and τ^{dom} , identify the specification region or point, s^* , in the s -plane.

Step 2. Select the lead zero to move the rightmost pole(s) to s^* . If necessary, apply series cancellation compensators of the form $\frac{s + D}{s + E}$ to shift the leftward interfering poles.

Step 3. Select the lead pole = $\frac{\text{Lead zero}}{\alpha}$ ($\alpha \approx \frac{1}{10}$).

Step 4. Implement the lead compensator cascade with unity feedback, and select K for the acceptable transient response.

An example illustrating the design steps for the lead compensator is presented here.

EXAMPLE 6.20 Lead Compensator

A lead compensator is designed to modify the behavior of the plant;

$$G_x(s) = \frac{1}{(s^2 + 1)}$$

such that the following performance specifications are met.

- $\zeta \geq .707$.
- $\tau \leq .1$ sec.
- Determine that the system is stable.

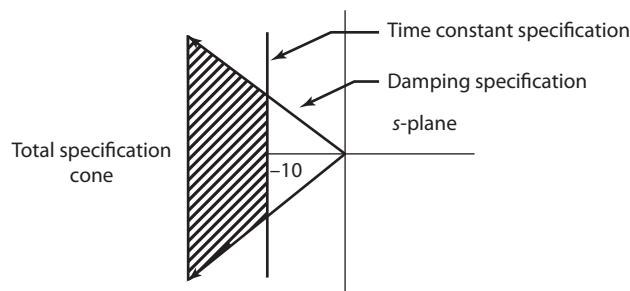
Solution:

Step 1. Given the specifications on the desired closed-loop performance in terms of ζ^{dom} and τ^{dom} , identify the specification region or point, s^* , in the s -plane. The specification region is the cone shown in Figure 6-22.

We can design to any point in the total specification cone. For demonstration purposes, we will select our design point at the corner of the cone as $s^* = -10 + j 10$.

Step 2. Select the lead zero to move the rightmost pole(s) to s^* . The rightmost poles are the dominant poles (that is, they are the poles that govern the fundamental parts of the response). In this example, the plant only has two poles located at $s_{1,2} = \pm j 1$. If a zero were placed at about $s = -10$, the plant poles would tend

FIGURE 6-22



to travel leftward in a circular path with the center at the zero. Along their trajectory, the poles would be dragged through the design point, $s^* = -10 + j 10$, so this zero seems to be satisfactory at this point in the design.

Step 3. Select the lead pole = $\frac{\text{Lead zero}}{\alpha}$ ($\alpha \approx \frac{1}{10}$). The pole becomes $s = -100$

Step 4. The lead compensator transfer function becomes

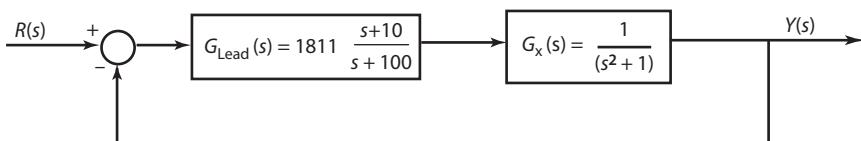
$$G_{\text{Lead}}(s) = K \frac{s + 10}{s + 100}$$

To find the gain value, K , we simply solve the root locus magnitude condition at s^* . This calculation is performed as

$$\left| \frac{1}{K} \right| = |F(s)| = \frac{|s + 10|}{|s + 100| \cdot |s^2 + 1|} \Bigg|_{s=s^*=-10+j10} = 0.00055 \Rightarrow K = 1811$$

This compensator is implemented in the cascade configuration shown in Figure 6-23.

FIGURE 6-23



The value computed for K is only a starting value, which likely will need to be tuned or slightly modified (based on simulation results).

Rate Feedback Compensator Design A rate feedback compensator serves the same purpose as the lead compensator. It is used to increase system stability and speed its response. The rate feedback compensator transfer function is

$$H(s) = K \cdot s \quad (6-41)$$

Design Steps:

Step 1. Given the specifications on the desired closed-loop performance in terms of ζ^{dom} and τ^{dom} , identify the specification region or point, s^* , in the s -plane.

Step 2. Select the rate feedback zero ($-A/K$) to move the rightmost pole(s) to s^* . If necessary, apply series cancellation compensators of the form $\frac{s + D}{s + E}$ to shift the leftward interfering poles.

Step 3. Draw the root locus for $K \cdot \left(s + \frac{A}{K} \right) \cdot Gx$; select K for the desired dominant closed-loop pole location(s).

Step 4. Compute A since you know K and $(-A/K)$.

Step 5. Implement in *rate feedback configuration*—no adjustment is needed. The rate feedback configuration is described in Example 6-21.

EXAMPLE 6.21 Rate Feedback Compensator

A rate feedback compensator is designed to modify the behavior of the plant:

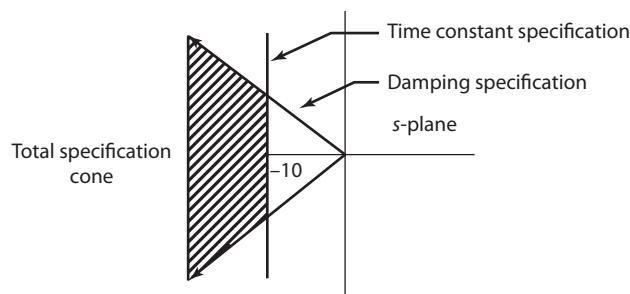
$$G_x(s) = \frac{1}{(s^2 + 2s + 101)}$$

such that the following performance specifications are met.

- $\zeta \geq .707$.
- $\tau \leq .1$ sec.
- Determine if the system is stable.

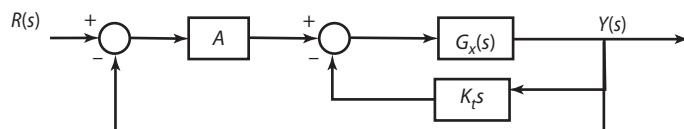
Solution

Step 1. Given the specifications on the desired closed-loop performance in terms of ζ^{dom} and τ^{dom} , identify the specification region or point, s^* , in the s -plane. The specification region is the cone shown in Figure 6-24.

FIGURE 6-24

We can design to any point in the total specification cone. For demonstration purposes, we will select our design point at the corner of the cone as $s^* = -10 + j 10$.

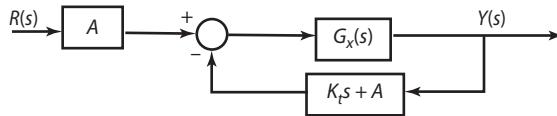
Step 2. The rate feedback compensator is implemented in the rate feedback configuration shown in Figure 6-25.

FIGURE 6-25

The rate feedback configuration is normally applied to systems which have two outputs; one is the controlled variable and the other is the derivative of the controlled variable. Therefore, the inner-loop feedback path, $K_t s$, becomes just K_t , which is the input of the derivative of the controlled variable.

The rate feedback structure can be simplified by moving the A gain leftward through the summing junction and combining the resulting two feedback loops, which are now parallel paths. The simplified diagram is shown in Figure 6-26.

FIGURE 6-26



The feedback loop transfer function can be factored into monic form to produce

$$K_t s + A = K_t \left(s + \frac{A}{K_t} \right)$$

In this form, the zero added by the rate feedback configuration can be clearly seen at $s = -\frac{A}{K_t}$.

Select the lead zero to move the rightmost pole(s) to s^* . The rightmost poles are the dominant poles (that is, they are the poles that govern the fundamental parts of the response). In this example, the plant only has two poles located at $s_{1,2} = -1 \pm j 10$. If a zero were placed at about $s = -20$, the plant poles would tend to travel leftward in a circular path with the center at the zero. Along their trajectory, the poles would be dragged through the design point at $s^* = -10 + j 10$, so this zero seems to be satisfactory at this point in the design.

Step 3. Draw the root locus for $K \cdot \left(s + \frac{A}{K} \right) \cdot G_x$ and select K for desired dominant closed-loop pole location(s). For brevity, we are using K in place of K_t . The gain is calculated as

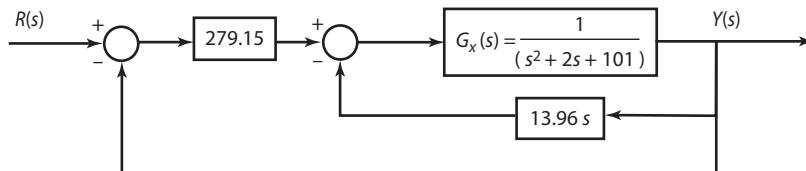
$$\left| \frac{1}{K} \right| = |F(s)| = \frac{|s + 20|}{|s^2 + 2s + 101|} \Big|_{s=s^*=-10+j10} = 0.0716 \Rightarrow K = 13.96$$

Step 4. Compute A , since you know K and $(-A/K)$.

$$20 = \frac{A}{K} = \frac{A}{13.96} \Rightarrow A = 279.15$$

Step 5. The rate feedback controller is shown in Figure 6-27.

FIGURE 6-27



Although the value computed for K is only a starting value, it should be fairly close to the final design value. The design should be simulated to determine whether additional tuning is required.

6.8 Bode Plots

Bode plots provide another method for the design and analysis of controls. While the root locus method is a complex variable method, the Bode plot method is a frequency-based method. Since most signals can be equivalently represented in the frequency domain, the Bode plot method is often preferred for control system design.

Consider a transfer function given in monic form:

$$G(s) = K \cdot \frac{N(s)}{D(s)}$$

The transfer function is rewritten in *Bode form* as

$$G(jw) = K \cdot \frac{N(jw)}{D(jw)} = K \cdot \frac{(1 + jw/A_1)(1 + jw/A_2)\dots}{(1 + jw/B_1)(1 + jw/B_2)\dots}$$

The important property of the Bode form representation is that the transfer function always has a gain of K at zero frequency. Once a transfer function is converted to Bode form, it is easy to compute the magnitude and phase of the transfer function.

The log-magnitude (LM) of $G(jw)$ equals the sum of the numerator LMs minus the sum of the denominator LMs. This addition operation is made possible due to the use of LM. The phase of $G(jw)$ equals the sum of the numerator angles minus the sum of the denominator angles. Both the LM and phase consist of the sum of individual terms, and these terms can be categorized being any of the following.

1. Gain term.
2. Determine the pole or zero at the origin.
3. Determine the first-order pole or zero (distinct).
4. Determine the second-order pole or zero (complex).

Straight line *sketches* of Bode plots can be performed quickly with a pencil on paper and can provide a fairly good representation of the actual Bode plot. Table 6-12 summarizes the necessary steps in the sketching process.

TABLE 6-12 BODE PLOT SKETCHING STEPS

Step	Description
1	Put $G(s)$ into Bode form, $G(jw)$.
2	Set up the Bode plot axis, identify the rightmost and leftmost pole or zero in $G(jw)$, and set the Bode frequency range to go from .01* at the rightmost pole, zero to 100* at the leftmost pole, or zero.
3	Set $K = 1$ and plot the magnitude response of each term in $G(jw)$.
4	Plot the phase response of each term in $G(jw)$.
5	Compute the <i>total</i> LM by summing the slopes for each term in step 3, and scale the result by $LM(K)$.
6	Compute the total phase by summing the phases for each term in step 4.

6.8.1 Controls

The open-loop Bode frequency response can be used to *approximate* closed-loop system behavior provided the open-loop frequency response satisfies three conditions:

- The gain at low frequencies is relatively high for low steady-state errors.
- The gain at high frequencies is relatively low for disturbance attenuation.
- The slope of the LM plot is approximately -1 LM/decade in the vicinity $.4\omega_c \rightarrow 4\omega_c$, where ω_c = crossover frequency of the LM plot which is the frequency that the LM trace passes through the 0 LM point.

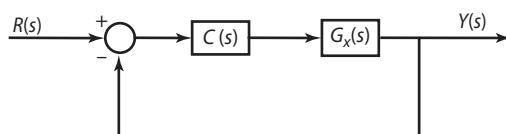
Provided these three conditions are approximately satisfied, the following two correlations (Equations 6-42 and 6-43) are valid.

$$\tau \approx \frac{1}{\omega_c} \equiv \text{time constant, seconds} \quad (6-42)$$

$$\zeta \approx \begin{cases} .01 \cdot \phi_{pm} & \text{for } \phi_{pm} < 60^\circ \\ .027 \cdot \phi_{pm} - 1 & \text{for } 75^\circ > \phi_{pm} > 60^\circ \end{cases} \quad (6-43)$$

There are several basic types of compensators used in classical control applications. The majority of these compensators are normally placed in *series* with the plant, and feedback is applied around the combination. This configuration is called *cascade* and is illustrated in Figure 6-28.

FIGURE 6-28 CASCADING CONFIGURATION



In Figure 6-28, the system to be compensated (the uncompensated system) is $G_x(s)$, and the compensator is $C(s)$. Negative feedback is applied around the series combination. The signal $R(s)$ is a reference signal, and the signal $Y(s)$ is the output signal.

Four basic types of compensators are discussed in this section:

1. Proportional—Derivative (PD) compensator
2. Lead compensator
3. Proportional—Integral (PI) compensator
4. Lag compensator

The design steps for each of these compensators are based on the bode plots of the plant, $G_x(s)$, and are presented in the following sections.

Lag Compensator Design The primary function is to increase the low frequency gain to reduce steady-state errors and decrease sensitivity to noise and parameter variations. The general form of the lag compensator transfer function is presented as

$$\text{Lag transfer function: } C(s) = K \frac{(Ts + 1)}{(\alpha Ts + 1)} \quad (6-44)$$

The quantity, $1/\alpha$ is called the lag ratio, and it is responsible for the amount of gain added by the compensator.

Design Steps

Step 1. Determine the frequency, ω^* , which defines the region, $\omega < \omega^*$, in which the gain is to be increased. Determine how much gain is to be added, K^* , using the error coefficient specifications for the system.

Step 2. Identify the crossover frequency and phase margin from the uncompensated system Bode plots, ω_c and ϕ_{pm} .

Step 3. Select the zero of the lag compensator, $1/T$, to satisfy the steady-state error specifications; however, do not allow it to move higher than $.4\omega_c$ if a phase margin problem exists (phase margin less than 30°).

Step 4. Compute the lag ratio, $\frac{1}{\alpha} = \frac{1}{K^*}$, and compute the lag pole as $\frac{1}{\alpha} \cdot \text{Lag zero}$.

Step 5. Begin with $K = 1$ and tune the compensator by *slightly* adjusting K until the desired performance is achieved.

An example illustrating the design steps for the lag compensator is presented here.

EXAMPLE 6.22 Bode Lag Compensator Design

A lag compensator is designed to achieve the following performance specifications.

- Lag zero selected as $-\frac{1}{T} = -0.1$.
- Select $\alpha = 10$ for a gain factor of 10. The lag ratio becomes $\frac{1}{\alpha} = \frac{1}{10}$.
- Select $K = \alpha$, initially.

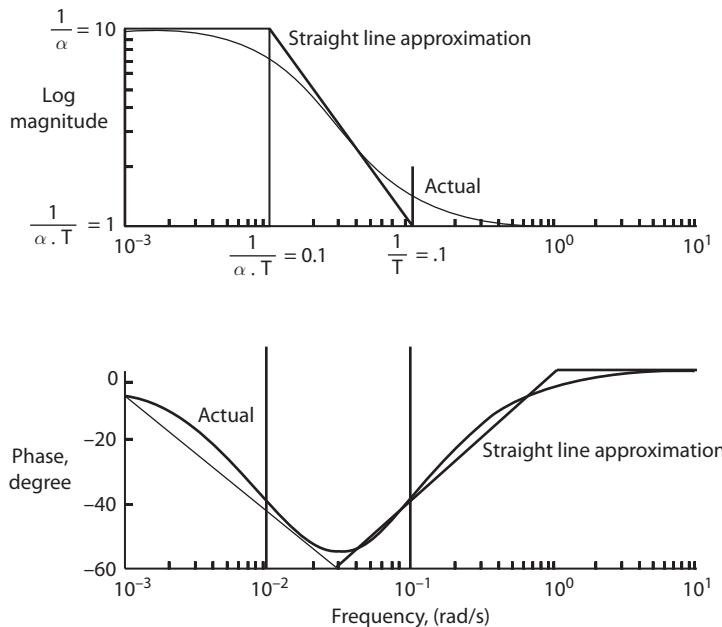
Applying this information to the general lag compensator produces the transfer function in Equation 6-45 for the lag compensator.

$$C(s) = K \frac{(Ts + 1)}{(\alpha Ts + 1)} = \frac{s + .1}{s + .01} \quad (6-45)$$

Solution

The Bode plots, log magnitude, and phase plots for the lag compensator are presented in Figure 6-29.

In Figure 6-29, the straight line approximations along with the actual LM and phase traces have been included to illustrate the error between the two. The lag compensator accomplishes its purpose—that of

FIGURE 6-29 LAG COMPENSATOR BODE PLOTS

providing a factor of 10 times more gain at low frequencies (less than .01 rad/sec), and no gain change at higher frequencies (above .1 rad/sec).

Proportional Integral (PI) Compensator Design The primary function is to increase the low frequency gain to reduce steady-state errors and decrease sensitivity to noise and parameter variations. The general form of the PI compensator transfer function is

$$\text{PI transfer function: } C(s) = K \frac{(Ts + 1)}{s} \quad (6-46)$$

Since the function of the PI compensator is to increase low frequency gain, the design steps are somewhat different than the previous compensator. The design steps for the PI compensator are outlined in the following steps.

Design Steps

Step 1. Determine the frequency, ω^* , which defines the region, $\omega < \omega^*$, in which the gain is to be increased. Determine how much gain is to be added, K^* , using the error coefficient specifications for the system.

Step 2. Identify the crossover frequency and phase margin from the uncompensated system Bode plots, ω_c and ϕ_{pm} .

Step 3. Select the zero of the PI compensator to satisfy the steady-state error specifications; however, do not allow it to move higher than $.4\omega_c$ if a phase margin problem exists (phase margin less than 30°).

Step 4. Initially, select $K = 1$ and tune the compensator slightly by modifying the value of K for the desired performance.

An example illustrating the design steps for the PI compensator is presented here.

EXAMPLE 6.23 Bode PI Compensator Design

A PI compensator is designed to achieve the following performance specifications.

- Select $T = 10$ to add a gain factor of 10.
- Select $K = 1$ initially, this will produce a high frequency gain of $K \cdot T = 10$.

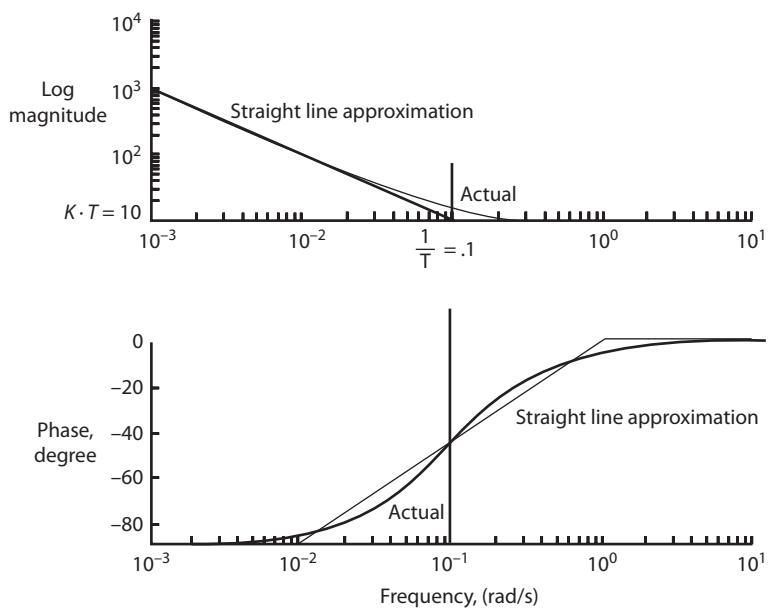
Applying this information to the general PI compensator produces the transfer function in Equation 6-47 for the PI compensator.

$$C(s) = K \frac{(Ts + 1)}{s} = 10 \cdot \frac{s + .1}{s} \quad (6-47)$$

Solution

The Bode plots, log magnitude, and phase plots for the PI compensator are presented in Figure 6-30.

FIGURE 6-30 BODE PLOTS FOR THE PI COMPENSATOR



In Figure 6-30 the straight line approximations along with the actual LM and phase traces have been included to illustrate the error between the two. The PI compensator provides infinite gain at zero frequency with a phase lag of 90° . At high frequencies, above .1 rad/sec, the compensator was designed to provide a gain of 10 with zero phase loss. This value could be modified by selecting a different K , for example, had unity gain been desired at high frequencies K would be selected as .1.

Lead Compensator Design Primary function: Add, phase lead to a system for faster and less oscillatory response and control the amount of gain added at high frequencies. The general form of the lead compensator transfer function is presented in Equation 6-48.

$$\text{Lead transfer function: } C(s) = K \frac{(Ts + 1)}{(aTs + 1)} \quad (6-48)$$

The quantity, $1/\alpha$ is called the *lead ratio* and it is responsible for the amount of phase added by the lead compensator. The larger the lead ratio, the more spread apart the pole and zero of the lead compensator and consequently the more phase added. The peak phase added lies at a frequency halfway between the pole and zero of the compensator at $\frac{1}{\sqrt{\alpha} \cdot T}$. The relationship between the lead ratio and phase for the compensator is summarized in Table 6-13.

TABLE 6-13 RELATIONSHIP BETWEEN LEAD RATIO AND PHASE FOR A LEAD COMPENSATOR

Lead Ratio, $1/\alpha$	Phase, deg
1	0
2	17.5
5	37.5
10	55
20	62.5
50	72.5
100	80

Design Steps

Step 1. Determine the frequency, ω^* , where phase is to be added and determine how much phase is to be added, ϕ^* based on the Bode plots of the uncompensated system.

Step 2. Compute the lead ratio; $\frac{1}{\alpha}$, given the ϕ^* using the lead table.

Step 3. Compute the lead pole and zero;

$$\text{Zero: } \frac{1}{T} = \sqrt{\alpha} \cdot \omega^*$$

$$\text{Pole: } \frac{1}{\alpha} \cdot \frac{1}{T}$$

Step 4. Implement the lead compensator and set K slightly less than 1 to counteract the additional gain of the compensator. Adjust the K for desired transient response.

An example illustrating the design steps for the lead compensator is presented here.

EXAMPLE 6.23 Bode Lead Compensator Design

A lead compensator is designed to achieve the following performance specifications.

- Select the frequency to add phase lead at $\omega^* = 10 \text{ rad/sec}$.
- Add 55° of phase at ω^* by selecting the lead ratio as $\frac{1}{\alpha} = 10$.
- Compute the lead zero, $-\frac{1}{T} = \sqrt{\alpha} \cdot \omega^* = -3.16$.
- Compute the lead pole, $-\frac{1}{\alpha \cdot T} = -31.6$.
- Select $K = 1$ initially.

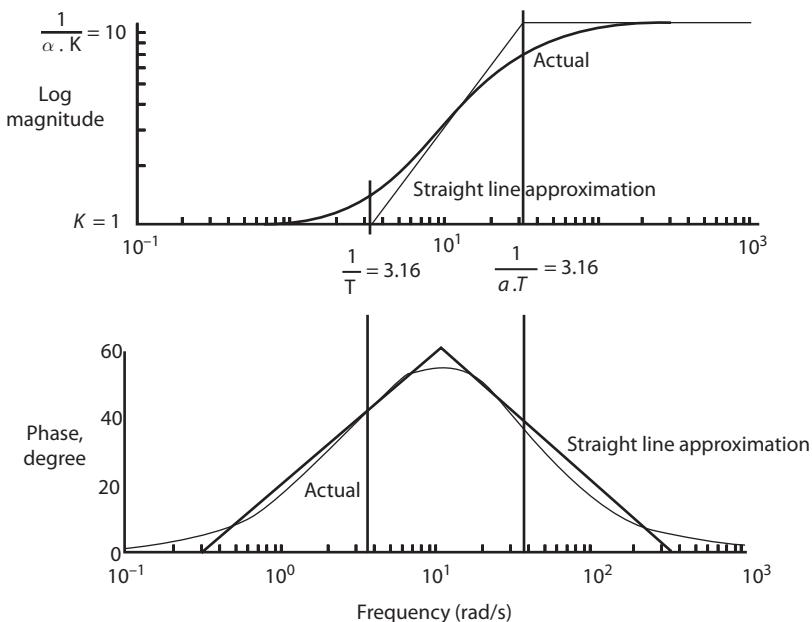
Applying this information to the general lead compensator produces the transfer function, Equation 6-49, for the lead compensator.

$$C(s) = K \frac{(Ts + 1)}{(\alpha Ts + 1)} = 10 \cdot \frac{s + 3.16}{s + 31.6} \quad (6-49)$$

Solution

The Bode plots, log magnitude and phase plots, for the lead compensator are presented in Figure 6-31.

FIGURE 6-31 BODE PLOTS FOR A LEAD COMPENSATOR



In Figure 6-31 the straight line approximations along with the actual LM and phase traces have been included to illustrate the error between the two.

Proportional Derivative (PD) compensator Design Primary function: , Add phase lead to a system for faster and less oscillatory response. The general form of the PD compensator transfer function is

$$\text{PD transfer function: } C(s) = K(Ts + 1) \quad (6-50)$$

The characteristics of the PD compensator phase contributions are summarized in Table 6-14.

TABLE 6-14 PD COMPENSATOR PHASE RESPONSE

Frequency	LM	Phase
<<< 1/T	K	0
1/T	K	45
>>> 1/T	∞	90

Design Steps

Step 1. Determine the frequency, ω^* , where phase is to be added and determine how much phase is to be added using the PD table.

Step 2. Select the PD zero at the frequency which produces the desired phase (interpolate where necessary).

Step 3. Implement the compensator and adjust the K gain for a satisfactory transient response, beginning with $K = 1$.

Although the PD compensator provides phase lead at frequencies above $1/T$, it also adds more and more amplification as frequency increases. When noise is present (as it usually is), this is an undesirable property, because the PD compensator amplifies the noise to a large degree.

An example illustrating the design steps for the PD compensator is presented here.

EXAMPLE 6-24 Bode PD Compensator Design

A PD compensator is designed to achieve the following performance specifications.

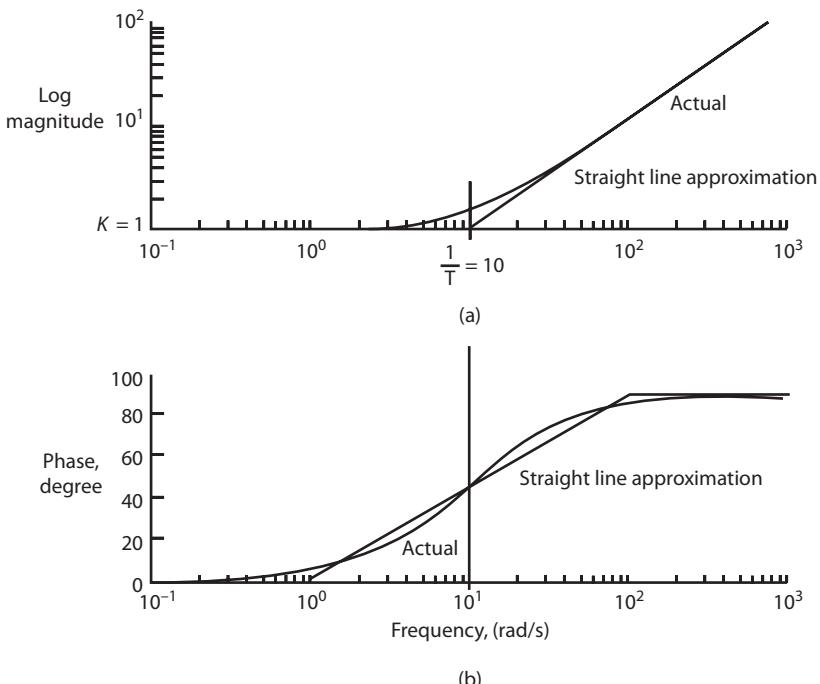
- Select the frequency to add phase lead at $\omega^* = 10 \text{ rad/sec}$
- Add 45° of phase at ω^* by selecting the *PD* zero at, $-\frac{1}{T} = -10$
- Select $K = 1$ initially

Applying this information to the general PD compensator produces the transfer function, Equation 6-51, for the PD compensator.

$$C(s) = K(Ts + 1) = .1 \cdot (s + 10) \quad (6-51)$$

Solution

The Bode plots, log magnitude and phase plots, for the PD compensator are presented in Figure 6-32.

FIGURE 6-32 BODE PLOTS FOR PD COMPENSATOR

In Figure 6-32 the straight line approximations along with the actual *LM* and phase traces have been included to illustrate the error between the two.

6.9 Controller Design Using Pole Placement Method

This section introduces the pole placement method of designing the controller. In some of the previously discussed methods like Bode Plot or Root Locus, the system performance characteristic is obtained for different values of gain ‘K’ and the value of ‘K’ that gives the desired system performance is selected. Unlike those methods in Pole Placement, the desired closed loop characteristic equation is obtained based on the system performance characteristic and then the unknown gains are computed by comparing the desired characteristic equation with the actual closed loop characteristic equation.

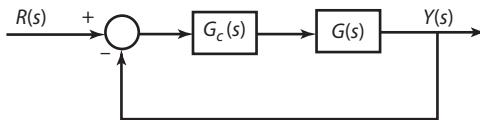
The steps involved in the controller design using pole placement are outlined in the following;

Step 1. Determine the desired closed loop characteristic equation based on the required system performance characteristic. The characteristic equation for a second order system is given by

$$s^2 + 2\zeta\omega_n s + \omega_n^2 = 0 \quad (6-52)$$

Step 2. Obtain the closed loop characteristic equation. For a unit feed back system shown in Figure 6-33 the characteristic equation is given by

$$1 + G(s)*G_c(s) = 0 \quad (6-53)$$

FIGURE 6-33 UNIT FEEDBACK SYSTEM

where $G(s)$ is the transfer function of the system and $G_c(s)$ is the transfer function of the controller.

Step 3. Equating both characteristic equations (Equations 6-52 and 6-53) will give the unknown parameters of the controller.

EXAMPLE 6.25 DC Gear Motor Controller Design

Obtain the closed-loop characteristic equation for the PM DC Gear Motor system described in Chapter 4 (see Figure 4.2). Also design the controller that will control the motor position with the following system performance characteristics.

- Settling time = 0.5 sec
- Percent overshoot should not exceed 16%
- Steady-state error should be less than 2%

Solution

The open-loop transfer function of the system is obtained from Equation 4-16 and can be expressed as

$$G(s) = \frac{k}{\frac{mr^2}{G^2} L_a s^5 + R_a \frac{mr^2}{G^2} s^2 + k^2 s} \quad (6-54)$$

Let us introduce a PID controller in the system with a unit feedback as shown in Figure 6-33. The transfer function for the PID controller is

$$G_c(s) = P + D * s + \frac{I}{s} = \frac{Ds^2 + Ps + I}{s} \quad (6-55)$$

From Equations 6-53, 6-54, and 6-55, the closed-loop characteristic equation can be obtained as

$$1 + \left(\frac{Ds^2 + Ps + I}{s} \right) \left(\frac{k}{\frac{mr^2}{G^2} L_a s^5 + R_a \frac{mr^2}{G^2} s^2 + k^2 s} \right) = 0 \quad (6-56)$$

$$\begin{aligned} & \left[\frac{mr^2}{G^2} L_a s^3 + R_a \frac{mr^2}{G^2} s^2 + k^2 s \right] s + k(Ds^2 + Ps + I) = 0 \\ & \frac{mr^2}{G^2} L_a s^4 + R_a \frac{mr^2}{G^2} s^3 + (k^2 + kD)s^2 + kPs + kI = 0 \end{aligned} \quad (6-57)$$

Using the values from the given data (see Chapter 4, pg. 303), we get

$$\frac{mr^2}{G^2} L_a = 3.8 \times 10^{-11} \approx 0$$

and

$$R_a \frac{mr^2}{G^2} = 4.65 \times 10^{-6} \approx 0$$

Hence, neglecting the first and second term of Equation 6-57, we get a second-order characteristic equation for the closed loop system given as

$$\begin{aligned} s^2 + \frac{kP}{(k^2 + kD)} s + \frac{kI}{(k^2 + kD)} &= 0 \\ s^2 + \frac{P}{(k + D)} s + \frac{I}{(k + D)} &= 0 \end{aligned} \quad (6-58)$$

On comparing Equations 6-52 and 6-58, we get

$$\frac{P}{(k + D)} = 2\zeta\omega_n \quad (6-59)$$

and

$$\frac{I}{(k + D)} = \omega_n^2 \quad (6-60)$$

Equations 6-59 and 6-60 can be used to find the PID gains. However, we have two equations and three unknowns. Hence, considering the derivative gain to be zero, we get

$$P = 2k\zeta\omega_n \quad (6-61)$$

and

$$I = k\omega_n^2 \quad (6-62)$$

ζ and ω_n can be calculated using the system performance requirements using the expression for settling time and percent overshoot for a unit step response of a second order system as

$$\zeta = \frac{|\ln(\text{Percent Overshoot}/100)|}{\sqrt{\pi^2 + [\ln(\text{Percent Overshoot}/100)]^2}} = \frac{|\ln(0.16)|}{\sqrt{\pi^2 + [\ln(0.16)]^2}} = 0.5$$

and

$$\omega_n = \frac{|\ln(\text{steady state error}/100)|}{\text{settling times } X \zeta} = \frac{|\ln(0.02)|}{0.5 \times 0.5} = 15.65/\text{sec}$$

Hence,

$$P = 2k\zeta\omega_n = 2 \times 0.032 \times 0.5 \times 15.65 = 0.5/\text{sec}$$

$$I = k\omega_n^2 = 0.032 \times 15.65^2 = 7.83/\text{sec}^2$$

The required PI controller is

$$G_c(s) = \frac{0.5s + 7.83}{s}$$

MATLAB Code

Using MATLAB simulate the response of a closed loop PM DC gear motor system described in Chapter 4 (see Figure 4.2), which uses the PI controller designed with a negative feedback, as shown in Figure 6-33. The desired position to reach is 200 radians.

```

Clear
clc
Ra = 20.5; %Armature Resistance, ?
La = 168E-6; %Armature Inductance, H
k = 0.032; %Motor Constant, Nm/A (or V/rad/sec)
G = 49; %Gear Ratio
m = 1.125; %Mass, KG
r = 0.022; %Radius of the pulley, m
g=9.81; %Acceleration due to gravity, m^2/sec
P=0.5; %Proportional Gain
I=7.83; %Integral Gain
thetha=200; %desired position
%open loop transfer function
Gs=tf(k,[m*r^2*La/G^2 Ra*m*r^2/G^2 k^2 0]);
%controller transfer function
Gcs=tf([P I],[1 0]);
sys1=Gcs*Gs;
sys2=1;
sys=feedback(sys1,sys2); %Step Response of the closed loop system
t=0:0.01:10;
U=thetha*ones(size(t));
lsim(sys,U,t)
sys3=feedback(Gcs,Gs); %Output from the PI Controller
figure
lsim(sys3,U,t)
ylabel('Angular displacement of the motor shaft (rad)')

```

Result

Figure 6-34 shows the response of the closed loop system. As seen from the figure, we have not yet reached the desired system performance. The overshoot is 33.5% ((267 – 200)/200) and the settling time is 0.7 sec. This is because the open loop plant is a forth order system and we reduced it to a second order system. Also, the expressions used for calculating ζ and ω_n are for a unit step response of a second order system. Hence, we would need some tuning in the PI gains to reach the desired value.

The other factor we have to consider while designing the controller is how close we are able to create an environment for the mathematical model (dynamical system) when compare to the environment in which the real system will be operating. Let us check the response of the controller output before tuning the PI Gains. Figure 6-35 shows the response of the PI controller.

As seen from Figure 6-35, initial output from the PI is 103 which is actually the voltage that goes into the motor. Hence, we have to limit the PI output depending on the available power supply. For example if we are using a 24V DC power supply we have to limit the PI output to ± 24 V. This will also help in reducing

FIGURE 6-34 STEP RESPONSE OF THE CLOSED-LOOP SYSTEM FOR $P = 0.5$ and $I = 7.83$

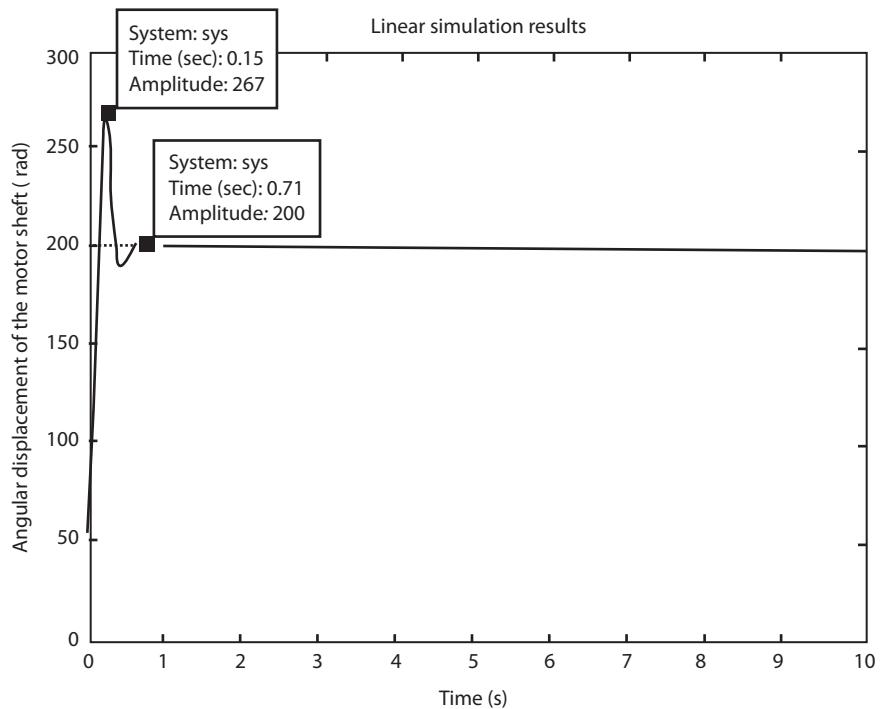
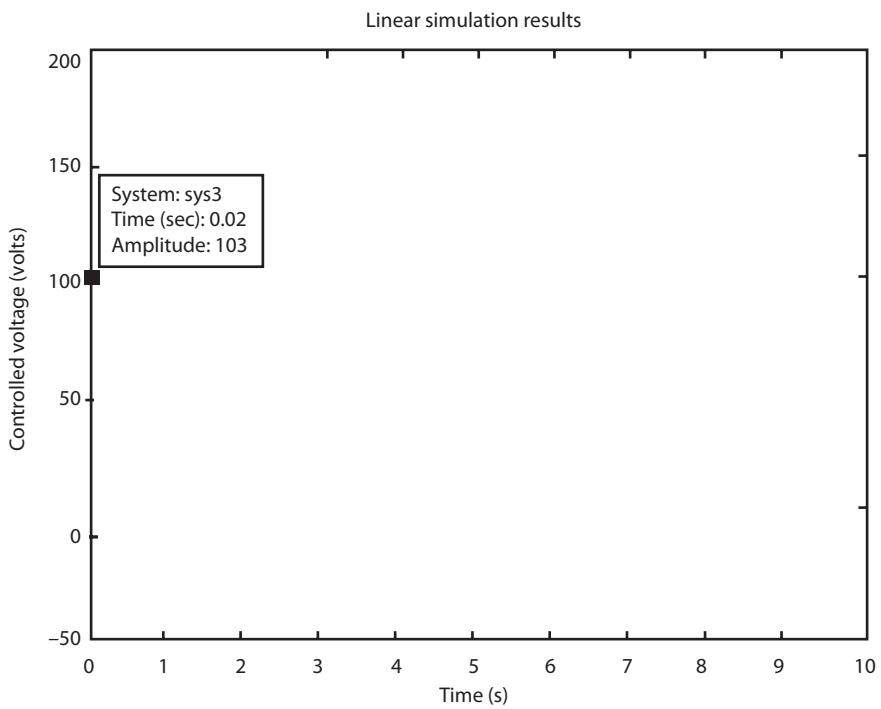


FIGURE 6-35 PI CONTROLLER RESPONSE



the overshoot value. Also, we would need to incorporate the voltage $\left(R_a \frac{mgr}{kG}\right)$ used by the weight of the body attached to the gear motor (see Figure 4.4) in the dynamical system for getting the response close to the real system. The tuning of the PI should also compensate for the assumption that the inertia and damping losses in the sysyem is negligible. In Chapter 8, we will cover in details how to implement the mathematical model of the PM DC motor closed-loop system for the environment similar to the real system and perform the required tuning of the PI gains. Also, we will see how to implement the real system and compare the response of both dynamical and real system.

6.10 Summary

During the design stage of a mechatronics system it is necessary to understand the performance aspects of the individual system components, such as the sensors and actuators, as well as the overall integrated system performance. Modeling plays a critical role in this stage of development and the ability to construct accurate nonlinear models from illustrations is essential. This chapter has introduced the modified analogy approach to satisfy this requirement. The approach differs from the basic analogy approach in its ability to incorporate nonlinear behavior directly into the model. In addition to this feature, the modified approach results in a block diagram system model which is the best format for multidisciplinary (mechatronic) applications.

REFERENCES

- | | |
|--|--|
| Kuo, Benjamin C., <i>Automatic Control Systems</i> , Third Edition. Prentice-Hall Inc., New Jersey, 1975. | Raven, Francis H. <i>Automatic Control Engineering</i> , Third Edition. McGraw-Hill., New York, 1978. |
| D'Azzo, John J. and Constantine, Houpis H., <i>Linear Control System Analysis and Design Conventional and Modern</i> , 2d Edition. McGraw-Hill., New York, 1981. | Dorf, Richard C. <i>Modern Control Systems</i> , Sixth Edition. Addison-Wesley Publishing Co., New York, 1992. |

PROBLEMS

- 6.1. Compute the steady-state step error for a *BFS* system using

$$G(s) = \frac{1}{s + 1} \text{ and } H(s) = \frac{s + 2}{s + 10}$$

- 6.2. Compute the steady-state ramp error for a *BFS* system using

$$G(s) = \frac{1}{s(s + 5)} \text{ and } H(s) = 1$$

- 6.3. Compute the steady-state parabolic error for a *BFS* system using

$$G(s) = \frac{10}{s^2} \text{ and } H(s) = 1$$

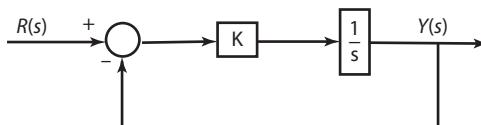
6.4. Compute the steady-state error for a *BFS* system using

$$G(s) = \frac{1}{s(s + 5)} \text{ and } H(s) = 1$$

with the input defined as $R(t) = 2u(t) + 10r(t - 1)u(t - 1)$, where $r(t)$ is the unit ramp function and $u(t)$ is the unit step function.

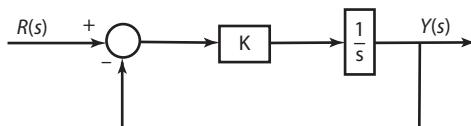
6.5. Use the root-locus magnitude condition to find the K -range for stability for the system in Figure P6-5.

FIGURE P6-5



6.6. The situation illustrated in Figure P6-6 occurs frequently when a digital computer is used to control a continuous system. The digital computer is modeled as a pure time delay of T seconds. The continuous system is a simple integrator. Use the root locus magnitude condition to find the maximum time delay such that the system remains stable. Hint: Use a Pade approximation to represent the time delay as a rational transfer function and substitute $K = \frac{2}{T}$ to simplify your calculations.

FIGURE P6-6



6.7. Design a lag controller using root locus techniques for the plant,

$$G_x(s) = \frac{1}{s(s^2 + 20s + 101)}$$

such that the following performance specifications are met.

- $e_{ss}(\text{step}) = 0$
- $e_{ss}(\text{ramp}) = 0.01$
- System is stable

6.8. Design a PI controller using root locus techniques for the plant,

$$G_x(s) = \frac{1}{s^2 + 9s + 18}$$

such that the following performance specifications are met.

- $e_{ss}(\text{step}) = 0$
- $\zeta = 0.707$
- $\tau = 0.5\text{s}$
- System is stable

6.9. Design a lead controller using root locus techniques for the plant,

$$G_x(s) = \frac{100}{s^2 + 100}$$

such that the following performance specifications are met.

- $\zeta \geq 0.707$
- $\tau \leq 0.1\text{s}$
- System is stable

6.10. Design a lead controller using root locus techniques for the plant,

$$G_x(s) = \frac{1}{(s + 1)(s^2 + 1)}$$

such that the following performance specifications are met.

- $\zeta \geq 0.707$
- $\tau \leq 0.5\text{s}$
- System is stable

6.11. Design a lead controller using root locus techniques for the plant,

$$G_x(s) = \frac{1}{s(s^2 + 2s + 2)}$$

such that the following performance specifications are met.

- $\zeta = 1.0$
- $\tau \leq 0.2\text{s}$
- System is stable

6.12. Design a lead controller using root locus techniques for the plant,

$$G_x(s) = \frac{1}{s(s + 10)(s^2 + 1)}$$

such that the following performance specifications are met.

- $\zeta = 0.707$
- $\tau \leq 0.2\text{s}$
- System is stable

6.13. Design a rate-feedback controller using root locus techniques for the plant,

$$G_x(s) = \frac{101}{s^2 + 2s + 101}$$

such that the following performance specifications are met.

- $\zeta \geq 0.707$
- $\tau \leq 0.5\text{s}$
- System is stable

6.14. Design a PI controller using Bode techniques for the plant,

$$G_x(s) = \frac{1}{s + 10}$$

such that the following performance specifications are met.

- $e_{ss}(\text{step}) = 0$
- $e_{ss}(\text{ramp}) \leq 0.05$
- $\zeta = 1$
- $\tau \leq 0.1\text{s}$
- System is stable

6.15. Design a lead controller using Bode techniques for the plant,

$$G_x(s) = \frac{10}{s(s + 1)}$$

such that the following performance specifications are met.

- $\zeta = 0.5$
- $\tau = 0.1\text{s}$
- System is stable

6.16. Design a lead controller using Bode techniques for the plant,

$$G_x(s) = \frac{1}{s^3}$$

such that the following performance specifications are met.

- $\zeta = 0.5$
- $\tau = 0.1\text{s}$
- System is stable

CHAPTER 7

SIGNAL CONDITIONING AND REAL TIME INTERFACING

- 
- 7.1 Introduction
 - 7.2 Elements of a Data Acquisition and Control System
 - 7.2.1 Overview of the I/O Process
 - 7.2.2 General Purpose I/O Card (GPIO)
 - 7.2.3 Installation of the I/O Card and Software
 - 7.3 Transducers and Signal Conditioning
 - 7.4 Devices for Data Conversion
 - 7.4.1 Operational and Instrumentation Amplifiers
 - 7.5 Data Conversion Process
 - 7.5.1 The Analog-to-Digital Converter
 - 7.5.2 Successive Approximation—Type of A/D Converter
 - 7.6 Application Software
 - 7.6.1 LabVIEW Environment
 - 7.6.2 LabVIEW Applications
 - 7.6.3 VisSim Environment
 - 7.6.4 VisSim Applications
 - 7.7 Summary
 - References

This chapter presents theoretical and practical aspects of computer interfacing and real-time data acquisition and control. Besides the computer and the real-world system, the remaining devices are sensors, actuators, and a general-purpose input/output interface card that includes the A/D and D/A devices. Signal-conditioning accessories amplify low-level signals and then isolate and filter them for more accurate measurements. Amplifier selection and analog-to-digital conversion techniques are the focus of the analog electronics section. Mechatronics integrates signal conditioning, hardware interfacing, control systems, and microprocessors. Signal processing and data interpretation also are handled using the visual programming approach. The versatility of our visual programming environment allows us to present three popular systems: LabVIEW, MATLAB, and VisSim.

7.1 Introduction

Real-time interfacing is a general term used to describe the aspects of connecting a computer with a real-world process and communicating data between the two. Monitors, keyboards, printers, disks, modems, and CD's are familiar (but specific) examples of real-time interfacing. A more general approach categorizes the interfacing process into four major components: sensors, actuators, the computer, and a real-world process. For example, the process of entering data into a computer satisfies this categorization if the human operator is viewed as the real-time process. The sensor is the keyboard; it transfers information from the real-time process to the computer. The monitor is the actuator, transferring information from the computer back to the real-time process. In this chapter, we focus on a broader class of real-time processes which can be human as well as machine based. Sensors and actuators then become components which transfer information between the electrical computer discipline and others (including electrical, mechanical, fluid, thermal, and human). This chapter is designed to be self-contained and does not rely on prior programming knowledge.

7.2 Elements of a Data Acquisition and Control System

A data acquisition (DA) system is a collection of add-on hardware and software components that allow your computer to receive real-world information from sensors. Although sensors can be based on electrical, mechanical, optical, or other principles, they all perform the same function: to convert real-world information (such as motion, temperature, and pressure) into low-power electrical signals which can be read by the computer. Once the data resides within the computer, any of three operations may be performed: plotting, processing, and writing to a file.

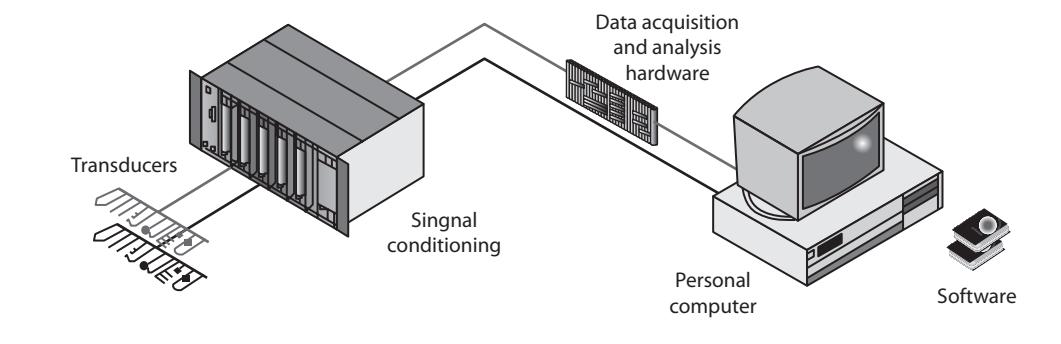
A data acquisition system also can be thought of as a monitoring system. It can receive data from a real-world process and display the data. It can also display the features of the data extracted through its processing. In situations where it is necessary to acquire and process data and to also send data back to the real-time process, we make use of a data acquisition and control (DAC) system. A DAC system is a superset of a DA system and requires both sensors and actuators. The purpose of the actuator is to convert low-power computer signals real-world signals, into resulting in motion, heat, pressure, etc. Common actuators include stepper motors, solenoids, relays, hydraulic motors, speakers, and piezoelectric actuators.

As an example that illustrates the difference between a DA and a DAC system, consider the measurement of the speed–flow–pressure characteristics of a variable speed compressor at 20 different speed values. The compressor has been connected to a variable speed motor—the speed of which can be manually varied by turning a knob. Signals from flow, pressure, and speed sensors attached to the compressor outlet and motor are read and plotted by a computer. The DA system reads and displays the three sensor input values. At each speed, the operator must vary the speed knob until the desired speed is reached and then record the sensor readings for speed, flow, and pressure. In the DA system just described, there is a fair amount of variation in the speed reading due to the speed adjustment at each point by the operator. A DAC version of the same system could be used to remove this speed variation. The DAC application could be created by adding an actuator that moves the speed knob in response to a signal produced by the computer. If the value of this signal is programmed as a function of the error between the actual motor speed and the desired speed, the knob will be moved in the direction that reduces this error. Eventually, the actual motor speed will agree precisely with the desired speed. An application program could be written to automate this process further—to reset the desired speed value to each of the 20 desired values, wait for the flow and pressure signals to stabilize, and record and save a measurement.

The current trend is to use personal computers (PCs) with DAQ hardware for data acquisition in areas of laboratory research, testing and measurement, and industrial automation. The DAQ hardware which act as an interface between the computer and the outside world could be in the form of modules that can be connected to the computer's ports (parallel, serial, USB, etc.) or cards connected to slots (PCI, ISA, PCI-Express, etc.) in the mother board. The newest DAQ devices offer connectivity over wireless and cabled ethernet for remote or distributed DAQ applications. The DAC system presented in Figure 7-1 shows a screw terminal panel with I/O devices.

The PC-based DAQ system depends on each of the following system elements.

- Computer
- Transducers and sensors
- Signal conditioning
- DAQ hardware

FIGURE 7-1 THE TYPICAL PC-BASED DAQ SYSTEM

- Screw terminal panel(s)
- General purpose input/output (GPIO) card
- Software

7.2.1 Overview of the I/O Process

The input/output (I/O) process is the means by which a computer communicates with real-world phenomena via the DAQ device. The performance of the I/O process therefore is dependent on the available computer, selected DAQ device, and the bus architecture. Today's computer (with high-speed processor coupled with high-performance bus architecture) has the capability of transferring data by any of the following methods.

- **Direct Memory Access (DMA):** With this mechanism data is transferred between the DAQ device and computer memory without the involvement of the CPU. This mechanism makes DMA the fastest available data transfer mechanism. Also, the processor is not burdened with moving data, and hence, it can engage in more complex processing tasks.
- **Interrupt Request (IRQ):** IRQ transfers rely on the CPU to service data transfer requests. The device notifies the CPU when it is ready to transfer data. Hence, the data transfer speed is tightly coupled to the rate at which the CPU can service the interrupt requests.
- **Programmed I/O:** This is a data transfer mechanism in which a buffer is not used—instead the computer reads and writes directly to the device.
- **Memory Mapping:** It is a technique for reading and writing to a device directly from the program, which avoids the overhead of delegating the reads and writes to kernel-level software.

However, the data transfer mechanism the computer can use depends upon the selected data-acquisition device and its bus architecture. For example, while PCI and FireWire devices offer both DMA and interrupt-based transfers, PCMCIA and USB devices use interrupt-based transfers.

The available hard drive is the limiting factor for real-time storage of large amounts of data. Hard drive access time and hard drive fragmentation can significantly reduce the maximum rate at which data can be acquired and streamed to disk. For systems that must acquire high-frequency signals, high speed hard drive with large memory space is needed.

In general, the overall communication speed is directly proportional to

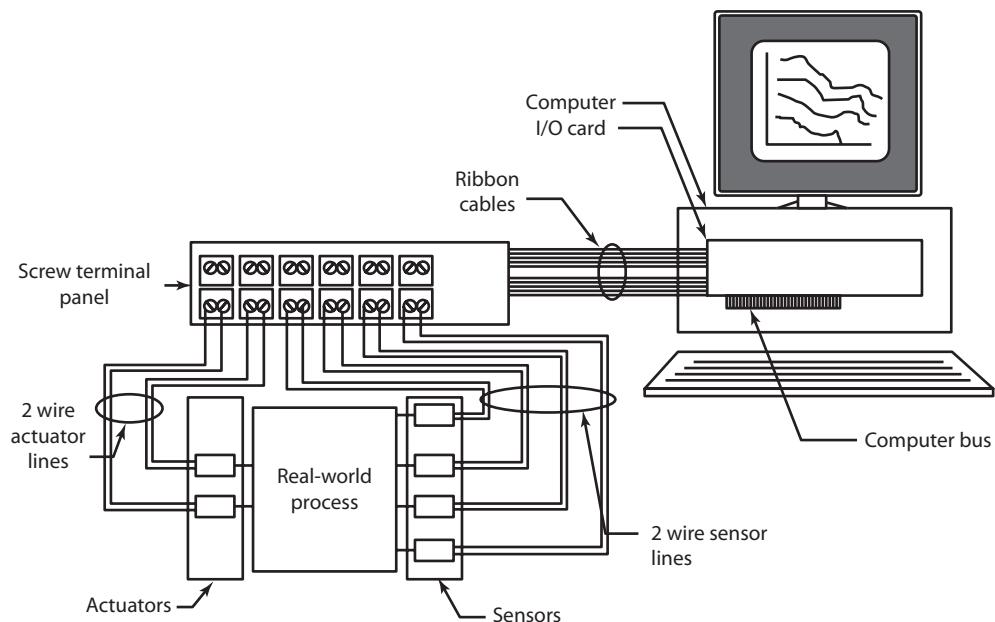
- The clock frequency of the processor chip
- The bit length of the bus (i.e., 8 bit, 16 bit, 32 bit, ...)

and inversely proportional to

- The bit length of the processor (i.e., 16 bit, 32 bit, ...)

Figure 7-2 presents these major components and their interconnections in a four-sensor/two-actuator DAC system. The number of I/Os varies, depending on the manufacturer, as does the functionality. Some screw terminal panels have resistors that can be cut or soldered to change the gain range of a single or group of channels. The screw terminal is attached to the GPIO card via ribbon cables. Figure 7-2 shows two cables; however, the number may vary depending on the type of the card. On some cards four cables are required, two for the analog channels and two for the digital channels. The application softwares function is to provide the engineer with an easy means of reading sensors, writing to actuators, and processing data (plotting, control algorithms, saving data, and data manipulation).

FIGURE 7-2 COMPONENTS OF A DAC SYSTEM AND INTERCONNECTIONS FOR A SYSTEM WITH FOUR SENSORS AND TWO ACTUATORS



7.2.2 General Purpose I/O Card (GPIO)

The necessary ingredients for the general I/O process are the PC and operating system software, general purpose I/O (GPIO) card and software driver, and the proper termination panel(s) and cabling for the GPIO card. The GPIO card is installed into a free expansion slot in the PC bus. Its address is specified both on the card (using micro-switches) and in its driver software.

The termination panel is connected to the GPIO card by one or more cables. At this point, the system is ready to operate, except for the application software.

Whatever application software is selected for the mechatronic programming tasks, it must provide the programmer with the ability to create open-loop as well as closed-loop applications. Most GPIO card manufacturers (including Computer Boards, Inc., Advantech, Data Translation, and Metrabyte) offer their own Windows-based software for controlling their cards. An important limitation of this type of software is that it only works on cards provided by one manufacturer, making multi-card, multi-brand applications impossible.

An alternative to board-manufacturer software is a general Windows application software package designed to work with many GPIO cards of different brands. The success of this approach is evident by the popularity of such well-known packages as LabTech Notebook, LabVIEW, and Snapshot. Most of these packages are aimed at the data acquisition market and, as such, are often not suitable for closed loop applications. Listed packages capable of both open- and closed-loop operation tend to be more oriented towards control systems and include LabVIEW, MATRIXX, Simulink, and VisSim. Table 7-1 presents some of the most popular graphical application software.

TABLE 7-1 POPULAR GRAPHICAL-BASED APPLICATION SOFTWARE

Name	Description
Labtech Notebook	General purpose DAC with analysis
LabWindows	General purpose DAC with analysis
Workbench® PC	General purpose DAC
Snap-Master™	General purpose DAC with analysis and display
EasyEst	General purpose DAC with analysis
Unkel Scope	High speed DA
Snapshot	High speed DA
Acquire	General purpose DAC
LabVIEW	General purpose DAC with analysis
Hyperception	High speed DAC with analysis and display
MATRIXX	High speed DAC with analysis and display
Simulink	High speed DAC with analysis and display
Visual Designer	General purpose DAC with analysis
XANALOG	High speed DAC with analysis and display
VisSim	General purpose DAC with analysis and display

7.2.3 Installation of the I/O Card and Software

Before finalizing the data acquisition card, some thought should be given as to how frequently the I/O must be sampled. Most GPIO cards operate in the 1 to 3 kHz range, which means that you can expect anywhere from 1000 to 3000 samples per second, depending on the card. Application software based on Windows may suffer a slight degradation in this rate, but the difference is small and often can be regained by modifying the algorithm or moving to a more powerful processor. Each GPIO card takes up one slot and provides inputs, outputs, or a combination of both. Most cards come with digital I/O, which can be configured either as input, output, or a combination of both. Analog I/O is harder to come by because it requires a D/A or A/D converter, but most card manufacturers offer many configurations that combine digital and analog I/O.

Once you have selected a manufacturer and identified the potential card(s) which could be used, you must determine the type and precision of the I/O channels needed. I/O channels may be either inputs or outputs and are classified based on the type of data transferred. This leads to three channel types:

- Analog
- Digital
- Frequency (counter timer)

Precision of the I/O channel pertains to the accuracy and transient characteristics of the D/A and A/D converter employed. Accuracy of the converter is a function of bit length. Most converters have 12-bit resolution, which is ample for most applications; however, this depends on your application and should be a consideration prior to purchasing. A list of some of some popular I/O cards is presented in the Appendix.

In mechatronic applications, the I/O typically operates in either of two modes, open loop or closed loop. Open-loop operation exists when inputs are read and/or outputs are written, but no relationship or dependency exists between the inputs and outputs.

EXAMPLE 7.1

A manufacturer of compressors may use the open-loop method for automatic compressor testing. A time-based sequence is often employed. At time t_1 , a signal is sent to the compressor motor which resets the motor speed to the specified value. Since this signal is sent *from* the computer, it is called an *output*. After waiting a preselected time, T seconds, for the equipment (motor and compressor) to hopefully stabilize, the pressure, temperature, flow, and speed are read and saved in memory. Because these four signals are read *into* the computer they are called *inputs*. The process then is repeated with a new motor speed command. This approach is called open loop because the output signal (motor speed command) occurs at T second intervals regardless of the values of the input signals. Had stabilization not been reached in the T second period, it would not be known. Also note that this process reduces to a classic data logger if the output signal is omitted.

In a closed-loop operation, the output signal(s) are dependent on the input signal(s). The same example can be made to operate in closed loop by changing the time-based logic on when a reading is taken to event-based logic—the event being stabilization of the four input signals. Clearly, this requires a little more programming effort, because the time differential of each of the four signals must be computed and combined in such a way that the read input command is initiated at the instant when all four at the derivatives are less than a prespecified magnitude. Aside from the extra programming, the closed-loop approach will automatically vary the time between measurements—each measurement is taken at a precisely controlled operating condition, which is something that the open-loop approach cannot claim.

In mechatronic applications, the I/O typically operates by either, an open-loop or closed loop operation. An open-loop operation exists when inputs are read and/or outputs are written, but no relationship or dependency exists between the inputs and outputs. In a closed-loop operation, the output signal(s) are dependent on the input signal(s).

7.3 Transducers and Signal Conditioning

Transducers sense physical phenomena and produce electrical signals that the DAQ system measures. For example, thermocouples, resistance temperature detectors (RTDs), thermistors, and IC sensors convert temperature into an analog signal that an analog-to-digital converter (ADC) can

measure. Other examples include strain gauges, flow transducers, and pressure transducers, which measure force, rate of flow, and pressure, respectively. In each case, the electrical signals produced are proportional to the physical parameters they monitor.

The electrical signals generated by the transducers must be optimized for the input range of the DAQ device. Signal conditioning accessories amplify low-level signals and then isolate and filter them for more accurate measurements. In addition, some transducers use voltage or current excitation to generate a voltage output. The analog input specifications give you information on both the capabilities and accuracy of the DAQ product. Basic specifications, which are available on most DAQ products, tell you the number of channels, the sampling rate, the resolution, and the input range.

Amplification The most common type of conditioning is amplification. For example, the low-level thermocouple signals should be amplified to increase the resolution and reduce noise. The signal should be amplified so that the maximum voltage range of the conditioned signal equals the maximum input range of the ADC.

Isolation Another common signal-conditioning application is isolating the transducer signals from the computer for safety purposes. The system being monitored may contain high-voltage transients that could damage the computer without signal conditioning. Isolation ensures that the readings from the plug-in DAQ device are unaffected by differences in ground potentials or common-mode voltages. When the DAQ device input and the signal being acquired are each referenced to “ground,” problems occur if there is a potential difference in the two grounds. This difference can lead to ground loop, which may cause inaccurate representation of the acquired signal. If the difference is too large, it may damage the measurement system. Using isolated signal-conditioning modules eliminates ground loops and ensures that the signals are accurately acquired.

Multiplexing A common technique for measuring several signals with a single measuring device is multiplexing. Signal-conditioning hardware for analog signals often provides multiplexing for use with slowly changing signals like temperature. The ADC samples one channel, switches to the next channel, samples it, switches to the next channel, and so on. Because the same ADC samples many channels instead of one, the effective sampling rate of each individual channel is inversely proportional to the number of channels sampled.

Filtering The purpose of a filter is to remove unwanted signals from the signal that you are trying to measure. A noise filter is used on DC-class signals, such as temperature, to attenuate higher frequency signals that can reduce the measurement accuracy. If the noise signals were not removed, they would erroneously appear as signals within the input bandwidth of the device. AC-class signals, such as vibration, often require a different type of filter known as an anti-aliasing filter. Like the noise filter, the anti-aliasing filter is also a low-pass filter; however, it requires a very steep cut-off rate, so it almost completely removes all signal frequencies that are higher than the input bandwidth of the device.

Excitation Signal conditioning also generates excitation for some transducers. Strain gauges, thermistors, and RTDs (for example) require external voltage or current excitation signals. Signal-conditioning modules for these transducers usually provide these signals. RTD measurements are usually made with a current source that converts the variation in resistance to a measurable voltage. Strain gauges, which are very low-resistance devices, typically are used in a Wheatstone bridge configuration with a voltage excitation source.

Linearization Another common signal-conditioning function is linearization. Many transducers, such as thermocouples, have a nonlinear response to changes in the phenomena being measured. Several software systems are available, including linearization routines for thermocouples, strain gauges, and RTDs.

Number of Channels The number of analog channel inputs is specified for both single-ended and differential inputs for devices with both input types. Single-ended inputs are all referenced to a common ground reference. These inputs typically are used when the input signals are high level (greater than 1 V), the leads from the signal source to the analog input hardware are short (less than 15 ft), and all input signals share a common ground reference. If the signals do not meet these criteria, you should use differential inputs. With differential inputs, each input has its own ground reference; noise errors are reduced because the common-mode noise is canceled out.

Digital I/O Digital I/O interfaces are often used on PC DAQ systems to control processes, generate patterns for testing, and communicate with peripheral equipment. In each case, the important parameters include the number of digital lines available, the rate at which you can accept and process digital data on these lines, and the drive capability of the lines. If the digital lines are used for controlling events (such as turning on and off heaters, motors, or lights), a high data rate is usually not required because the equipment cannot respond very quickly. The number of digital lines, of course, should match the number of processes to be controlled. In each of these examples, the amount of current required to turn the devices ON and OFF must be less than the available drive current from the device.

A common Digital I/O application is to transfer data between a computer and equipment such as data loggers, data processors, and printers. Because this equipment usually transfers data in one byte (8-bit) increments, the digital lines on a plug-in DIO device are arranged in groups of eight. In addition, some boards with digital capabilities will have handshaking circuitry for communication-synchronization purposes. The number of channels, data rate, and handshaking capabilities are all-important specifications that should be understood and matched to the application needs.

7.4 Devices for Data Conversion

In this section, the basic principles of the operational amplifiers are discussed followed by some of the basic circuits that are used in signal conditioning. Major components reviewed in this section are amplifiers, zero and span circuits, and A/D and D/A converters.

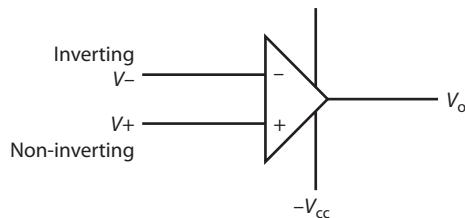
Let us consider an example of a real-time mechatronic system. It is assumed that the output power to be controlled is of a continuous nature and measured with an analog-type transducer. This information is passed to the analog-to-digital (A/D) component. The analog signal is converted into a binary format in order to be compatible with the data acquisition system. The output is determined based on the model of the process and the control algorithm. If an electro-hydraulic actuator does the final actuation, an electro-hydraulic servo valve at the interface modulates the fluid power variable and produces the desired output at the fluid power actuator-load interface. In this case, integration of electronic, mechanical, and fluid power components are necessary for successful implementation of a computer-controlled system.

7.4.1 Operational and Instrumentation Amplifiers

Operational Amplifiers An amplifier is an electronic device which increases the size of voltage or current signal without altering the signal's basic characteristics. Operational amplifiers (op-amps) are termed *operational* because of their early growth in a computing area—mainly to

perform mathematical operations. The operational amplifier has emerged as a basic building block in electronic circuits and is made up of many transistors fabricated on a processor chip. Its use arises from its ability to provide accurate and stable results under a wide range of operational conditions. Figure 7-3 is the schematic symbol for an operational amplifier showing its principle terminals. It is also characterized by very high gain, high input impedance, and very low output impedance. Without feedback, the amplifier would be unstable because of its excessive gain.

FIGURE 7-3 OPERATIONAL AMPLIFIER SCHEMATIC SYMBOL



The operational amplifier has three terminals known as the inverting input, non-inverting input, and the output. Depending on the magnitude of the components and the configuration used, different characteristics are obtained.

Referring to Figure 7-3, the terminal identification is

- + V_{cc} and - V_{cc} = power supply usually with magnitudes in the range of 10 to 15 V
- V_o = output voltage terminal
- V_+ and V_- = input voltages

Thus,

$$A = \text{voltage gain} = \frac{V_{out}}{V_{in}} = \frac{V_o}{(V_+ - V_-)} \rightarrow \infty$$

R_o = output resistance (≈ 0)

R_i = input resistance ($\approx \infty$)

Basic characteristics of operational amplifier are

- (-) Inverting input terminal; A voltage applied to this terminal is amplified with an 180° phase shift.
- (+) Non-inverting terminal; A voltage applied to this terminal is amplified without a phase shift.
- The voltage gain is so great that the voltage difference between the inputs is zero
- The input impedance is so great that the inputs draw no current.
- Zero output impedance.

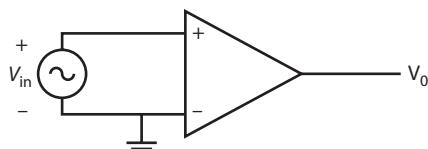
From these parameters, the operating characteristics of the device are determined. For example, the infinite input resistance implies a zero input current at either input terminals:

$$I_+ = I_- = 0$$

Also, the zero output resistance implies that, regardless of the load on the output terminal, there will not be any loss in the output signal.

To study the effect of the infinite gain on operational amplifier, consider the circuit shown in Figure 7-4. Here the input voltage, V_{in} is actually equal to the difference between V_+ and V_- , between the input terminals. If V_{in} is varied over a wide range, the output voltage changes, as in Figure 7-4.

FIGURE 7-4 TEST CIRCUIT FOR OP-AMP

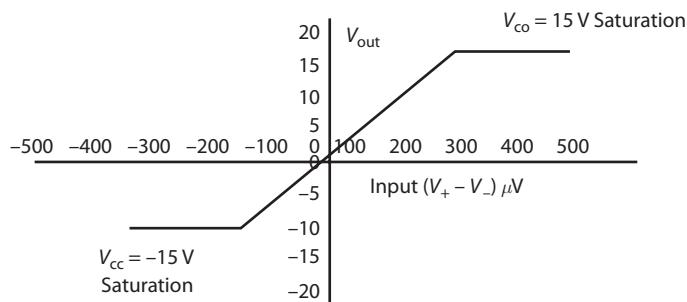


In this example, a supply of ± 15 V and an op-amp with a voltage gain equal to 200,000 is assumed. From the output characteristics in Figure 7-5, we see the output voltage never exceeds the power supply ± 15 V. When operating in this region, the device is saturated. However, in the region between the two horizontal lines in the figure, the characteristics are very linear and have a slope equal to the gain (in this case $200,000 \frac{V}{V}$). The main consideration here is the magnitude of input voltage required to drive the operational amplifier into saturation, namely, either $+75$ or $-75 \mu\text{V}$. In other words, a voltage difference of 0.000075 V at the input terminals will cause the device to saturate. A very useful fact about operational amplifiers operating in their linear region is that the difference in potential across the input terminals approaches zero (due to the higher gain).

$$(V_+ - V_-) = 0; \therefore V_+ = V_-$$

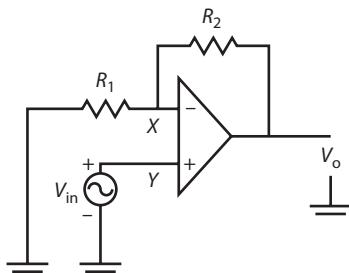
Basic op-amp circuits for signal conditioning purposes are discussed next.

FIGURE 7-5 OP-AMP OUTPUT CHARACTERISTICS



Basic Op-Amp Circuits As an example of a basic op-amp circuit, let us consider the non-inverting amplifier shown in Figure 7-6. It should be noted that the amplifier is non-inverting, because the sign of the output voltage relative to the ground is the same as the input voltage.

FIGURE 7-6 NON-INVERTING AMPLIFIER



The output is considered to be taken from across a potential divider circuit consisting of R_1 and R_2 in series. The voltage V_x is a fraction of the output voltage.

$$V_- = \frac{R_2}{R_1 + R_2} V_o \quad (7-1)$$

Since there is no current through the operational amplifier between the two inputs, there is no potential difference between them. Therefore,

$$V_x = V_{in} \quad (7-2)$$

Equating Equations 7-1 and 7-2 and solving for the output voltage yields

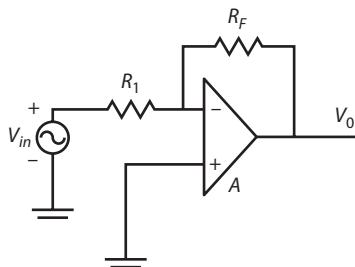
$$V_o = \left(1 + \frac{R_1}{R_2}\right) V_{in} \quad (7-3)$$

Equation 7-3 may be expressed as voltage gain:

$$\frac{V_o}{V_{in}} = \left(1 + \frac{R_2}{R_1}\right) \quad (7-4)$$

It can be noted that the closed-loop gain of the non-inverting amplifier depends only on the external resistors R_1 and R_2 . Introducing the negative feedback through the resistors has reduced the overall closed-loop gain. While this feedback has reduced the gain, it has made the amplifier independent from the operation amplifier itself.

Another basic op-amp circuit is the inverting amplifier, as shown in Figure 7-7. The signal is applied to the inverting terminal through a resistor R_1 with the non-inverting input connected to the

FIGURE 7-7 INVERTING AMPLIFIER

ground. This results in an output that is out of phase with the input. A feedback path is provided from the output, through the resistor R_F , and to the inverting input. It is to be noted that, since the + input terminal is at ground potential, then the - input terminal also must be at the ground; therefore,

$$V_- = 0$$

Since the voltage at the junction of the resistors R_1 and $R_F = 0$ and the input current to the op-amp is also 0, the currents in the resistors must be equal; therefore,

$$I_1 = -I_2 \quad (7-5)$$

$$\frac{V_{in}}{R_1} = -\frac{V_o}{R_F} \quad (7-6)$$

Solving Equation 7-6 for the output voltage gives

$$V_o = -\frac{R_F}{R_1} V_{in} \quad (7-7)$$

It can be noted that the closed-loop gain of the inverting amplifier depends only on the external resistors R_1 and R_F . Furthermore, the gain is negative (180° phase shift). This is an important difference between the inverting and non-inverting amplifiers. In the non-inverting configuration, the signal is applied to the op-amp input terminal, which has an infinite input resistance, and therefore, no current is drawn from the signal source. On the other hand, in the inverting amplifier, the signal is applied through the resistor combination, and as indicated above, a current is drawn from the source. For best results, it should be driven by a low output resistance source.

This current is

$$I_{in} = \frac{V_{in}}{R_1} \quad (7-8)$$

From this, we may state:

$$R_{in}|_{\text{non-inverting}} = \infty \quad (7-9)$$

$$R_{in}|_{\text{inverting}} = R_1 \quad (7-10)$$

This difference in the input resistance plays an important part in instrumentation amplifiers.

EXAMPLE 7.2

The circuit in Figure 7-7 has $R_1 = 20 \text{ k}\Omega$, $R_F = 100 \text{ k}\Omega$, and $V_{\text{in}} = 0.5 \text{ V}$. Find V_o .

Solution

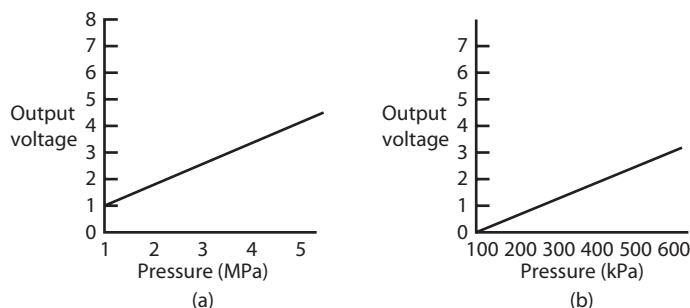
$$V_o = -\frac{R_F}{R_1} V_{\text{in}} = -\frac{100}{20} (0.5) = -2.5 \text{ V}$$

Zero and Span Circuits Depending on specific applications, the output signals of transducers may need to be raised to the level needed. For example, consider a pressure sensor that has a sensitivity of 0.001 V/kPa over a range of 0 to 5000 kPa (abs.) and has an output of 1.2 V at 0 kPa (abs.). If this sensor were used to measure a pressure of 0 to 1000 kPa (abs.) and energize an A/D converter at voltage levels of 0-5 V, then some modification to the signal would be required.

This situation is shown in Figure 7-8. In order to provide the desired signal for the A/D converter, the y -axis intercept and the slope of the sensor output are modified. For considering the slope, the output equation of the inverting amplifier is considered. From Equation 7-7,

$$V_o = -\left(\frac{R_F}{R_1}\right)V_{\text{in}}$$

FIGURE 7-8 (A) TRANSDUCER OUTPUT (B) A/D INPUT

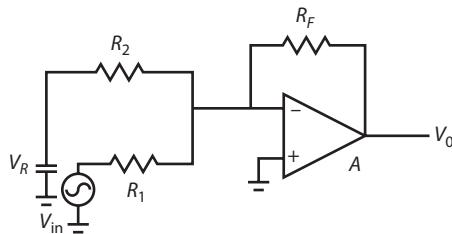


Equation 7-7 is in the form $y = mx + b$, (the equation for a straight line), where the slope is the coefficient of the term on the right side is shown as

$$\text{Slope (span)} = -\left(\frac{R_F}{R_1}\right) \quad (7-11)$$

The b or y -intercept term can be obtained by adding a second input to the inverting amplifier as shown in Figure 7-9. Here V_R is a constant reference voltage equal in magnitude to the y intercept. The output is obtained by considering each input separately and adding them. The result of this calculation is

$$V_o = -\left(\frac{R_F}{R_1}\right)V_{\text{in}} - \left(\frac{R_F}{R_2}\right)V_R \quad (7-12)$$

FIGURE 7-9 MULTIPLE INPUT INVERTING AMPLIFIER

The addition of the inverting amplifier with a gain of -1 will get rid of the negative output. The output expression for this amplifier is

$$V_o = \left(\frac{R_F}{R_1} \right) V_{in} + \left(\frac{R_F}{R_2} \right) V_R \quad (7-13)$$

As an example of a zero/span design, a pressure sensor and A/D converter are shown next. Let V_T be pressure sensor output and V_o = amplifier output. The sensor output is

At $p = 0$, $V_T = 1.2\text{ V}$ and $V_o = 0\text{ V}$

At $p = 1000\text{ kPa (abs.)}$ $V_T = 1.2 + (0.001)1000 = 2.2\text{ V}$ and $V_o = 5\text{ V}$

Consider the slope as per Equation 7-11,

$$\text{Slope} = \frac{\Delta V_o}{\Delta V_T} = \frac{5 - 0}{2.2 - 1.2} = 5 \frac{V}{V}$$

We select $R_F = 10\text{ k}\Omega$, then $R_1 = 2\text{ k}\Omega$ for the amplifier, the zero may be set by applying Equation 7-13:

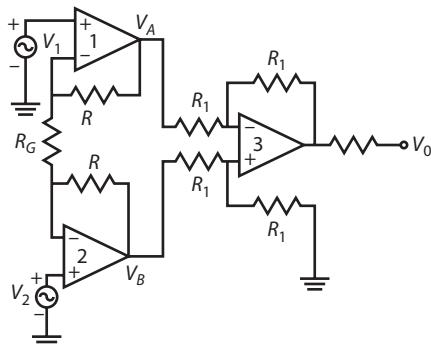
$$V_o = \left(\frac{R_F}{R_1} \right) V_T - \left(\frac{R_F}{R_2} \right) V_R \quad (7-14)$$

Substituting for R_F and R_1 and letting $R_2 = R_F$ yields:

$$V_o = \left(\frac{R_F}{R_1} \right) V_T + \left(\frac{R_F}{R_2} \right) V_R \quad (7-15)$$

At $p = 0$ Equation 7-15 becomes $0 = (5)(1.2) + V_R$ and $V_R = -6\text{ V}$, which completes the design of the amplifier.

Instrumentation Amplifier The output signals of transducers are rarely at the levels required for the job at hand. The desired signal is often achieved using a versatile amplifier known as an instrumentation amplifier. The gain of this amplifier can be precisely set by the addition of a single external resistor, and it has a very high input resistance. It also has the ability to amplify small signals in the presence of noise. Figure 7-10 shows the schematic of an instrumentation amplifier.

FIGURE 7-10 INSTRUMENTATION AMPLIFIER

Referring to Figure 7-10, op-amp1 and op-amp2 are the input sections of the amplifier. OP-amp3 is a unity gain amplifier in the output section, which converts the output to the single end. The instrumentation amplifier shown in the figure has two inputs, which is useful for applications (such as strain gauge bridge circuits). The resistor R_G is the external gain setting resistor. Letting $V_1 - V_2 = \Delta V_{in}$ and $V_A - V_B = \Delta V_o$ we can proceed with the following analysis:

Since the basic relation $V_+ = V_-$ holds, the current through the gain setting resistor, R_G , is

$$I_G = \frac{V_{A1} - V_{B2}}{R_G} = \frac{\Delta V_{in}}{R_G} \quad (7-16)$$

The input currents to the op-amps are zero; I_G must flow through both resistors R and R_G ; therefore,

$$\Delta V_o = IR = \frac{\Delta V_{in}}{R_G} (2R + R_G) = \left(1 + \frac{2R}{R_G}\right) \Delta V_{in} \quad (7-17)$$

This expression may be solved for the gain:

$$A = \left(1 + \frac{2R}{R_G}\right) \quad (7-18)$$

In practice, the designer selects the value for R and the external resistor R_G to suit the gain requirement. In addition to providing the gain necessary for a given application, most commercial instrumentation amplifiers also allow for zero setting. The characteristics of the instrumentation amplifier are high input resistance and their ability to amplify small signals in the presence of noise.

Amplifiers are susceptible to errors, such as nonlinearity errors, hysteresis errors and thermal stability errors. The three-op-amp configuration for instrumentation amplifiers is a popular design. Commercial units with these characteristics are available in the form of monolithic, modular single IC chips. Some models of instrumentation amplifiers are provided with power supply and digitally programmable resistor network units representing R_G . Since the gain of these units can be changed in conjunction with data acquisition systems, they are known as *programmable gain instrumentation amplifiers*.

7.5 Data Conversion Process

Data acquisition systems in the real-time environment, especially using microcomputers, has become an important need in many automated manufacturing situations. If it is required to monitor several pieces of data at precisely the same time, sample and hold devices are necessary. A sample and hold device is used to hold each sampled value of the signal until the next pulse occurs.

In many mechatronic applications, the sensor signal is a complex, time-varying voltage that is considered a combination of many sine and cosine waves of different frequencies and amplitudes. A process of filtering eliminates some of the frequencies. Some of the sources of noise are the following.

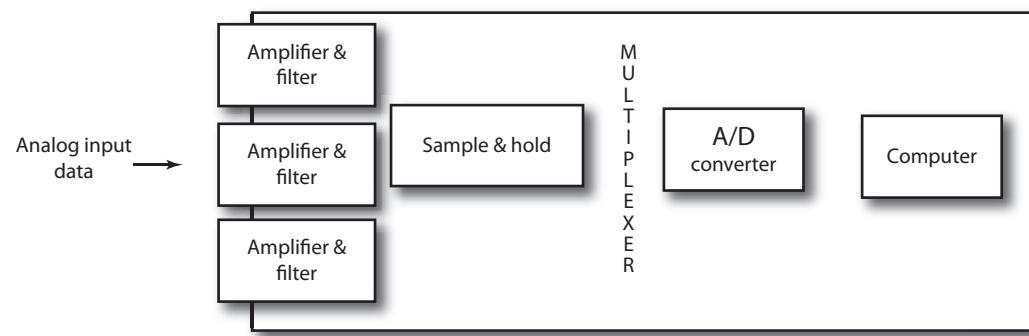
- Noises from the actuator motion, example being commutator brushes and relay contacts.
- Noise due to electromagnetic interference of transmission lines.
- Cross talk.

Filtering is used to remove certain band of frequencies from a signal, permitting others to be transmitted. A low-pass filter has a band, which allows all frequencies from zero to a certain level frequencies to be transmitted. A high-pass filter has a pass band, which allows all frequencies from a certain level up to infinity. A bandpass filter attenuates signals at both high and low frequencies and allows a range of frequencies to pass without attenuation.

In several cases, the computer reads the information from several channels one at a time using a device called multiplexer. The multiplexer is a switching device which enables each of the inputs to be sampled in turn. It is a data selector that allows only one of the inputs to get through to the output. If it is necessary to control the experiment, the computer must supply the outputs in digital or analog form. Data acquisition systems require the use of analog-to-digital converters to input the sensed information coming from analog transducers and sensors to computer system. Figure 7-11 shows a diagram of analog-to-digital conversion using multiplexer, sample-and-hold circuits.

If the needed control signal is analog (mainly for control valves and motors), digital output from the computer has to be processed through a digital-to-analog converter. In a computer-aided process control system, the process control actions are recommended to the operator and displayed on the operator terminal. Automatic closed-loop process control systems require that the sensed data be manipulated to decide the control action, which should be subsequently and automatically

FIGURE 7-11 MULTIPLEXER IN AN ANALOG-TO-DIGITAL CONVERSION ENVIRONMENT



implemented by means of suitable actuators. The majority of the actuators (other than stepping motors and relays) require analog voltages and currents for their operation. It is therefore necessary to convert the digital binary output from the microcomputer into an analog format.

The real-time interface handles the required digital data from the A/D or D/A converters. The control program is designed upon an overall analysis of the closed-loop control system. This control algorithm is executed during each sampling interval. If the bandwidth of the system is relatively high and multiple signals are sampled and converted, then efficient interface software is necessary.

Typical control programs can be generated in higher-level languages, in assembly level languages and in visual simulation languages. Assembly languages are useful in some instances where execution time is small and critical. Visual simulation procedures are used for situations where real-time data acquisition and control are done simultaneously using process models and sensor data. In these cases, block diagram-based models of manufacturing or industrial process are constructed and used in an atmosphere of hardware and software in one loop.

7.5.1 The Analog-to-Digital Converter

Although most of the sensors provide a direct signal output, a large number of transducers convert a dynamic variable into an analog electric signal. It is necessary to use an analog-to-digital converter (ADC) to transform an analog voltage into a binary number through a process called quantization.

Quantization Quantization is a process of taking continuous analog signals and breaking it into a number of discrete steps. The conversion is discrete and takes place one at a time. The A/D converter has two sides: one is the analog side and the other is the digital side. The analog-side specification includes a full-scale reference voltage range (V_R). The A/D converter will operate under this range. The digital side is specified in terms of the number of bits of its register. A n -bit A/D converter will output n -bit binary numbers. The three functions of sampling, quantizing, and encoding are involved in this process, which are shown in Figure 7-12.

FIGURE 7-12 FUNCTIONS OF ANALOG-TO-DIGITAL CONVERSION



The conversion process involves:

- Sampling the continuous signal.
- Storing this voltage, and, before the next sample is taken.
- Converting the stored number to the binary number, which typically consists of a n -bit binary output word length.

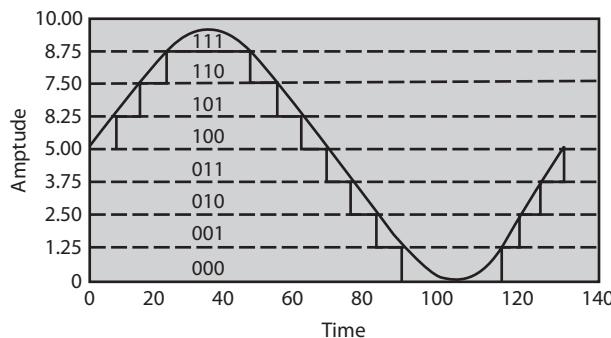
The quantization of a sampled analog signal involves the assignment of a finite number of amplitude levels corresponding to discrete values of input signals V_X between the range of 0 and the full-scale value of V_R . If the A/D converter has a range of 10 bits, 1024 (2^{10}) different values of the input voltage can be represented. A typical voltage range for these systems is 10 V. Therefore, the resolution would be $\frac{10}{1024}$ or about 10 mV, which yields about 0.1% accuracy.

Sampling Rate This parameter determines how often conversions can take place. A faster sampling rate acquires more data in a given time and therefore often can form a better representation of the original signal.

Resolution The number of bits that the ADC uses to represent the analog signal is the resolution. The higher the resolution, the larger the number of divisions the range is broken into and, therefore, the smaller the detectable voltage change.

Figure 7-13 shows a sine wave and its corresponding digital image as obtained by an ideal 3-bit ADC. A 3-bit converter (as a simple example) divides the analog range into 2^3 , or eight divisions. Each division is represented by a binary code between 000 and 111. The resolution of the A/D converter is the number of bits used to digitally approximate the analog value of the input. The number of possible states is equal to the number of bit combinations that results from the converter (which is equal to 2^n , where n is the number of bits). Clearly, the digital representation is not a good representation of the original analog signal, because information has been lost in the conversion. By increasing the resolution to 16 bits, however, the number of codes from the ADC increases from 8 to 65536 to obtain an extremely accurate digital representation of the analog signal.

FIGURE 7-13 DIGITIZED SINE WAVE WITH A RESOLUTION OF THREE BITS



The error that results from the quantization process is called the quantization error. The quantization error can be as large as one-half the quantization level spacing.

$$\text{Quantization error} = \pm \frac{1}{2}$$

As a typical example, in a 8-bit A/D converter, the number of quantization levels are 256 which is (2^8) . For a maximum possible range of the voltage signal 0 to 10 V, the resolution is $\frac{10}{256}$ or 0.0391 V, and the quantization error = 0.0195 V. In the same way, a typical 12-bit A/D converter with a reference voltage of 10 V would be able to represent analog voltages in the range between 0 and 10 V with 2^{12} different binary values.

The main requirements for selecting an A/D converter include resolution, range of voltage needed, and speed of conversion.

$$V_X = V_R (b_1 2^{-1} + b_2 2^{-2} + \dots + b_n 2^{-n}) \quad (7-19)$$

where,

V_X = analog voltage input

V_R = reference voltage

b_1, b_2, \dots, b_n are n -bit digital outputs

Minimum V_X is zero and maximum V_X is determined by the size of binary digits. For a 8-bit word, $V_{\max} = V_R(2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-6} + 2^{-7} + 2^{-8}) = 0.9961 V_R$.

Resolution The conversion resolution is identified smallest possible change as $\Delta V = V_R 2^{-n}$ (approximately) and is the smallest possible output voltage. For example, if 5 bits are used, the reference voltage is 10 V, $\Delta V = 10(2)^{-5} = 0.3125$ V/bit.

EXAMPLE 7.3

Find the binary equivalent word that results from a 6.424 V input to a 5-bit A/D converter with a 10 V reference.

$$V_X = V_R(b_1 2^{-1} + b_2 2^{-2} + \dots + b_n 2^{-n}) = \pm \frac{6.434}{10} = 0.6434$$

Solution

Using the successive approximation method, the results of each multiplication will have a fractional part and whole number part (0 or 1). This determines whether the digit is 0 or 1. The first multiplication gives the most significant digit (MSB) and the last multiplication gives the least significant digit (LSB).

$$0.6434(2) = 1.2868, b_1 = 1$$

$$0.2868(2) = 0.5756, b_2 = 0$$

$$0.5756(2) = 1.1512, b_3 = 1 \quad \text{Output} = 10100_2$$

$$0.1512(2) = 0.3024, b_4 = 0$$

$$0.3024(2) = 0.6048, b_5 = 0$$

EXAMPLE 7.4

Determine how many bits a converter must have to provide output increments of 0.02 volts or less. The reference is 10 V.

$$\Delta V = 0.02 = 10(2)^{-n}$$

Solution

Taking logarithms,

$$\log(0.02) = \log((10(2)^{-n}))$$

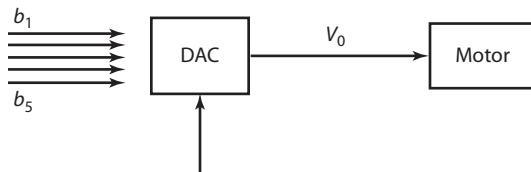
$$\log(0.02) = \log(10) - n \log 2$$

$$n = \frac{\log(10) - \log(0.02)}{\log 2} = 8.96$$

$n = 9$ will be satisfactory. It can be proved as $\Delta V = 10(2)^{-9} = 0.0195$ V.

EXAMPLE 7.5

A motor to be driven by a digital signal has a speed variation of 200 rev/min per volt with minimum rpm at 5 V and maximum at 10 V. Find the minimum speed word, maximum speed word, and speed change for change of 1 bit. Use a 5-bit, 15 V reference, D/A converter, as shown in Figure 7-14.

FIGURE 7-14**Solution**

The minimum speed is at 5 V

$$\begin{aligned}V_0 &= 5 = 15(b_12^{-1} + b_22^{-2} + \dots + b_52^{-5}) \\0.3333 &= (b_12^{-1} + b_22^{-2} + \dots + b_52^{-5})\end{aligned}$$

By using the successive approximation method,

$$\begin{aligned}b &= 0 \\b_2 &= 1 \\b_3 &= 0 \quad \text{Output for minimum speed} = 01010_2 \\b_4 &= 1 \\b_5 &= 0\end{aligned}$$

The maximum speed is at 10 V

$$V_0 = 10 = 15(b_12^{-1} + b_22^{-2} + \dots + b_52^{-5})$$

By using the successive approximation method,

$$\text{Output for maximum speed} = 10101$$

$$\text{Voltage change for change of 1 bit} = \Delta V = V_{\text{ref}}(2)^{-n} = 15(2)^{-5} = 0.469 \text{ V}$$

$$\text{Speed change for change of 1 bit} = \Delta S = (0.469 \text{ V})(200 \text{ rpm/V}) = 93.8 \text{ rpm}$$

The sampling interval, ΔT , is determined primarily by the bandwidth of the system being controlled. In typical industrial actuation systems where inertial loads are relatively large, typical bandwidths of 10 to 100 Hz are common. Using a sampling theorem, a sample time, T , of the order of 0.01 to 0.001 s is required. Commercially available A/D converters today have sampling rates up to several hundred megahertz.

Range Range refers to the minimum and maximum voltage levels that the ADC can quantize. DAQ devices offer selectable ranges, so the device is configurable to handle a variety of voltage levels. With this flexibility, you can match the signal range to that of the ADC to take advantage of the available measurement resolution.

Code Width The range, resolution, and gain available on a DAQ device determine the smallest detectable change in voltage. This change in voltage represents one least significant bit (LSB) of the digital value and is often called the *code width*. The ideal code width is found by dividing the voltage range by the gain times two raised to the order of bits in the resolution.

Critical Considerations of Analog Inputs Although the basic specifications may show that a DAQ device has a 16-bit resolution ADC and a 100 kHz sampling rate, one may not sample at full speed on all 16 channels and get full 16-bit accuracy. For example, some products on the market today with 16-bit ADCs get less than 12 bits of useful data. While evaluating DAQ products, also consider the differential nonlinearity, relative accuracy, settling time of the instrumentation amplifier, and noise.

Settling Time Settling time is the time required for an amplifier, relays, or other circuits to reach a stable mode of operation. The instrumentation amplifier most likely will not settle when you are sampling several channels at high gains and high rates. Under such conditions, the instrumentation amplifier has difficulty tracking large voltage differences that can occur as the multiplexer switches between input signals. Typically, the higher the gain and the faster the channel switching time, the less likely the instrumentation amplifier is to settle.

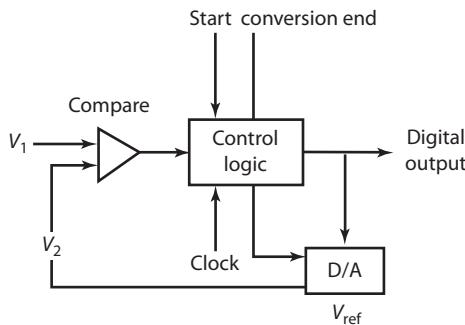
Noise Any unwanted signal that appears in the digitized signal of the DAQ device is noise. Because the PC is a noisy digital environment, acquiring data on a plug-in device takes a careful layout on multiple-layer DAQ devices by skilled analog designers. Simply placing an ADC, instrumentation amplifier, and bus interface circuitry on a one or two-layer board will likely result in a very noisy device. Designers can use metal shielding on a DAQ device to help reduce noise. Proper shielding not only should be added around sensitive analog sections on a DAQ device but also must be built into the layers of the device with ground planes.

Triggers Many DAQ applications need to start or stop a DAQ operation based on an external event. Digital triggers synchronize the acquisition and voltage generation to an external digital pulse. Analog triggers, used primarily in analog input operations, start or stop the DAQ operation when an input signal reaches a specified analog voltage level and slope polarity.

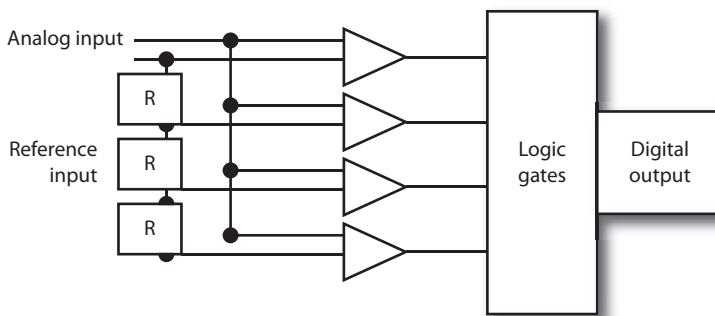
Another important consideration is the number of analog input channels. In a multichannel A/D system, the processor is programmed to determine which analog channel to convert next in sequence. When several analog input channels are to be converted, the A/D converter is shared, or multiplexed, by sequentially sampling and converting the voltage on each of the channels.

7.5.2 Successive Approximation Type of A/D Converter

The successive approximation type of A/D converter (as displayed in Figure 7-15) is a popular type of A/D converter. It uses a trial-and-error approach to estimate the input voltage to the A/D, which needs conversion. The pulses generated by the output of the clock are counted in a binary manner. A D/A converter changes it to analog voltage. A voltage comparator compares the clock-generated voltage to the input analog voltage of the sensor. In this type of A/D converter, a series of known analog voltages are generated and compared to the input voltage. When the analog input from the sensor equals the clock-generated voltage, the pulses from the clock are stopped. The counter output represents the digital equivalent of the analog input from the sensor.

FIGURE 7-15 A/D CONVERSION (SUCCESSIVE APPROXIMATION CONVERSION)

Flash type analog-to-digital converters (as shown in Figure 7-16) are fast. The converter consists of a set of input comparators that act in parallel with each having an analog input voltage as one input. The output of the latches is in digital form. The code converter consists of combinational logic circuits. For analog input voltage, all of those comparators for which the analog voltage is greater than the reference voltage will provide high output, and those for which it is less will provide low output.

FIGURE 7-16 A/D CONVERSION (FLASH CONVERTER)

Digital-to-analog converters reconstruct the digital signals into continuous-time analog signals for the purpose of actuation and display. Some of the devices used in the mechatronic hardware area are analog devices (such as solenoids or valves). To actuate these devices from a computer, the computer output signals have to be converted to analog signals. A digital-to-analog converter is similar to a digitally controlled potentiometer that is calibrated in the range of operation. The digital-to-analog converters typically consist of a precise reference voltage, a weighed resistor network comprised of switches that can be closed or opened in response to changes in the digital code for the word, and an operational amplifier.

Digital-to-analog conversion using software techniques involves the generation of a series of pulses by the microcomputer, which represents the digital information. The pulses are then applied

to a resistor-capacitor network, which converts the digital data into an averaged DC signal. This software digital-to-analog conversion technique has to be specially designed for high-speed applications. Monolithic single-chip digital-to-analog converters, which use hardware to carry out high-speed conversion, are also used in mechatronic devices.

7.6 Application Software

Application software is required for the DAQ hardware to work in coordination with the computer. The application software would either register directly from the DAQ hardware or a low-level software driver packaged with the DAQ hardware develops the higher level applications to register data coming from the hardware. Also, there are some off-the-shelf application softwares (like MATLAB, LabVIEW, VisSIM, etc.) that provide an interface for programming means to the following.

- Acquire data at required sampling rate.
- Analyze and display the acquired data.
- Stream data to and from disk.
- Integrate different DAQ boards in a computer and use various functions of a DAQ board from a single user interface.

The application software discussed in this book are LabVIEW and VisSim. These are the popular platforms for simulation and data acquisition, and control. Laboratory examples are discussed in this chapter and industrial case studies are presented in Chapter 8.

7.6.1 LabVIEW Environment

LabVIEW, short for Laboratory Virtual Instrument Engineering Workbench, is a graphical programming language that uses icons instead of lines of text to create applications. Unlike text-based programming languages like C, Pascal, or BASIC, where instructions determine program execution, LabVIEW uses dataflow programming, where the flow of data determines execution order.

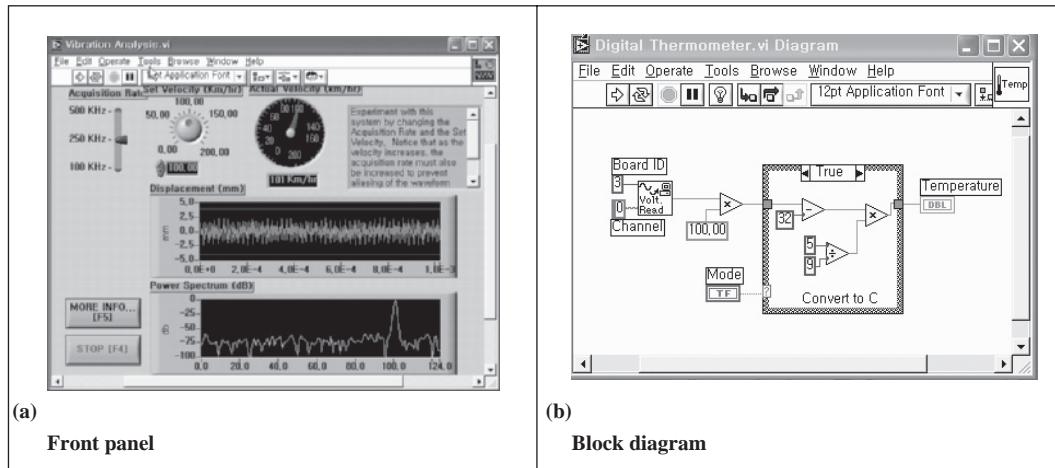
With LabVIEW, one can develop a program in a graphical environment with an effective user interface. It also simplifies the task of interfacing computers with the instruments, thereby providing an easy means for collecting, storing, analyzing, and transmitting data.

LabVIEW can work on PCs running Microsoft Windows, Mac OS, Sun SPARCstations, and HP 9000/700 series workstations. The programs created are independent of the type of machine that they are created for, so programs can be transferred between different operating systems.

Virtual Instruments (VI) LabVIEW programs are also known as virtual instruments (VIs). VIs have three main parts: the front panel, the block diagram, and the icon connector (Figure 7-17).

Front Panel The front panel provides the user interface with the VI. It is analogous to a front panel of a real instrument. Through the front panel, the user can control the program, change inputs, and see the results (outputs). In the LabVIEW environment, inputs are called *controls*, and outputs are called *indicators*.

Controls simulate instrument input mechanisms and supply data to the block diagram of the VI. Indicators simulate instrument output mechanisms and display data the block diagram acquires or generates. The front panel window has an extensive library of controls and indicators. Controls include knobs, push buttons, dials, and other input mechanisms. Indicators include graphs, LEDs, and other

FIGURE 7-17 WEIGHT MEASUREMENT SYSTEM

output displays. Every front panel control or indicator has corresponding terminal on the block diagram.

Block Diagram The block diagram contains the source code also known as G code or block diagram code, which are created using graphical representations of functions to control the front panel objects. The block diagram window has an extensive library of functions. The controls, functions, and indicators are wired together to define the data flow.

The primary block diagram objects include the following.

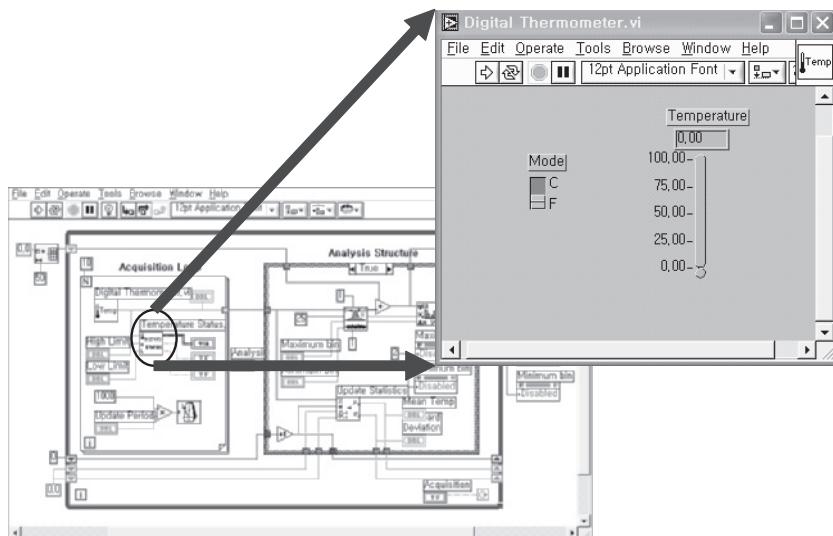
Terminals The terminals represent the data type for the control or indicator. By default, front panel objects appear as icon terminals, but one can also configure front panel controls or indicators to appear as an icon on the block diagram. Terminals are entry and exit ports that exchange information between the front panel and block diagram.

Nodes Nodes are objects on the block diagram that have inputs and/or outputs and perform operations when a VI runs. They are analogous to statements, operators, functions, and subroutines in text-based programming languages. The *Add* and *Subtract* functions are examples of nodes.

Wires The data is transferred among block diagram objects through wires. Each wire has a single data source, but one can wire it to many VIs and functions that read the data.

Structures Structures are graphical representations of the loops and case statements of text-based programming languages. They are used on the block diagram to repeat blocks of code and to execute code conditionally or in a specific order.

Icon/Connector The icon/connector is used to turn the VI into an object that can be used as a subroutine. The icon graphically represents the VI in the block diagrams of other VIs (Figure 7-18).

FIGURE 7-18 BLOCK DIAGRAM WITH ICON/CONNECTOR

Toolbar Both the front panel and block diagram windows contain a toolbar of command buttons and status indicators that one uses to control the VI.

Front Panel Toolbar:

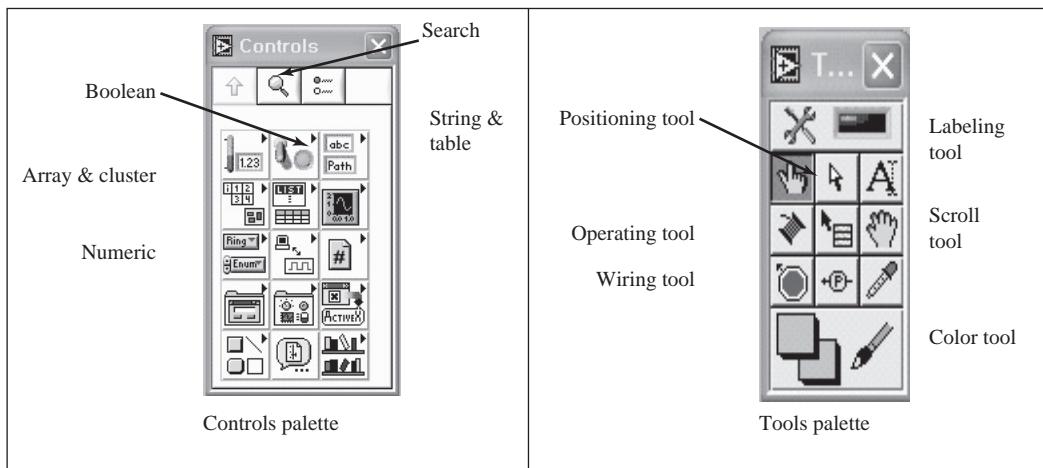
- *Run Button*—Click to run the VI.
- *Broken Run*—Replaces *Run Button* and indicates that VI has errors. Click this button to list the errors.
- *Continuous Run*—Click on it to execute the VI repeatedly.
- *Abort Execution*—This button appears while VI is executing and is used to stop the VI execution.
- *Pause Button*—This button halts the execution. It should be pressed again for the VI to continue.

Block Diagram Toolbars: These have additional items in the toolbar other than the items discussed previously.

- *Execution Highlighting*—For debugging to see the flow of data.
- *Step Into*—Helps to step into a loop, structure, function or node.
- *Step Out*—Helps to step out of a loop, sub VI, or other block-diagram loop node.

Palettes LabVIEW uses graphical and floating palettes to aid in the creation and operation of VIs. The three palettes include the *Tools*, *Controls*, and *Functions* palettes. You can create, modify, and debug VIs using the Tools located in the floating *Tools* palette.

Tools Palette (Figure 7-19(b))

FIGURE 7-19 CONTROLS AND TOOLS PALETTE

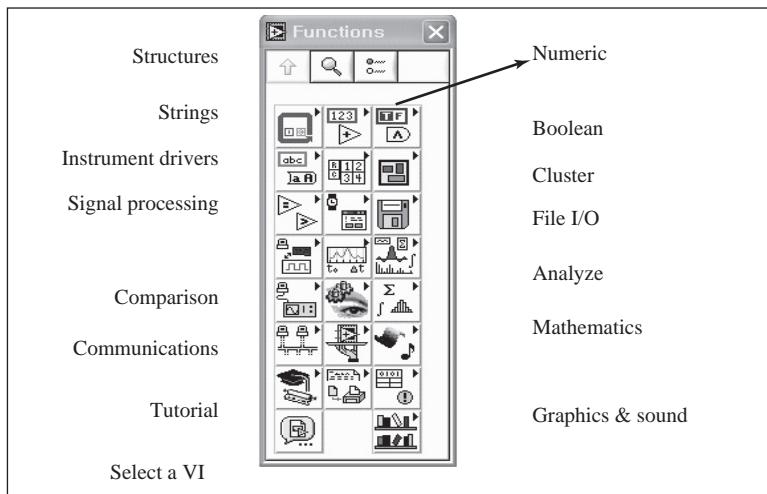
- *Operating tool*—*Operating tool* to manipulate values of front panel controls.
- *Positioning tool*—Use the *Position tool* to select, move, resize objects.
- *Labeling tool*—Use it to enter text into labels.
- *Wiring tool*—To wire objects together in block diagram.
- *Scrolling tool*—Scroll through windows.
- *Breakpoint tool*—Use *Breakpoint tool* to set breakpoints on sub VIs, wires, functions.

Controls Palette (Figure 7-19(a))

- Numeric sub-palette—Consists of controls and indicators.
- Boolean sub-palette—Controls and indicators for boolean.
- String & path—Consists of controls and indicators for strings and paths.
- Array & cluster—Consists of controls and indicators that group set of data types.
- List & table—Controls and indicators for list boxes and tables.
- I/O sub-palettes—Consists of controls and indicators for accessing instruments and data acquisition hardware connected to the system.

Functions Palette (Figure 7-20) Block diagrams are built with the *Function palette* and are available with an active block diagram window.

- Structures sub-palette—Consists of program control structures for looping and sequencing.

FIGURE 7-20 FUNCTION PALETTE

- String sub-palette—Consists of functions to create and manipulate strings.
- String—Consists of functions to create and manipulate strings.
- Array sub-palette—Consists of functions to create and process arrays.
- Cluster sub-palette—Consists of functions to create and manipulate clusters.
- Time and Dialog—Consists of controls structures for looping, sequencing, etc.
- String sub-palette—Consists of functions for dialog windows, timing, and error handling.
- Data acquisition—VIs for performing analog/digital I/O signal conditioning.

LabVIEW has extensive libraries of functions and subroutines to help with most programming tasks without the use of pointers, memory allocation, and other programming problems found in conventional programming languages. LabVIEW also contains application-specific libraries of code for data acquisition, General Purpose Interface Bus (GPIB) and serial instrument control, data analysis, presentation, and storage. The analysis library contains functions including signal generation, signal processing, filters, windows, statistics, regression, linear algebra, and array arithmetic. Because of its graphical nature, the output appears in a form such as charts, graphs, and user-defined graphics.

Procedure

Selecting Objects:

- Click the left mouse button while the positioning tool is over the object.
- To select more than one object, $<\text{shift}> + \text{click}$.
- To select multiple objects, click and drag the dashed box surrounding the desired objects.

Moving Objects:

- Click on object with the positioning tool and drag it to desired location.
- Select object and press arrow key $<Shift>$ +arrow key to move the object faster.
- To move groups of objects, select the objects and drag it to desired location.

Deleting Objects:

- Select the object(s) and choose Clear from the *Edit* menu.
- Select the object(s) and press $<Delete>$ key.

Duplicating Objects:

- Select Copy and Paste from *Edit* menu.
- Select the object and $<ctrl>+drag$ the object to its new location.
- Select the object and drag and drop it between VIs.

Labeling Objects:

- There are free label 3 and owned label 3.
- An owned label belongs to and move with a particular object.
- A Free label is not attached to any object. You can create, move, or dispose of it independently.
- Create labels using labeling tool.
- Click anywhere, type the desired text, and click the *Enter* button.

Changing Fonts and Sizes:

- Using the text setting in the *Toolbar*, you can change the font, style, and size of the object.
 - Size (9 to 36) smaller: $<ctrl>+<->$ larger: $<ctrl>+<=>$
 - Style: plain, bold, italic, etc.
 - Justify: left, center, right.
 - Color.
 - Font: Times New Roman, etc.

Positioning and Deleting:

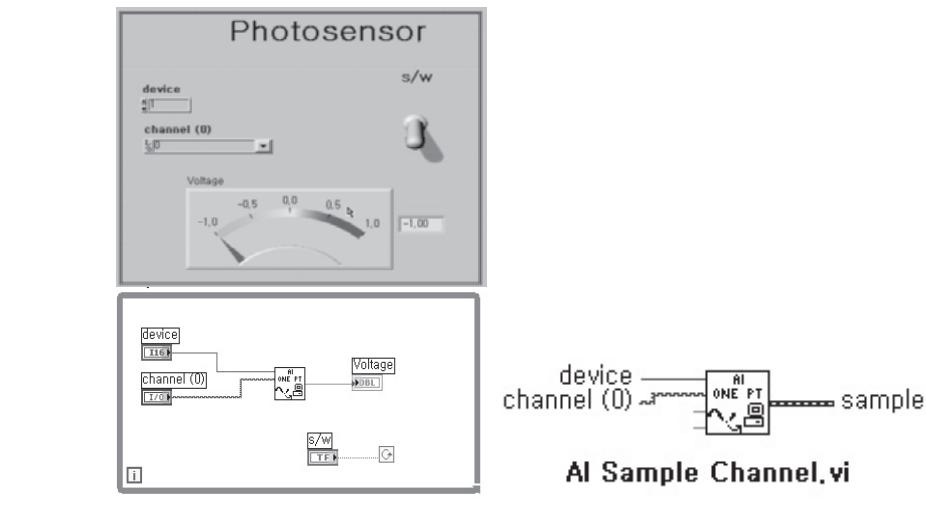
- Select the wire segment by clicking with positioning tool.
- To delete wires, select the broken wires.
 - press $$ key.
- Select Remove Bad Wires from the *Edit* menu.

7.6.2 LabView Applications

EXAMPLE 7.6 Analog Input from a Photo Sensor

This example looks at the procedure of taking an input from a photo sensor as shown in Figure 7-21.

FIGURE 7-21 FRONT PANEL OF THE TEMPERATURE SENSOR (EXAMPLE 7.9)



Solution

The voltage produced by the single point photo sensor is processed by a DAQ device. A single-point analog input reads the value from the input channel. The *Analog Input* palette found on the *Data Acquisition* palette contains the VIs that perform the single-point acquisitions. The AI Sample Channel VI is used for single-point acquisition; that is, it takes a single sample of the analog signal attached to a specified channel and returns the measured voltage. The two inputs for the VI are (1) device with the device number of the DAQ and (2) channel specifies the analog input channel number.

EXAMPLE 7.7 Temperature Measurement Using a K-Type Thermocouple

Temperature is a measure of the average kinetic energy of the particles in a sample of matter expressed in units of degrees on a standard scale. One can measure temperature in many different ways that vary in cost of equipment and accuracy. Thermocouples are one of the most common sensors used to measure temperature, because they are relatively inexpensive yet accurate sensors that can operate over a wide range of temperatures.

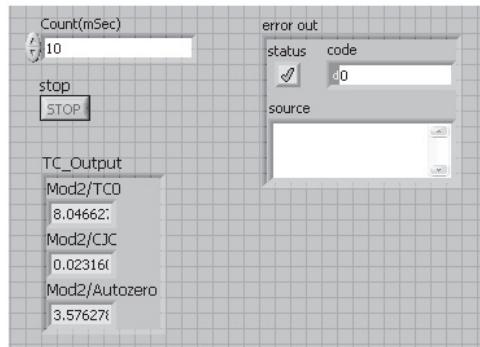
Solution

A thermocouple is created whenever two dissimilar metals touch and the contact point produces a small open-circuit voltage as a function of temperature. This thermoelectric voltage is known as the Seebeck voltage. The voltage is nonlinear with respect to temperature. However, for small changes in temperature, the voltage is approximately linear.

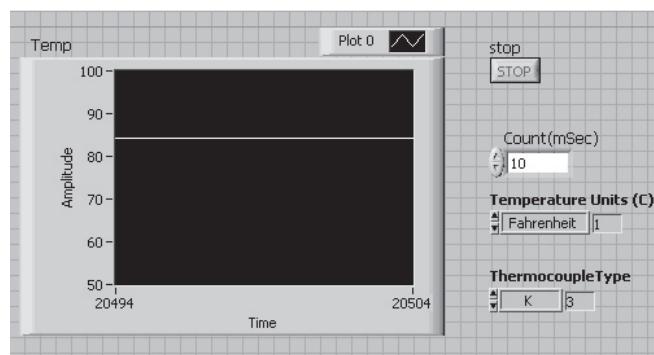
Several types of thermocouples are available, and different types are designated by capital letters which indicate their composition according to the American National Standards Institute (ANSI) conventions. For example, type K (chromel–alumel) is the most common general-purpose thermocouple with a sensitivity of approximately $41 \mu\text{V}/^\circ\text{C}$. It is inexpensive, and a wide variety of probes are available in its -200 to $+1350^\circ\text{C}$ range. Other types of thermocouples include B, E, J, N, R, S, and T. To measure a thermocouple Seebeck voltage, you cannot simply connect the thermocouple to a voltmeter or other measurement system, because connecting the thermocouple wires to the measurement system creates additional thermoelectric circuits. Hence, there is a need for some form of temperature reference to compensate for these unwanted parasitic “cold” junctions. The most common method is to measure the temperature at the reference junction with a direct-reading temperature sensor and subtract the parasitic junction voltage contributions. This process is called cold-junction compensation (CJC).

The example discussed here uses a K-type thermocouple interfaced to the system using NI cRIO 9004 and NI 9211 I/O. NI 9211 I/O has a 10-terminal, detachable screw-terminal connector that provides connections for four thermocouple input channels and has an internal CJC sensor. Figure 7-22 shows the front panel and Figure 7-23 shows the block diagram for the temperature measurement.

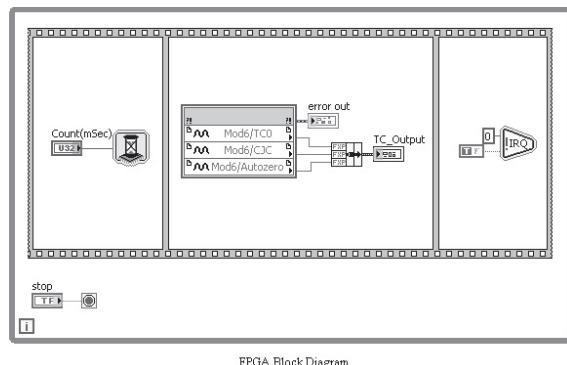
FIGURE 7-22 FRONT PANEL OF THE TEMPERATURE SENSOR (EXAMPLE 7.9)



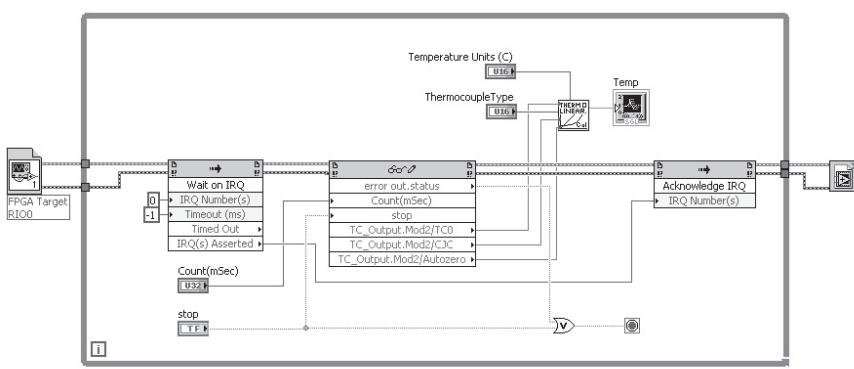
FPGA Front Panel



cRIO Front Panel

FIGURE 7-23 BLOCK DIAGRAM OF THE TEMPERATURE SENSOR (EXAMPLE 7.9)

FPGA Block Diagram



cRIO Block Diagram

EXAMPLE 7.8 Waveform Input

In many applications, acquiring one point at a time may not be fast enough. In addition, it is difficult to attain a constant sample interval between each point, because the interval depends on a number of factors: execution speed of the loop, software overhead of the call, and so on. With the *AI Acquire Waveform VI*, one can acquire multiple points at rates greater than the single point AI VIs can achieve. Furthermore, the VIs can accept user-specified sampling rates.

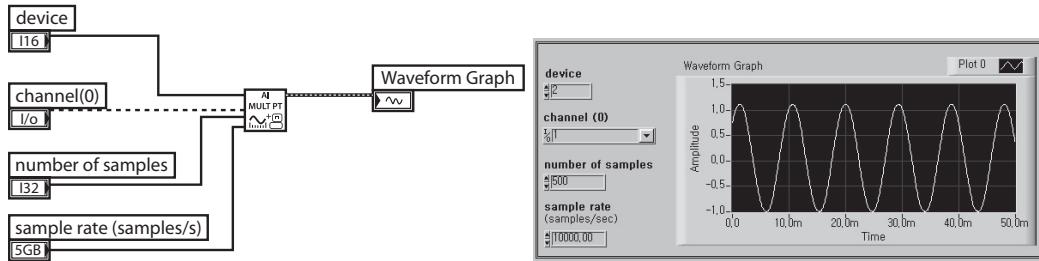
Solution

For this experiment, as shown in Figure 7-24, the analog input channel is connected to the sine wave output of the function generator.

EXAMPLE 7.9 Weight Measurement System with Strain Gauge

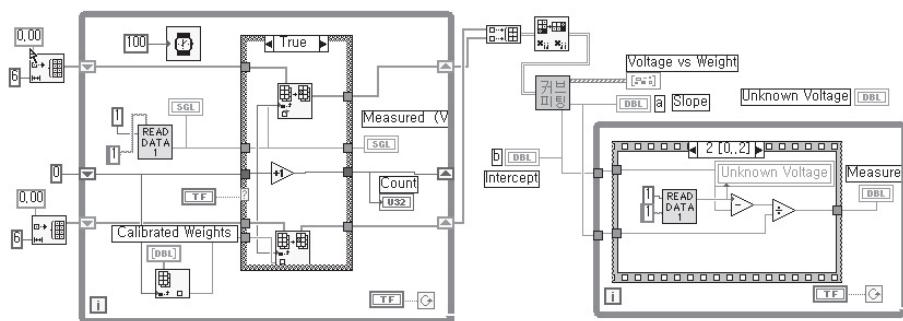
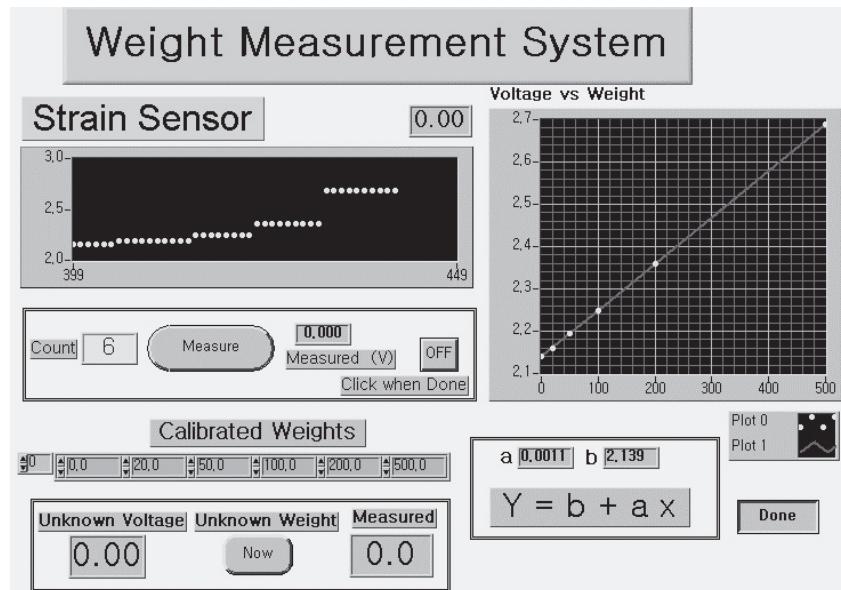
Strain gauges are important to any device under stress or strain. This example shows how the strain gauges are used to measure the displacement-strain relationship of the aluminum beam. Here, double strain gauges are mounted on the two sides of 15-cm long aluminum beam. As the known weight is placed at the end of the beam, the beam deflects as per the weight, and the virtual instrument monitors the strain signal from the

FIGURE 7-24 FRONT PANEL AND (B) THE BLOCK DIAGRAM OF THE WAVEFORM INPUT VI



strain gauges. The external bridge circuitry is used for the strain gauge measurements. The external voltage source from the power supply is used as the excitation voltage for the strain gauge. Figure 7-25(a) illustrates the front panel and Figure 7-25(b) illustrates the block diagram of the weight measurement system.

FIGURE 7-25 (A) FRONT PANEL OF THE WEIGHT MEASUREMENT SYSTEM (B) BLOCK DIAGRAM OF THE WEIGHT MEASUREMENT SYSTEM



Solution

At first, the calibration of the system is done using five different known weights. The virtual instrument gathers the input signal and plots the weight versus voltage as a graphical plot. After calibration, the user experiments with the unknown mass to the system. The virtual instrument displays the weight of the mass. This experiment provides an example of calibration, curve fitting, and measurement.

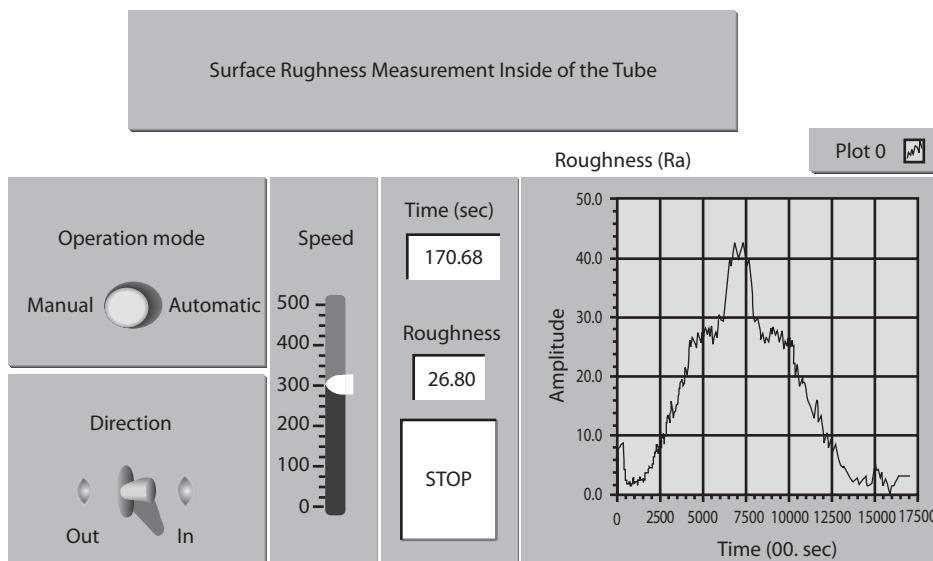
EXAMPLE 7.10 Using LabVIEW for Measurement of Surface Quality

An essential part of the quality control in the manufacturing of precision parts is the measurement of the quality of the surface finish of the part produced and, more particularly, the roughness of the surface. Conventional measurement techniques often require surface contact with the object being measured, which could potentially damage the surface. The evaluation of roughness through surface contact involves the use of a stylus device, which is drawn over the sample in order to detect and record variations in surface irregularities. Compared to the contact stylus method of inspection, the non-contact optical techniques can provide the same information in a quick, non-contact and flexible manner.

Solution

A non-contact laser probe can provide optical signature of the surface being scanned. The probe consists of a laser and photosensor, which provide calibrated electrical signals in proportion to the surface quality data. Figure 7-26 shows the LabVIEW adaptation of the instrument. It displays the front panel of the virtual instrument for the measurement of the surface roughness of the measured surface.

FIGURE 7-26 BLOCK DIAGRAM OF THE SURFACE QUALITY MEASUREMENT SYSTEM



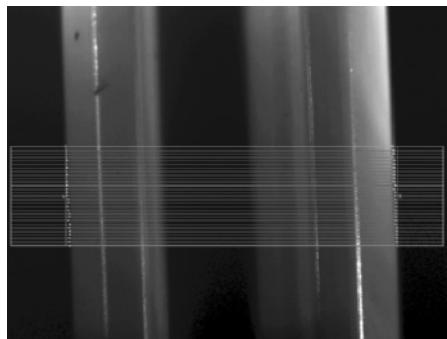
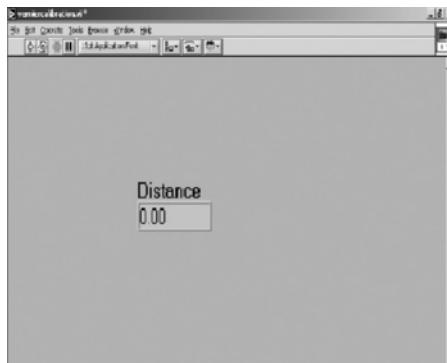
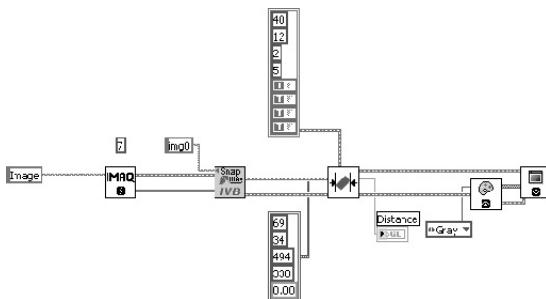
EXAMPLE 7.11 Vision-Based Dimensional Measurement

This case shows how vision-based systems can be applied to measure a component dimension. As an example, a plastic electrical raceway (i.e., conduit for electric wires in buildings) is chosen as a product example. The critical dimension under examination is the width of the raceway. A LabVIEW program can be used along with an *IMAQ* Vision module.

Solution

Figure 7-27(a) shows the block-diagram screen for the program. The image obtained using the digital camera is fed into the *IMAQ* software through the data acquisition device. Then using the “*CLAMP*” function, the distance between the edges is found. *CLAMP* function measures a distance in the horizontal direction, inside a set “*Region of Interest*,” and based on the set parameters (like measuring the distance in horizontal or vertical

FIGURE 7-27 VISION-BASED DIMENSIONAL MEASUREMENT SYSTEM



direction, image contrast, line spacing etc.) The front panel of the VI file, as shown in Figures 7-27(b) and (c), show the digital image after being processed by *IMAQ* software using the super imposed “Region of Interest” to find the distance between the edges.

EXAMPLE 7.12**Measurement of the Angular Position of a Motor Shaft Using Hall Effect Sensor**

It is important to understand that signal conditioning of any sensor output occurs at both hardware level and software level. In Chapter 3, while covering the constructional details of a Hall-effect sensor, it was shown how the low-voltage output signal of the Hall sensor is amplified using a differential amplifier and then the linear output signal is converted into digital signal by using the Schmitt trigger. Now, let us create a logic using LabVIEW that will take the digital signals from the inbuilt Hall sensors of the PM DC gear motor (Model No. IG420049-SY3754) and convert it into engineering units (radians).

Solution

We know from the discussion in Chapter 3 that, if the state of *A* or state of *B* is changing, we have to increment the motor position count by 1 if it is moving in the same direction or decrement it by 1 if it is moving in the opposite direction. Considering the counterclockwise direction of the motor to be positive, we would need to increment the count by 1 if the state of *A* is different from the previous state of *B* and decrement the count by 1 if the state of *A* is same as the previous state of *B*. Based on this discussion, a FPGA logic was developed to count the rotation of the motor shaft using LabVIEW 8.5, as shown in the Figure 7-28.

As shown in figure 7-28(a), whenever there is a change in the state of *A* (i.e., current state of *A* and previous state of *A*) or *B*, the counter ‘Encoder Position’ is incremented or decremented by 1 based on the current state of *A* and previous state of *B*. If the current state of *A* and previous state of *B* are the same, the counter ‘Encoder Position’ is incremented by 1, or it is decremented by 1. Figure 7-28(b) shows that the counter ‘Encoder Position’ is neither incremented nor decremented when there is no change in the state of *A* or *B*. This logic will be used in the real-time interface of the PM DC Motor system which will be covered in Chapter 8.

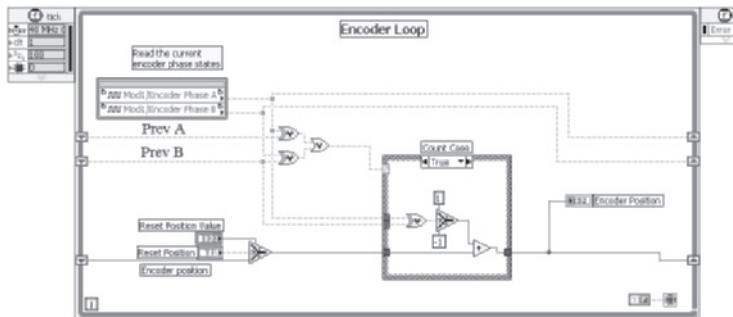
EXAMPLE 7.13**Pulse Width Modulation (PWM) of the PI Controller Output of the PM DC Gear Motor Position Control System**

Based on the error between the measured position of the motor shaft and the desired position of the motor shaft, the PI controller will decide what voltage should be given to the motor. Considering that we will be using a constant voltage power source (24 V DC supply) there is a need to develop some means to modulate the available power source. This was done by using NI cRIO 9004 with the FPGA logic for NI I/O module 9505. The NI 9505 module is a full H-bridge servo motor drive which enables a voltage to be applied across a DC motor in either direction. Figure 7-29 shows a basic H-bridge circuit (with reference to Figure 7-29). When switches S1 and S4 are closed and S2 and S3 are open, a positive voltage will be applied across the motor. On the other hand, by opening S1 and S4 switches and closing S2 and S3 switches, this voltage is reversed, allowing reverse operation of the motor.

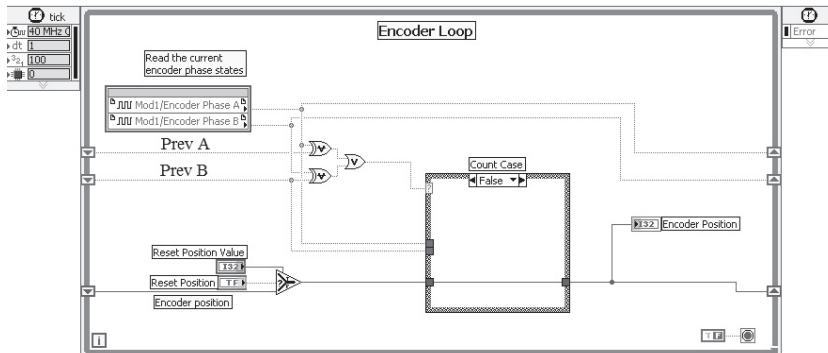
Solution

Then NI 9505 I/O module has an built-in hardware logic which takes signals through the variable name ‘*Motor*’ and ‘*Drive Direction*.’ If the signal to the variable ‘*Motor*’ is set to 1, the pair of switches S1 and S4 will be turned ON or the pair of switches S2 and S3 will be turned ON. If the signal to the variable ‘*Motor*’ is set to 0, all four switches will be open. The variable ‘*Drive Direction*’ decides which pair of switches to be turned ON when the variable ‘*Motor*’ is set to 1. If the signal to the variable ‘*Drive Direction*’ is set to 1, switches S1 and S4 will be turned ON, and the motor will rotate in one direction. If it is set to 0, switches S2 and S3 will be turned ON and the motor will run in the opposite direction. Thus, based on positive or negative voltage calculated by the PI, we can change the direction of the motor rotation. Now, the other task would be to

FIGURE 7-28 BLOCK DIAGRAM SCREEN FOR MOTOR SHAFT ANGULAR DISPLACEMENT MEASUREMENT

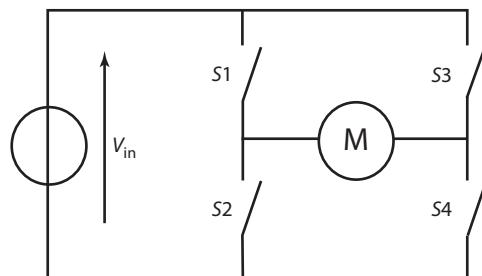


(a) Count Case when either state of signal A or signal B changes



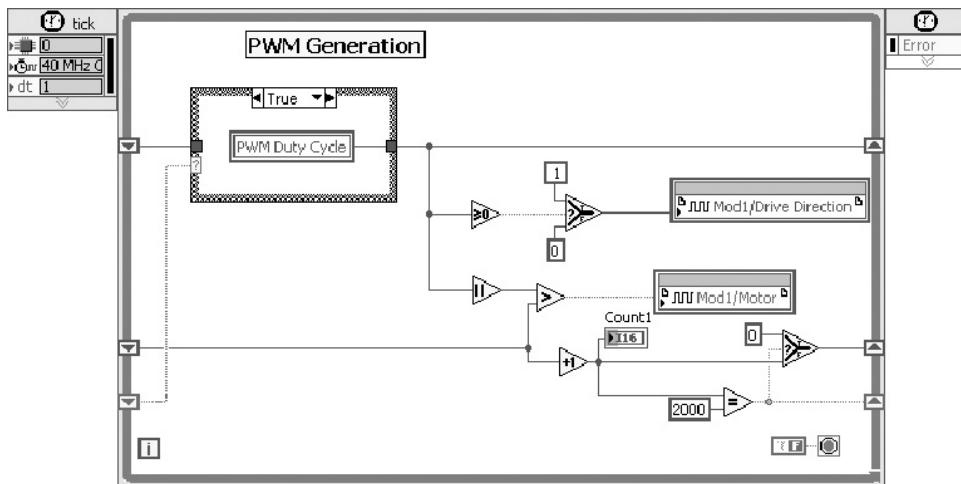
(b) Count Case when neither state of signal A nor signal B changes

FIGURE 7-29 H-BRIDGE CIRCUIT



vary the magnitude of the voltage given to the motor. This is done by generating a PWM signal, and the value of the duty cycle of the PWM is decided by the magnitude of the voltage calculated by the PI. Figure 7-30 shows the FPGA logic used to generate the PWM.

As seen from Figure 7-30, if the ‘*PWM Duty Cycle*’ is greater than or equal to 0 the signal to the variable ‘*Drive Direction*’ is set to 1 or else is set to 0. NI cRIO 9004 FPGA has an built-in clock of 40 MHz;

FIGURE 7-30 FPGA LOGIC FOR PWM

hence, the loop shown in Figure 7-30 is executed at every $1/40\text{ MHz} = 25\text{ ns}$. Thus, the counter ‘*Count1*’ is incremented by 1 at every 25 nanosecond and goes to 0 when it is equal to 2000. Thus, the period for the PWM is $25 \times 2000 = 50\text{ }\mu\text{s}$. If the absolute value of the ‘*PWM Duty Cycle*’ is greater than the counter value the variable ‘*Motor*’ is set to 1 or else it is set to 0.

For example, if the PID calculates a voltage value of 12 V, which corresponds to a duty cycle of 50% with a 24 V DC power supply and 1000 ticks with a counter made to go for a maximum of 2000 ticks, the variable ‘*Motor*’ is set to 1 for $25\text{ }\mu\text{s}$ and becomes 0 for the remaining 25 seconds (i.e., switches S1 and S4 are turned ON with switches S2 and S3 open for $25\text{ }\mu\text{s}$, sending 24 V to the motor in positive direction and all of the switches are open for the remaining $25\text{ }\mu\text{s}$, sending 0 V to the motor). Hence, the voltage for the entire cycle of $50\text{ }\mu\text{s}$ is $(24*25 + 0*25)/50 = 12\text{ V}$.

7.6.3 VisSim Environment

VisSim is a block-diagram-based simulation language designed to assist you in the analysis and design of systems (both continuous and discrete), controls, data acquisition, and system monitoring. Since VisSim is graphical in nature, no programming language experience is required (or even necessary). Applications are constructed as time-based block diagrams using the graphical editor. Block diagrams may contain several types of functions including: basic math (arithmetic, trigonometry functions, and algebra), Boolean logic, integration, time delays, and table look-ups.

After construction, the diagram may be simulated over a specified period of time called the time range. During simulation, the model is executed in a giant *do loop*, where time is directly proportional to the loop counter. The amount of time traversed in each iteration of the loop is usually constant and referred to as the “*TimeStep*” or simply the “*Step*” of the simulation.¹ Diagrams may be simulated in either of two modes: (1) simulated time and (2) real time.

In the simulated time solution, one second of elapsed time in the model will not correspond with one second of real time (real time is the time we live in as measured by your wristwatch). The speed at which the model is solved is a function of the PC processor and the complexity of the model. For example, a detailed model of an AC motor must be solved using extremely small ‘TimeSteps’ (0.1 milliseconds is common) and, as such, will likely run slower than real time. On the other hand, a model of an aircraft flight trajectory can be solved using a relatively large ‘TimeStep’ resulting in a faster than real-time solution. Models which simulate faster than real time can be forced to simulate in real time (usually this selection is through a switch or flag) by adding a fixed delay (wait time) to each time step.

This behavior is called underframing, because the time required to solve the model once is less than that amount of real time. On the other hand, models which simulate slower than real time can never be forced to simulate in real time unless they are first simplified and made to run faster. This behavior is called overframing because the time required to solve the model once exceeds that value of real time. The model time base cannot synchronize with the real time and will continually fall behind the real-time clock.

Whenever an integrator or time-delay block is used in a model, you must be aware that there is a potential that the model may become unstable. There are certain things you can do to minimize this potential, but first we examine model stability through an example.

Consider the difference equation

$$x_{k+1} = 0.5x_k \text{ with } x_0 = 1 \quad (7-20)$$

The results of several iterations of Equation 7-20 are presented in Table 7-2.

TABLE 7-2 ITERATION TABLE SOLUTION FOR EQUATION 7-20

k	x_k	x_{k+1}
0	1	0.5
1	0.5	0.25
2	0.25	0.125
3	0.125	0.0625
4	0.0625	0.03125

From the results in Table 7-2, a closed-form representation for the solution can be established and is

$$x_k = (0.5)^k x_0 \quad (7-21)$$

In Equation 7-21 the number 0.5 is called a pole (also called an eigenvalue or characteristic root). The “speed of response” or time constant of the system is governed by the value(s) of its pole(s). For discrete time systems, as the magnitude of the pole approaches 1, the speed of the response becomes slower. As the magnitude approaches 0, the response speed gets faster. For continuous time systems, as the real part of the pole approaches 0 but remains negative, the response becomes slower. As it becomes large and negative, the response becomes faster.

A stable equation is one which returns to a finite or bounded condition when perturbed. One way of applying a perturbation to an equation is through initial conditions. Consider Equation 7-21

¹This is a function of the numerical integration algorithm employed. Fixed step algorithms, such as Euler and the Runge Kutta methods, operate at regular time intervals. Variable step algorithms, such as the gears method and Adaptive Runge Kutta methods, operate at variable time intervals.

with an initial condition $x_o = 1$. As $k \rightarrow \infty$ the state, x_k , goes to 0. This is a stable response. Now let us increase the pole to a value of 2. Several iterations of Equation 7-21 with any non-zero initial condition reveal a constantly increasing value for x_k ; this is an unstable response.

For a discrete system to be stable, the magnitude of all system poles must be < 1 .

A general nonlinear differential equation, given in Equation 7-23 is

$$\dot{x}(t) = X(x(t), u(t))$$

where

$$x(t) \equiv \text{state} \quad \text{and} \quad u(t) \equiv \text{input}. \quad (7-23)$$

At a specified time, t_1 , Equation 7-23 may be linearized using the Taylor series approach discussed earlier.

$$\dot{x}(t) = Ax(t) + Bu(t); \quad A \equiv \left. \frac{\partial X}{\partial x} \right|_{\substack{x=x(t_1) \\ u=u(t_1)}} \quad \text{and} \quad B \equiv \left. \frac{\partial X}{\partial u} \right|_{\substack{x=x(t_1) \\ u=u(t_1)}} \quad (7-24)$$

The eigenvalues of the A matrix correspond to the system poles. When a differential equation is solved on a computer, it is solved at regularly spaced time points; each is separated by a ‘TimeStep’ of Δt seconds. We illustrate two discrete approximations which are often used; forward Euler and backwards Euler integrations. The effect of the methods and the choice of ‘TimeStep’ on the stability of a general equation (in this case we will use Equation 7-24) is investigated in the following solutions.

Solution of Equation 7-24 using the forward Euler method

$$\begin{aligned} x_{k+1} &= x_k + \Delta T \cdot \dot{x}_k \\ x_{k+1} &= x_k + \Delta T A x_k + \Delta T B u_k \\ x_{k+1} &= x_k(I + \Delta T A) + \Delta T B u_k \\ \text{Unforced Response } x_{k+1} &= (1 + \Delta T A)^k x_0 \\ \text{Stability Requirement: } |I + \Delta T A| &< 1 \end{aligned}$$

If A is stable: $\Delta T < 0$, where, $|A|$ is the real part of the largest pole of A .

Solution of Equation 7-24 using the backward Euler method

$$\begin{aligned} x_{k+1} &= x_k + \Delta T \cdot \dot{x}_{k+1} \\ x_{k+1} &= x_k + \Delta T A x_{k+1} + \Delta T B u_{k+1} \\ x_{k+1} &\equiv (I - \Delta T A)^{-1} x_k + \Delta T (I - \Delta T A)^{-1} B u_k \\ \text{Unforced Response } x_{k+1} &= (I - \Delta T A)^{-k} x_0 \\ \text{Stability Requirement: } |(I - \Delta T A)^{-1}| &< 1 \end{aligned}$$

If A is stable: $\Delta T > 0$, where $|A|$ is the real part of the largest pole of A .

Often one deals with stable A -matrices, meaning all poles have negative real parts. In these situations, the stability of the forward Euler method depends on the “TimeStep” value relative to the

value of the poles of A . The backward Euler method is not affected by the choice of the ‘*TimeStep*’ and cannot become unstable for any positive ‘*TimeStep*’ value.

To illustrate the effect of the ‘*TimeStep*’ on equation stability, consider the stable ordinary differential equation and initial condition:

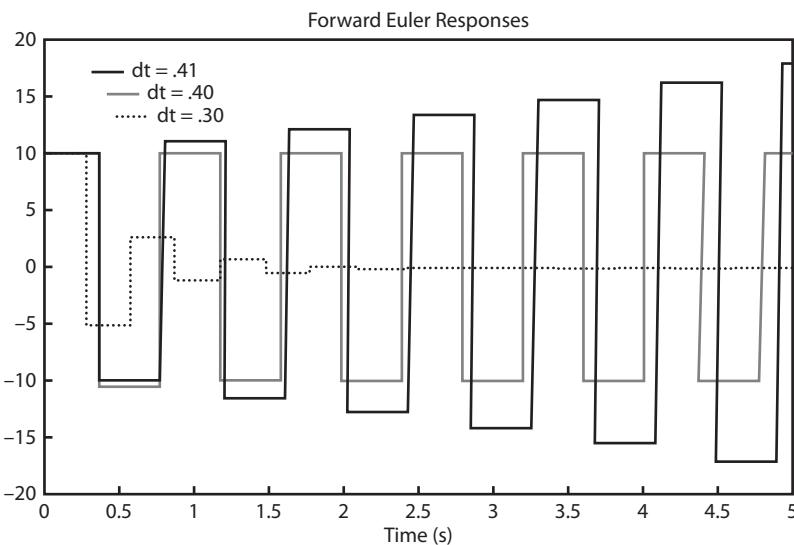
$$\dot{x}(t) = -5x(t) \text{ and } x(0) = 10 \quad (7-25)$$

Defining Δt as the update time, the forward Euler solution is solved in Equation 7-26.

$$x_{k+1} = (1 - 5 \cdot \Delta T_d)x_k \quad (7-26)$$

Our stability analysis indicates that the stability range for the forward Euler integration approximation from Equation 7-26 is $.4 >= \Delta t > 0$. Simulation results are presented in Figure 7-31 for three values of Δt , .41, .40, and .30. The results agree with the predicted results. The forward Euler method becomes unstable for $\Delta t > .40$ and remains stable for $.4 >= \Delta T_d > 0$.

FIGURE 7-31 FORWARD EULER RESULTS FOR EQUATION 7-26



Configuring VisSim for Real-Time Operation This section explains how to configure VisSim for real-time operation with one or more GPIO cards. We assume that VisSim and the real-time add-on have been installed on the computer. Figure 7-32 presents the VisSim Driver Setup dialog box. The meaning of the various settings is described here.

Analog Input Range: Available analog input ranges.

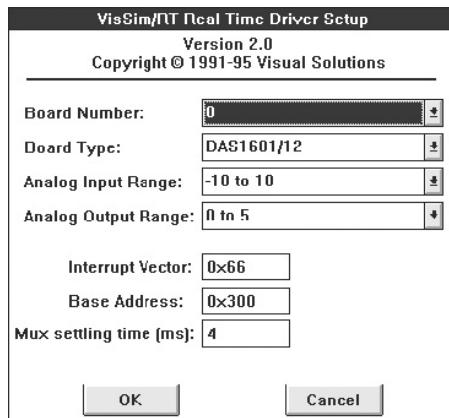
Analog Output Range: Available analog output ranges.

Base Address: The I/O port register address through which the driver commands the card, typically set to 0×300 (hexadecimal 300) and configurable between 0×220 and $0 \times 3FF$.

Board Number: The I/O card being configured, ranges from 0 to 15.

Board Type: The different board types available.

Mux Settling Time (ms): The time needed for the voltages to settle when using the multiplexer cards. The default is 4 ms.

FIGURE 7-32 VISSIM DRIVER SETUP DIALOG BOX

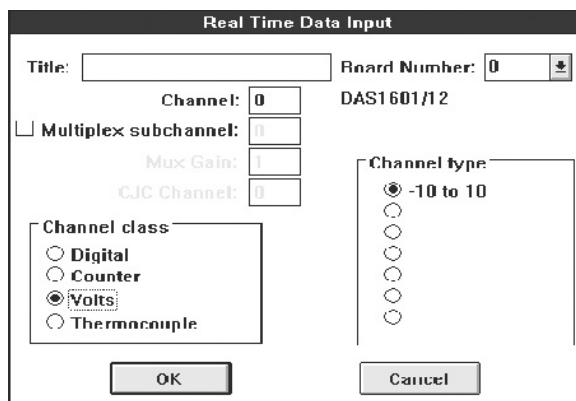
The VisSim *rt-DataIn* and *rt-DataOut* blocks are now *live* and must be individually configured for communication with analog, digital, or frequency channels on the card. The *rt-DataIn* block is used to read an input channel. Click the right-mouse button over the *rt-DataIn* to access a dialog box containing channel configuration parameters. This dialog box is shown in Figure 7-33. The meaning of the various settings is described here.

Board Number : Specifies the card number.

Channel: A number that must correspond with that channel number on the screw terminal supplied with the I/O board. VisSim uses channel 0 as the first channel, even if the documentation supplied by the board vendor describes the first channel as channel 1.

Channel Class: Any of six types of inputs: digital, voltage, current, quadrature encoder, thermocouple, and counter.

Counter: Provides a high frequency pulse-counting input. Pulses are generated by triggering on the leading edge of an incoming sinusoidal-shaped waveform. Most counters can count approximately

FIGURE 7-33 VISSIM RT-DATAIN DIALOG BOX

64,000 pulses before they overflow and reset. In VisSim, the counter value is reset at the end of every time step of the simulation. If you accumulate more than 64,000 counts, you should reduce the step size using the *Simulate* menu's *Simulation Setup* command. Most boards have at least one counter input. When using the counter channel, be particularly careful connecting the screw terminal to the signal source. Most board vendors re-use an existing digital channel for the counter input.

As an example of how the counter may be used, consider a sensor that generates a sinusoid with a frequency that corresponds to the velocity of a fluid. Typical frequencies may range from 1 kHz to 10 kHz. Assuming that you have set the simulation step size to 0.1 seconds, the number of pulses that occur in one step of the simulation can be computed by dividing the counter value (output of the *rt-DataIn* block configured to the counter) by the simulation time step.

Current: Supports current output applications and is set to a 4 to 20 mA range, providing a full 12 bits of resolution over that range.

Digital: Provides an ON/OFF channel input. When *Digital* is activated, the input behaves like a current sink. When the voltage level on the digital input line goes low, current flows from the 5 V power supply to ground. When the digital input channel is activated, the voltage level of the channel goes low (turns OFF). Most digital input channels are capable of sinking 10 mA. For digital inputs, VisSim uses the card manufacturer's specifications for TTL-level values.

Quadrature Encoder: Outputs the rotation count of the encoder. This parameter is applicable for the M5312 board only.

Thermocouple: Provides the thermocouple linearization input. A thermocouple produces a voltage corresponding to the temperature measured. When the Thermocouple parameter is activated, the *Channel Type* box shows different types of thermocouples (B, E, J, K, R, S, and T). VisSim/RT provides for cold junction compensation for the thermocouple linearization; it uses channel 7 on the I/O board for reading the temperature from the solid-state temperature device on the multiplexer card for cold junction compensation.

Volts: Provides a time-varying voltage input. The range of the voltage input is software-selectable for many boards. Boards with this characteristic are referred to as *programmable gain boards*. If you are using a programmable gain board and you activate *Volts*, the voltage ranges are presented under the *Channel Type* box. These typically range from ± 10 to $\pm .01$ V and lower. For boards without programmable gain, the input voltage range is normally set using a micro-switch on the I/O board itself. The corresponding voltage ranges, listed under the *Channel Type* box, are read-only settings. The voltage channel is often called an A/D channel. Analog input signals are converted to digital representations using a converter. The converter consists of registers whose numbers determine the resolution. Most boards use 12-bit resolution converters; however, some boards use higher resolution converters for greater precision. Channel resolution is proportional to the channel read time. The higher the channel resolution, the longer it takes to read the channel. In cases where accuracy is less important than speed, you may reduce the resolution.

Channel Type: Specifies the channel type. The channel type varies according to the possible input ranges for the board. The channel type is dependent on the selection under the *Channel Class*. When either digital or counter is activated, there are no channel type options; the channel type defaults to a special internal type. When *Volts* is selected for a programmable gain board, the *Channel Type* lists the allowable ranges from which you can choose. When *Volts* is selected for a non-programmable gain board, the *Channel Type* displays the hardware settings for the input voltage range on the board.

CJC Channel: Specifies the cold junction compensation for the card. VisSim uses the card manufacturer's specification for the cold junction compensation. Typically, this is 0. If you

physically change the cold junction compensation on the board, use this parameter to make the same change to the software. The *Cold Junction Compensation* parameter is available for use only when the *Thermocouple* parameter under *Channel Class* is active.

Multiplex Subchannel: Connects to an individual multiplex channel on a multiplex board. Each multiplex board physically connects to all of the analog input channels on the I/O board, letting you daisy chain them. Electrically, each multiplex board only connects to one of the I/O board's input channels. The electrical connection connects to only one of the analog input channels and depends on the jumper setting on the multiplex board. To set the jumper, refer to the documentation that accompanies the board. When you activate the *Multiplex Subchannel* parameter, you must also enter a subchannel number in the accompanying *Multiplex Subchannel* text box. VisSim/RT sends the multiplex subchannel number to the first four digital out channels, which the multiplex card uses as the multiplex subchannel specified and then reads the input channel. (You should not use the digital channels for other purposes.)

Mux Gain: Indicates the gain applied to a multiplexed signal when reading it into VisSim/RT. VisSim/RT divides the signal by the gain. When you change the mux gain on a channel, all of the gains on all of the multiplex subchannels are changed.

Title: Indicates an optional, non-executable, user-specified description of the channel. The title you specify appears on the *rt-DataIn* block in the VisSim diagram.

The *rt-DataOut* block is used to write data to an output channel. Click the right-mouse button over the *rt-DataOut* to access a dialog box containing channel configuration parameters. This dialog box is shown in Figure 7-34.

The meaning of the various settings is described here.

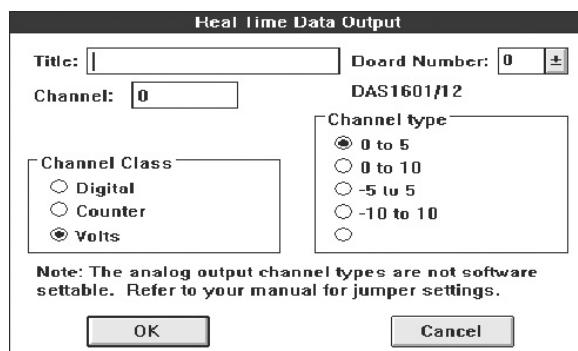
Board Number : Specifies the card number.

Channel: A number that must correspond with that channel number on the screw terminal supplied with the I/O board. VisSim uses channel 0 as the first channel, even if the documentation supplied by the board vendor describes the first channel as channel 1.

Channel Class: Either: digital, analog, or pulse generator.

Digital: Provides an ON/OFF channel output. When *Digital* is activated, the output behaves like a current sink. When the voltage level on the digital output line goes low, current flows from the

FIGURE 7-34 VISSIM RT-DATAOUT BOX



5 V power supply to ground. When the digital output channel is activated, the voltage level of the channel goes low (turns off). Most digital output channels are capable of sinking 10 mA. For real-time digital output, the inputs generated are Boolean in nature. The voltages corresponding to the ON/OFF states of the digital channel obey TTL-level values. Low is less than 0.7 V, and high is greater than 2.5 V. Digital I/O lines are especially useful for controlling equipment using ON/OFF signals. The channel itself will, most often, not have enough power to actuate. In these situations, an opto-isolated solid-state relay is used. These relays, supplied by the board vendor, are soldered onto the screw terminal panel. Instructions are provided by the board vendor. The opto-isolators have two sides: a low power side and a high power side. The digital I/O communicates with the low power side. By connecting a high power source plus equipment through the high power side, you may switch high AC or DC power using the digital I/O line. Most vendors offer opto-isolator modules ranging from 3 A with voltages to 280 V AC, to 45 A with voltages to 650 V AC.

Counter: Outputs high-frequency square waves. The counter output utilizes the board's counter timer and often re-uses an existing digital I/O channel. Refer to the documentation accompanying your board for information on wiring this channel. Data sampling rates of up to 20 kHz (counter-assisted) can be achieved. As an example of how the counter output operates, consider the control of a stepper motor/drive system. The drive is capable of receiving command pulses from 0 to 5 kHz to regulate its speed. By connecting a slider block with a range of 0 to 5000 to the *rt-DataOut* block configured for counter output, the motor speed can be controlled over its full range.

Volts: Provides a time-varying voltage output. The range of the voltage output is software-selectable for many boards. Boards with this characteristic are referred to as programmable gain boards. If you are using a programmable gain board and you activate *Volts*, the voltage ranges are presented under the *Channel Type* box. These typically range from ± 10 to ± 0.01 V and lower. For boards without programmable gain, the output voltage range is normally set using a micro-switch on the I/O board itself. The corresponding voltage ranges, listed under the *Channel Type* box, are read-only settings. The voltage channel is often called an D/A channel. Analog input signals are converted to digital representations using a converter. The converter consists of registers whose numbers determine the resolution. Most boards use 12-bit resolution converters; however, some boards use higher resolution converters for greater precision.

Channel Type: Specifies the channel type. The channel type will vary according to the possible output ranges for the board. The channel type is dependent on the selection under the *Channel Class*. When either *Digital* or *Frequency* is activated, there are no channel type options; the channel type defaults to a special internal type. If the board does not support programmable ranges or gains, the channel type will be set to the range value specified in the *File* menu's *Real Time Config* command. When *Volts* is selected for a programmable gain board, the *Channel Type* lists the allowable ranges from which you can choose. When *Volt* is selected for a non-programmable gain board, the *Channel Type* displays the hardware settings for the output voltage range on the board.

Title: Indicates an optional, non-executable, user-specified description of the channel. The title you specify appears on the *rt-DataOut* block in the VisSim diagram.

7.6.4 VisSim Applications

Three detailed examples which utilize VisSim real-time interface capability are presented in this section. The first example illustrates how the force on a cantilever beam is computed from a measurement of beam strain. The second example describes the implementation of a stepper motor actuation system.

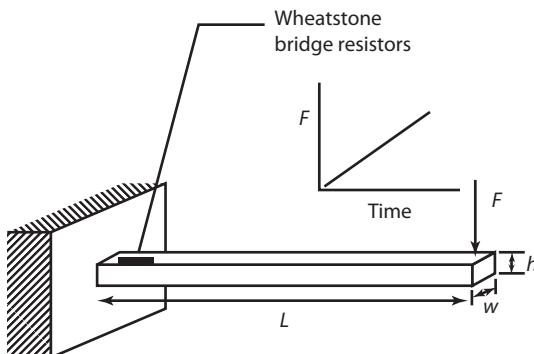
Stepper motors are popular for accurate positioning applications because they do not require feedback to achieve accuracy as do AC and DC servo motors. The third application presents a typical closed-loop test application. Derivatives of this example are plentiful in manufacturing and control applications.

EXAMPLE 7.14 Cantilever Beam Force Measurement

This system is a data acquisition system. A small steel horizontal beam is mounted at one end with the other end free to move. The free end is subjected to a linearly increasing vertical force. A strain gauge is attached to the fixed end and used to sense the strain due to the applied force. The force is then estimated from the strain gauge reading and the characteristics of the beam.

A diagram of the beam system is presented in Figure 7-35 along with the time history of the applied force, F . The location of the sensing resistors used in the Wheatstone bridge strain gauge are also shown.

FIGURE 7-35 CANTILEVER BEAM SYSTEM



Solution

Four symmetrically placed strain gauges, each having a gauge factor, G , are bonded to the beam at the support end. Two gauges, R_2 and R_3 , are bonded on the lower surface of the beam and forced into compression for positive F . The other two gauges, R_1 and R_4 , are bonded to the upper surface of the beam and forced into tension for the same applied force. The four gauges are connected in a Wheatstone bridge arrangement with an applied voltage, V_s . The bridge output voltage is amplified by an op-amp of gain A , producing an output voltage V_o suitable for sensing by the GPIO. The circuit diagram is presented in Figure 7-36.

One of the eight analog-input channels on an Advantech 711 GPIO card was used for the sensing operation. The DOS TSR for the card was loaded using the following statement.

```
advtech -d711 -b300
```

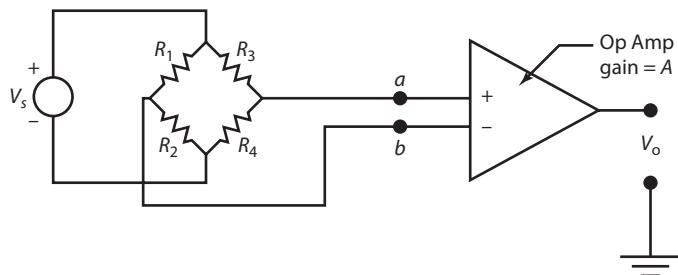
This statement loads the card driver and sets the card base address to 300 (hex). The relationship between V_o (the beam characteristics) and F is expressed as

$$F = \frac{w \cdot h^2 \cdot Y}{6 \cdot A \cdot V_s \cdot G \cdot L} \cdot V_o \quad (7-27)$$

where

$$w = 2.5 \text{ cm}$$

$$h = 0.25 \text{ cm}$$

FIGURE 7-36 WHEATSTONE BRIDGE CIRCUIT DIAGRAM

$$L = 30 \text{ cm}$$

$Y = 210 \text{ GPa}$, Young's modulus of elasticity

$$A = 10$$

$$V_s = 12$$

$$G = 2$$

An application program was created using VisSim to read the strain gauge value, compute the estimated force using Equation 7-27, and plot the results. Measurements are taken at 100 Hz for 10 seconds. The VisSim *Simulation Setup* is presented in Figure 7-37.

The application diagram which reads the input channel, computes Equation 7-27, and plots the estimated applied force in Newtons is presented in Figure 7-38.

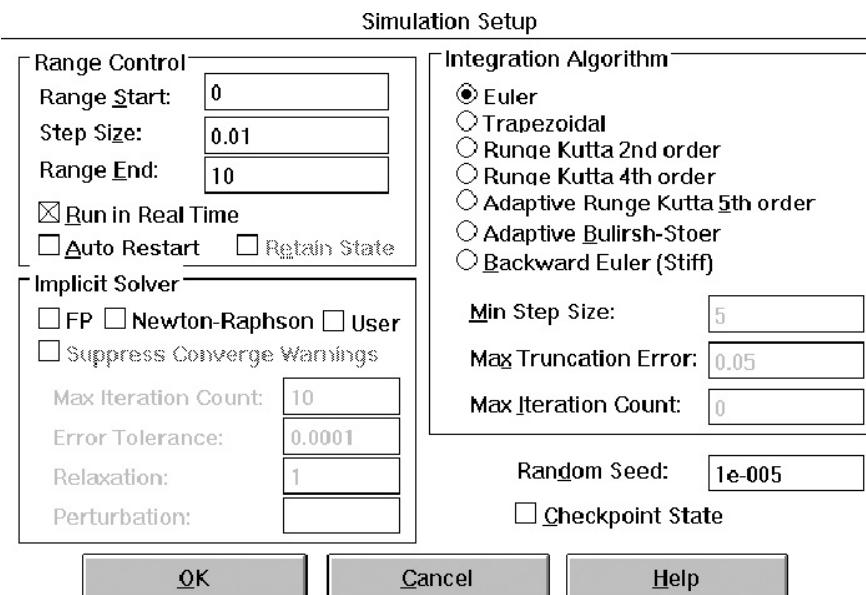
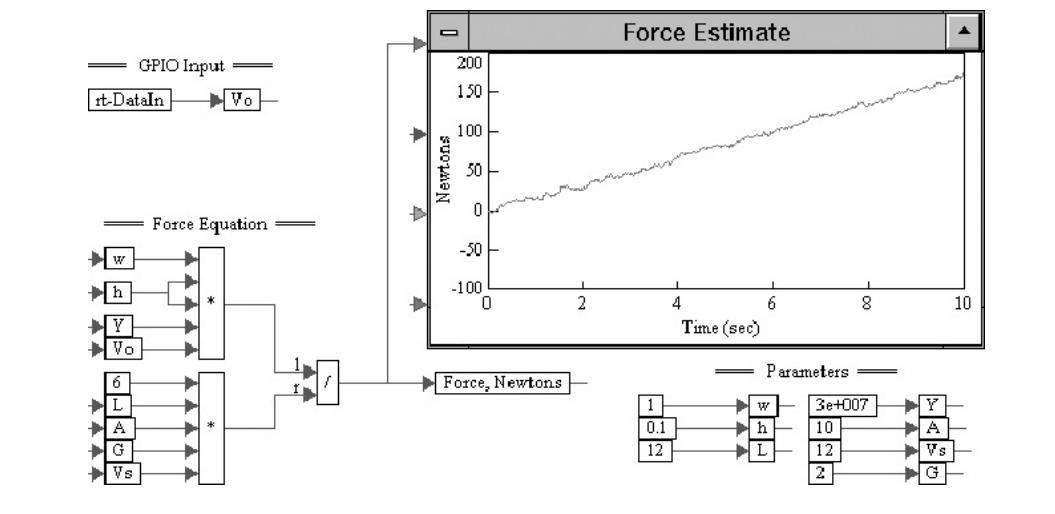
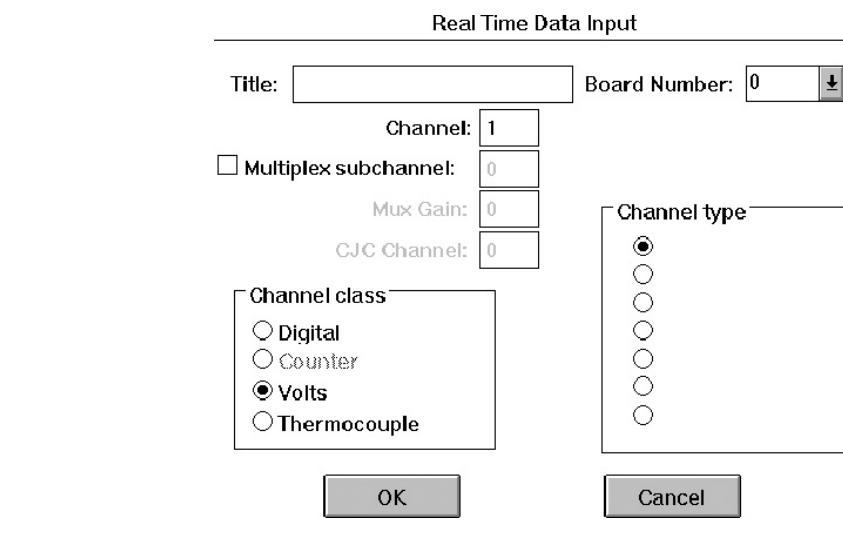
FIGURE 7-37 VISSIM SIMULATION SETUP DIALOG BOX FOR BEAM, EXAMPLE 7.16

FIGURE 7-38 APPLICATION DIAGRAM FOR BEAM FORCE ESTIMATION

VisSim variables have been used extensively in this diagram to simplify the wiring. In the lower-left section of the diagram, all required parameters are specified and loaded into meaningful variable names. The force equation is implemented in the lower left section of the diagram and the equation output is captured in the variable *Force, Newtons* and plotted on the graph. The GPIO input channel is read into the variable V_o in the upper-left section of the diagram. The setup of this *rt-DataIn* block is shown in Figure 7-39.

FIGURE 7-39 RT-DATAIN CONFIGURATION FOR BEAM IN EXAMPLE 7.16

Channel 1 of the eight possible analog-input channels on the 711 card is being used to communicate the V_o signal. The board number is defaulted to 0 because we are using only one 711 card, and the channel class is set to volts, indicating an analog input.

EXAMPLE 7.15 Controlling a Stepper Motor

There are five major components in a stepper motor actuation system:

1. The stepper motor
2. Step motor drive
3. Power supply
4. General purpose I/O card
5. Application program

The Stepper Motor

Most stepper motors are 8-wire motors and normally are configured to run in the parallel mode. We will use an oriental step motor: 8 wires, 1.2 A, 5Ω , $1.8^\circ/\text{step}$, which is connected in parallel for this demonstration.

Stepper Motor Drive

The IB462 bipolar step motor drive is a chopper drive capable of moving a stepper motor in full steps ($1.8^\circ/\text{step}$) or half steps ($.9^\circ/\text{step}$). This is a low-cost drive and has an enable/disable line that allows the stepper motor to be completely turned off from the computer. The drive is capable of utilizing input voltages to 40 V DC at up to 3.5 A. The pin locations for the drive are shown in Figure 7-40.

FIGURE 7-40 PIN LOCATIONS FOR THE IB462 DRIVE

1	● Enable	7
2	● Logic Ground	8
3	● H/F	9
4	● Step Clock	10
5	● CW/CCW	V
6	● Current Adjust	Ground
		12

Pin Description:

Enable: (0/1) = (ON/OFF).

Logic Ground: Digital ground on the screw terminal board.

H/F: (1/0) = (full step/ half step).

Step Clock: 0-1-0 pulse, moves motor one step on the falling edge.

CW/CCW: Direction, (1/0).

Current adjust: Connect to power supply ground with a resistor to lower the power used by motor. No resistor will apply a full 2 A to the motor.

Phases A, Abar, B, and Bbar: (pins 7 through 10): Connected according to motor setup instructions.

V: Supply voltage input 28 V, 3.5 A from the power supply.

Ground: Supply power supply ground, also connects to pin 6 (current adjust) through a resistor to limit the power consumed by the motor. No resistor will consume a full 2 A.

Pins 1, 3, 4, and 5 are connected to the screw terminal board, as are digital channels 1, 2, 3, and 4, respectively. Note, these must be connected to the IN side of each digital channel, not the OUT side. Pin 2 is connected to the screw terminal-panel digital ground. The current adjust (pin 6 to pin 12) resistor is not used, which allows a full 2 A to flow to the motor. If one is used, Table 7-3 summarizes some typical values.

TABLE 7-3 CURRENT-RESISTANCE RELATIONSHIP

Resistor Value (Ω)	Phase Current (amps)
133	0.5
402	1.0
1210	1.5
Open circuit (no resistor)	2.0

Power Supply

A universal 28 V, 3.5 A power supply was used for this application. Power supplies are available in different ranges and features.

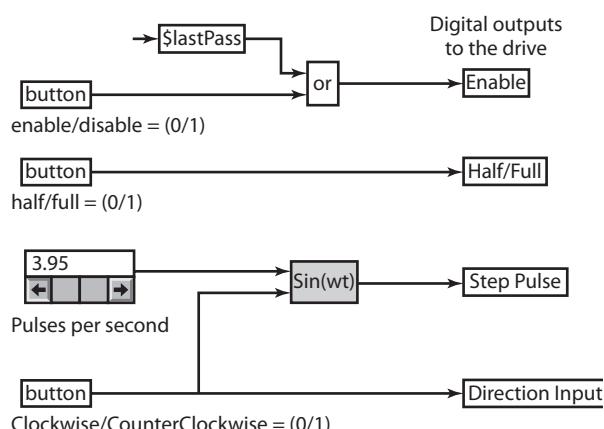
I/O Card:

Any general-purpose I/O card can be used to control the step motor from a computer. We will use a Strawberry Tree ACPC-Jr card with a T31 screw terminal panel. This card has eight analog inputs and 16 digital I/Os. Only four digital outputs are needed to control the stepper motor. The screw terminal wiring is covered in the previous chopper wiring section.

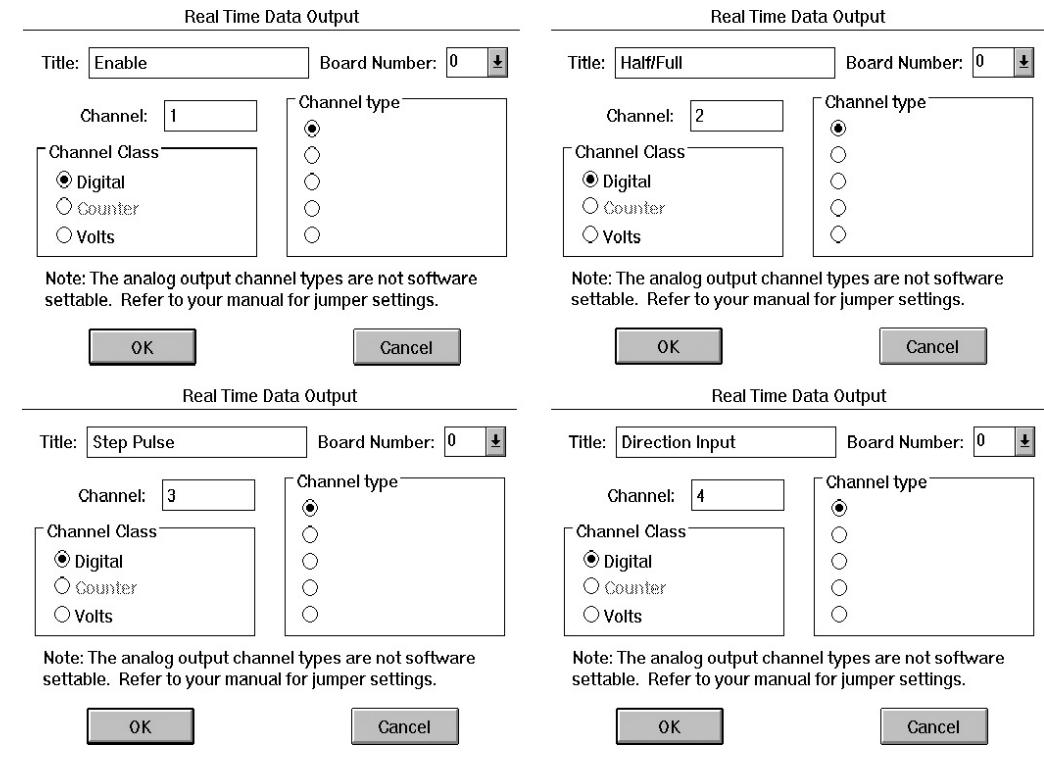
Solution

The purpose of the application program is to send digital signals to the four channels needed to actuate the stepper motor to achieve some desired motion. The application program is coded in VisSim and presented in Figure 7-41.

FIGURE 7-41 STEP MOTOR CONTROL PROGRAM



The four real-time digital outputs listed down the right side of Figure 7-41 are called *Enable*, *Half/Full*, *Step Pulse*, and *Direction Input*. Each of these four blocks is a *rt-DataOut* block internally configured as digital using channels 1, 2, 3, and 4 respectively. The configuration for each block is shown in Figure 7-42.

FIGURE 7-42 RT-DATAOUT BLOCK CONFIGURATIONS

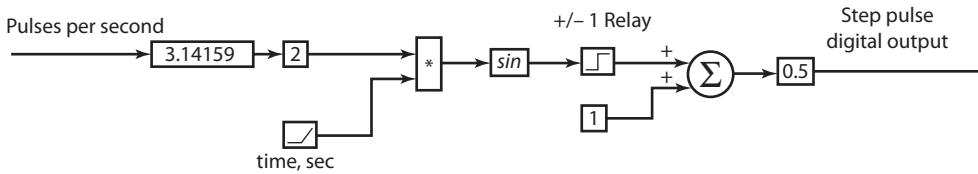
There are four things to notice here: the title, the board number, the channel number, and the channel class. Since this application uses only one GPIO card, it is automatically referred to as card 0. All four outputs are digital, so the *Digital* radio button is selected for the channel class. The title is the text that appears on the *rt-DataOut* block. Finally, the channel number is the channel on the screw terminal panel, so you must make sure that on the screw terminal you have wired digital output channel 1 to the enable pin on the drive, digital output channel 2 to the *Half/Full* pin, digital output channel 3 to the *Step Pulse* pin, and digital output channel 4 to the *Direction* pin.

In Figure 7-41, the *\$lastPass* variable is a VisSim “read only” discrete, which remains at 0 until the last pass through the diagram when it becomes 1. It is used in the diagram to turn the power OFF to the drive if the VisSim application is terminated for any reason.

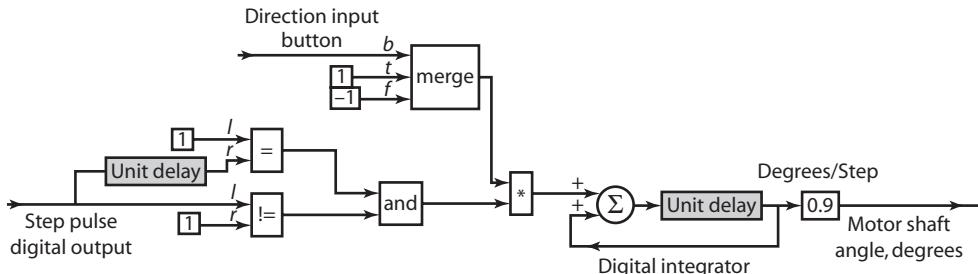
The *Step Pulse* channel requires some further explanation. The value in the slider block indicates the number of steps the motor should move per second. When using the half step mode, each pulse corresponds to 0.9° of motor shaft angle.

The *sin(wt)* block performs two functions: first it computes the *Step Pulse* waveform and second it computes the distance (in degrees) traveled by the motor by keeping track of the dictated pulses and direction. The contents of the *sin(wt)* block are presented in Figure 7-43.

The *Step Pulse* waveform is generated by creating a sinusoid using the slider value (pulses per second) as input and passing the sinusoid through a relay whose output is either -1 or 1 , depending on the sign of the sinusoid. This $-1, +1$, signal is a square wave and it is recentered to lie between 0 and 1 by first adding 1 to it and then dividing by 2 . The resulting signal is the *Step Pulse* waveform applied to the real-time output block.

FIGURE 7-43 SIN(WT) COMPOUND BLOCK

Since feedback is usually not used in stepper motor applications, it is necessary to integrate the *Step Pulse* waveform to keep track of the distance. The distance algorithm, which also resides in the *sin(wt)* compound block, is shown in Figure 7-44.

FIGURE 7-44 MOTOR SHAFT ANGLE CALCULATION

The logic to the left of the “and” block is used to isolate the “falling” edge of the step pulse signal. This is the condition where the signal changes from 1 to 0 in value. The “and” output is true (1) when the past value of the step pulse = 1 and the current value = 0. The block just to the right of the “and” block is a “multiply” block and is used to assign a sign to the direction of motion. The “merge” block output is either a 1 or -1 depending on the value of the *Direction* input button (either 0 or 1). To the right of the “multiply” block is a digital integrator used to sum the signed pulses, and finally, the accumulated pulse count is converted to degree units using the 0.9 °/step factor for half stepping.

The behavior of the *Step Pulse* signal and the *Motor Shaft Angle* signal during a 10 second test are shown in Figure 7-45 and Figure 7-46.

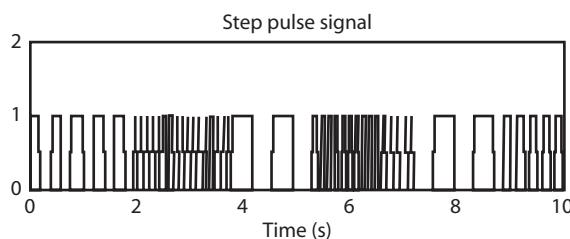
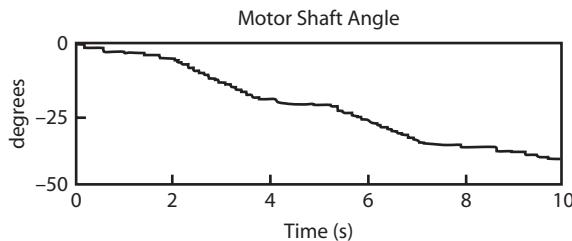
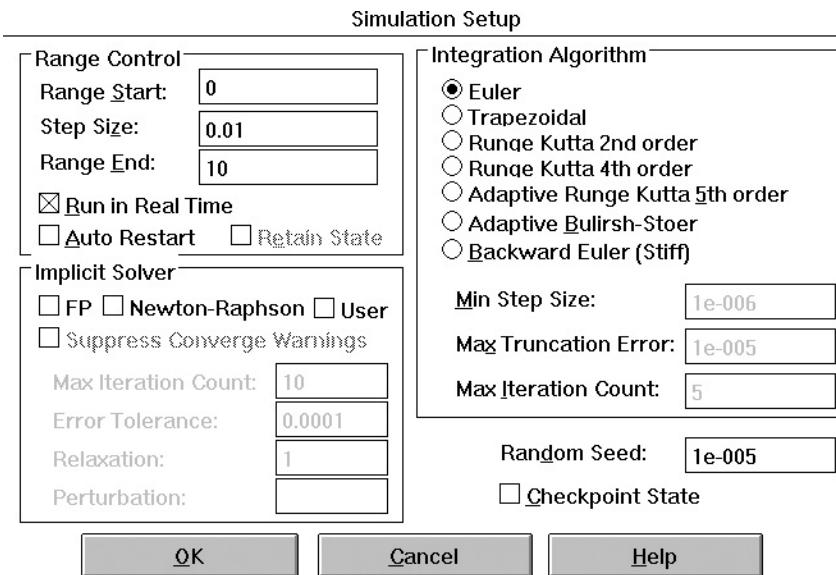
FIGURE 7-45 STEP PULSE SIGNAL BEHAVIOR

FIGURE 7-46 MOTOR SHAFT ANGLE BEHAVIOR

The setup information for the stepper motor application is shown in Figure 7-47. The *Run in Real Time* check box is selected to synchronize the application (model) clock with real time. The model will execute for 10 seconds of real time and will execute every 0.01 seconds (100 Hz). The *Integration Algorithm* selection only matters if continuous integrators are present in the model. In this example, all integrations are implemented as digital feedback loops, so the choice of integration algorithm doesn't matter. However, if you did have integrators in your model, the Euler algorithm is usually the best choice for real-time applications due to its speed.

FIGURE 7-47 SIMULATION SETUP SCREEN

EXAMPLE 7.16 Data Acquisition and Control System

This example illustrates the use of analog, digital, and counter timer (frequency) channels in a compressor-test rig system application. The use of two cards in the I/O system is also illustrated in this example, even though only one is needed. The system consists of a large, single speed induction motor directly connected to an air compressor. The pressure-flow characteristics of the compressor will be experimentally determined. The GPIO has five outputs, a digital ON/OFF channel to control the motor in either an ON or OFF mode and a step motor control

(which uses four digital output channels) to vary the restriction area on the compressor inlet. On the input side two signals are sensed: motor speed (using a counter timer attached to a laser tachometer) and the delta pressure, rise in the compressor (analog voltage channel). Since airflow is proportional to the area times the square root of the delta pressure, it is calculated as part of the application program in VisSim. The following relationships are used.

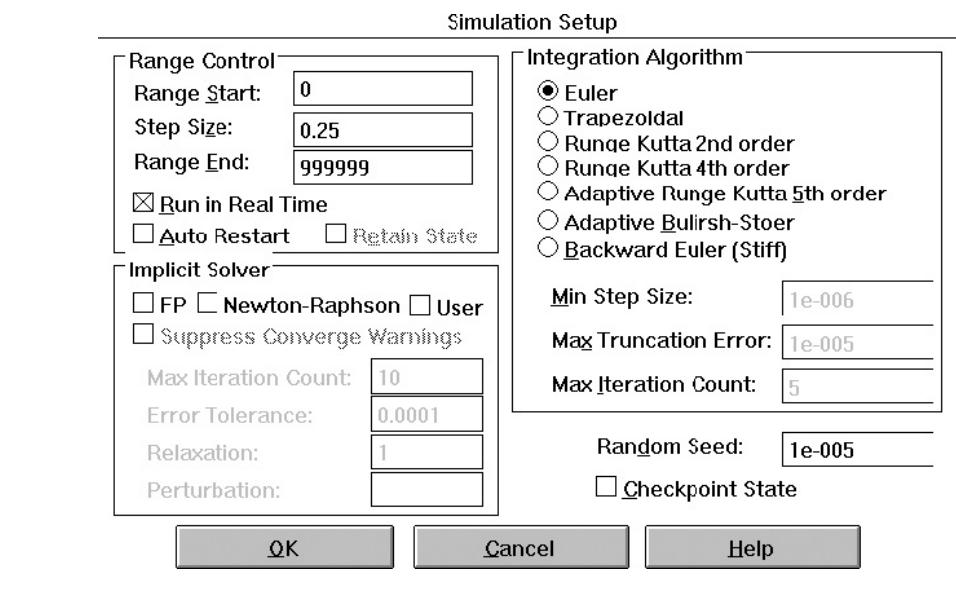
- The step motor moves in a range of 0 to 90° shaft angle. At 0°, the compressor inlet is closed off, and at 90° it is full-open. The scaling relationship between the two extremes was determined experimentally and is summarized in Table 7-4.
- Flow is computed from % Open and the delta pressure signals according to $F = K \cdot (\% \text{ Open}) \cdot \sqrt{\Delta P}$ where K is a constant used to convert units to flow in cubic feet per minute (cfm).

TABLE 7-4 SHAFT ANGLE–PERCENT OPEN RELATIONSHIP

Shaft Angle, deg	0	10	20	30	40	50	60	70	80	90
% Open	0	15.8	28.0	37.2	53.5	66.7	82.1	89.7	95.2	100

The test is expected to run over long periods of time, but VisSim demands that we specify this time. Since the test time is not known in advance, a time range much larger than ever expected is used—in this case the VisSim time range is selected to begin at 0 and end at 99999 seconds, this should be ample. The time step is selected to be .25 seconds (4 Hz). The VisSim setup is shown in Figure 7-48.

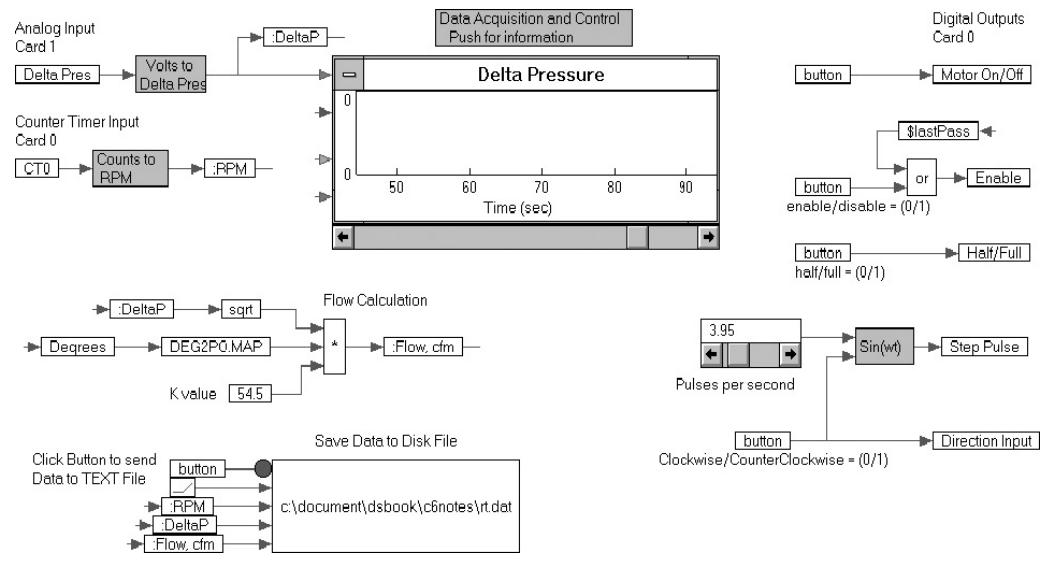
FIGURE 7-48 DATA ACQUISITION SIMULATION SETUP SCREEN



The VisSim model for the Data Acquisition application is shown in Figure 7-49.

Solution

The input signals are listed down the left side of the diagram and the output signals down the right side. The first input signal is an analog voltage input which reads the *Delta Pressure* sensor. The *rt-DataIn* block is called

FIGURE 7-49 DATA ACQUISITION DIAGRAM

DeltaPres in the diagram. We selected an Edwards Model 590 Pressure Sensor for measuring the delta pressure. The sensor outputs an analog voltage signal in the 0 to 10 V range. The scaling is linear over the range and 0 V corresponds to 0 T and 10 V to 1000 T. The second *rt-DataIn* input block is the counter timer, which is connected to a laser tachometer to measure the speed of the motor. For this application, we selected the LaseTach TEC 199LC tachometer. The unit is easy to mount, requiring that only a piece of reflective tape be adhered to the motor flywheel. The laser sensing unit can then be placed up to 30 ft away and, after some manual focusing of its beam on the tape, outputs a 0-5 V TTL pulse signal, which is connected directly to the GPIO counter timer (we used CT0, the first counter timer, however, the second one could also have been used) through a digital channel on the screw terminal. Access to counter timers in GPIO boards is usually through one of the card's digital I/O channels. The CBI universal screw terminal board is well labeled, and the layout is large, allowing easier connection of wires. According to the CBI manual, we attached the "high" line to pin 21 and the "low" line to the low-level ground (llgrnd) pin 7. The board provides several low-level ground pins, so you can use whichever one has the most convenient access. To set the counter timer up in the VisSim *rt-DataIn* block, the channel type is set to counter, and the channel number is set to 0 for CT0. Had we used the second counter timer, the channel number would be set to 1 for CT1.

Data that enters the VisSim diagram through a *rt-DataIn* block is in non-engineering units (such as volts, amps, or counts). Before using the data, it should be converted to engineering units. This operation, sometimes called scaling, is implemented as two "compound blocks" in our VisSim diagram. A compound block is simply a block with more blocks inside of it*. They are used to encapsulate detail in VisSim diagrams and save screen space. To enter a compound block, use the right button on the block. To return to the next higher level while in a compound block, use the right button on any "free" area of the screen. We look inside the two scaling blocks used in this application.

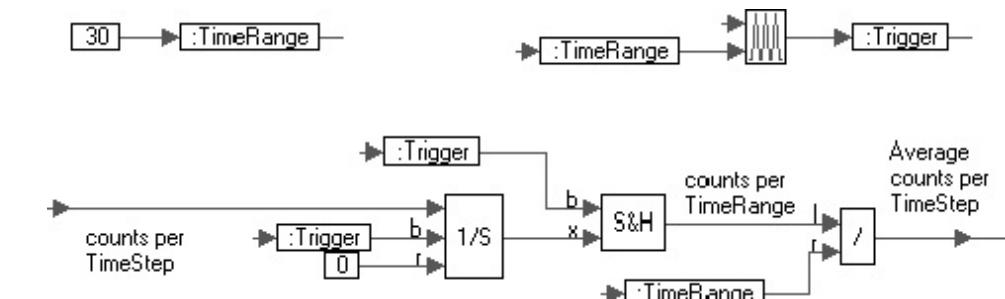
The first scaling block, named *Volts-to-Delta Press*, accepts a signal with the value of the sensor voltage in the 0 to 10 V range as input and produces a *Delta pressure* signal in units of inches mercury (in. Hg). The contents of this compound block are shown in Figure 7-50. Reading from left to right, the input signal is first converted from volts to torr using the linear scaling provided with the sensor (0 V = 0 T and 10 V = 1000 T). This is accomplished by multiplying the volt signal by 100. The VisSim "gain" block is used for this operation. A similar linear scale change is used to convert T to in Hg using a gain of 0.03937.

*VisSim provides unlimited depth in this hierarchy, eventually limited by RAM.

FIGURE 7-50 VOLTS-TO-DELTA PRESSURE COMPOUND BLOCK

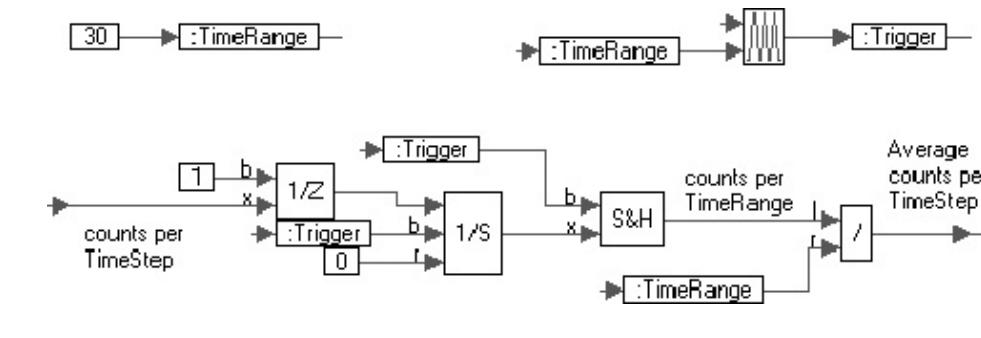
The second scaling block, called *Counts to RPM* converts the pulse input signal from the counter timer to RPM. Before this block is described, an explanation of counter timer operation is needed. When VisSim reads a counter timer signal into a diagram through a *rt-DataIn* block, the output of the *rt-DataIn* block is the number of counts that have occurred during that *TimeStep*. Once this value is read, the timer is reset to 0, and the process is repeated for the next *TimeStep*. The counter timer is of finite bitlength and will “wrap around” if you continue counting after the limit is reached. Generally, the count limit is 32,000 but does vary based on the card manufacturer. In this application, the nominal motor speed is 3600 rpm, and since the VisSim *TimeStep* was selected as 0.25 s, we would expect the counter to peak out at around 15 counts per *TimeStep*. To prevent wraparound, the *TimeStep* must never exceed $32,000/15 = 2133$ seconds. Sometimes it will be necessary to “smooth” the counter signal by collecting counts for longer periods of time. In this Data Acquisition application, we found it necessary to average counts over a 30 second period to achieve good rpm readings. We could have run the entire diagram at a *TimeStep* of 30 seconds to achieve this, but a 30 s update time would be too slow. The purpose of the *Counts to RPM* compound block is to emulate this 30 s update behavior in software while the diagram still runs at 0.25 s intervals. The contents of the *Counts to RPM* block are shown in Figure 7-51.

The use of variables greatly simplifies diagrams. Two variables are used in the *Counts to RPM* block: *TimeRange* and *:Trigger*. The colon preceding the variable name scopes the variable to the working screen. Variables can be made global by removing the colon. The *:TimeRange* variable is set to 30 (seconds) in the upper-left corner of the block. In the upper-right corner of the block, a variable named *:Trigger* is defined to be a “train” of unit height-zero width pulses with the time between pulses set to *:TimeRange* (30 seconds). The bottom half of the block does our averaging. The counts per *TimeStep* signal enters from the left and is integrated via the *1/S* block. This is a special integrator. Its state (accumulated value) can be reset externally by applying a 1 to its *b*-input (*b* means Boolean). The value it is reset (in this case 0) and is applied through the *r*-input (*r* means reset). Integration performs the same operation as accumulation or summing, so the output of the integrator is an accumulation of the counts sent to it. The integrator output will look like a saw tooth signal that increases linearly from 0 to some value over *:TimeRange* seconds (30 s). At the 30 s mark, the integrator is reset to 0, which forces its output to 0, and the process is repeated

FIGURE 7-51 COUNTS TO RPM COMPOUND BLOCK

for the next 30 seconds. The S&H block is a sample/hold block. It samples the x -input when the b -input (Boolean input) goes high (to 1). We use the sample/hold block to “snap” the integrator output just prior to 30 s and save it as the “counts per *TimeRange*” value. If this above diagram were run, it may not perform the intended function because it has no way of knowing that you want to snap the integrator output value *just before* the integrator is reset to 0. We need to supply this information to the diagram. A simple approach is to delay the *:Trigger* signal to the integrator by one *TimeStep* (0.25 s). In effect, this *advances* the sample/hold one *TimeStep* and the problem is solved. The modified *Counts to RPM* block is shown in Figure 7-52.

FIGURE 7-52 MODIFIED COUNTS TO RPM COMPOUND BLOCK



The $1/Z$ block is a unit delay block. It operates exactly as a sample/hold block, except it waits one *TimeStep* before outputting its value. Before using the *Counts to RPM* block in our real-time application, we should test it in simulated time. The counts per *TimeStep* input should be about 900 so we'll used a constant 900 value and add some random variation to it. We'll add a normally distributed random number with 0 mean and a standard deviation of 50 using the VisSim *Gaussian Random Generator* block. Since the counts per *TimeStep* must be an integer only the integer part of the random output is used. The test diagram is shown in Figure 7-53.

FIGURE 7-53 AVERAGE COUNT DIAGRAM

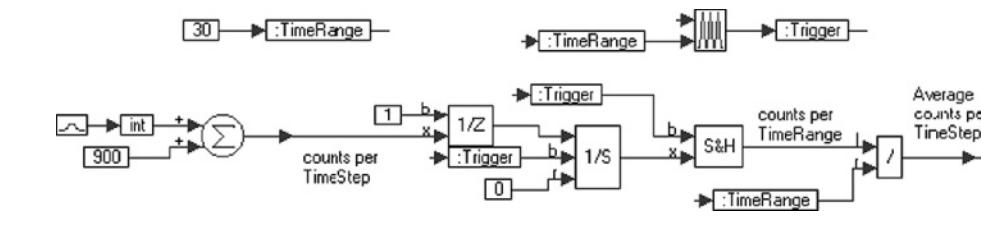
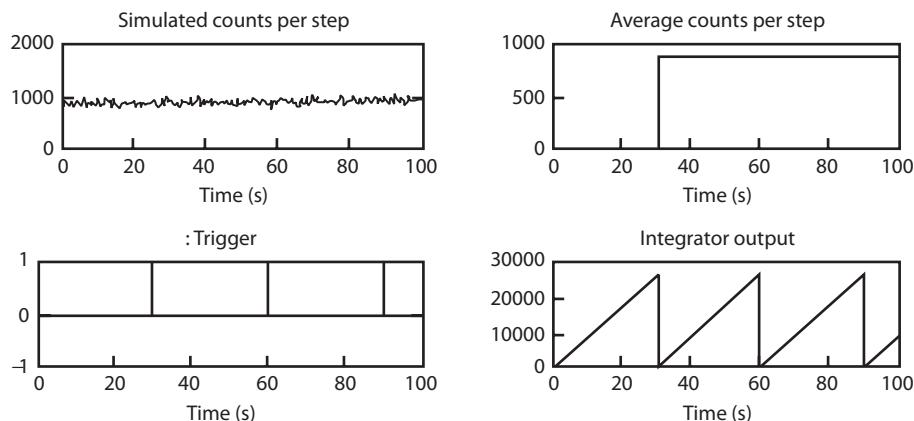
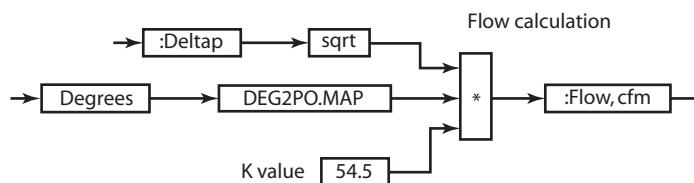


FIGURE 7-54 AVERAGE COUNT DIAGRAM SIGNAL RESPONSES

The behavior of four signals in Figure 7-53 is shown in Figure 7-54.

The process of converting the sensor signals from raw data to engineering units is now complete. The two input signals are available as the variables *DeltaP* and *:RPM* in the diagram. Of the five digital outputs, the last four used for the step motor control are identical to those discussed in the previous example. The top digital output is used to control the ON/OFF status of the motor. As mentioned, this output channel utilizes a solid-state relay to boost the low-power digital output signal from the TTL range (0 to 5 V) to the 0 to 230 V range required for powering the motor. The fact that a solid-state relay is being used makes no difference in our VisSim diagram.

On the left side of the screen, there is a calculation called *Flow Calculation*. This calculation uses the nonlinear relationship between the shaft angle of the stepper motor and the percent open to compute the airflow through the compressor. The calculation is shown in Figure 7-55.

FIGURE 7-55 FLOW CALCULATION PROCEDURE

This calculation uses three variables: *DeltaP*, and *:Flow, cfm* (which are both local variables by virtue of the colon) and *Degrees* (which is a global variable computed in the *sin(wt)* compound block used to generate the *Step Pulse* command to the motor). Calculation of the *Degrees* variable was covered in the last example on the step motor. The block entitled “*DEG2PO.MAP*” is a map (sometimes called a table lookup) routine. Map routines are based on an ordered table of input/output data. When an input is sent to the map, the corresponding output is returned. Linear interpolation and extrapolation is automatically applied to all maps in VisSim. The map data used for this application was entered using the Windows Notepad or Wordpad editor, and the contents of the map file is shown in Figure 7-56.

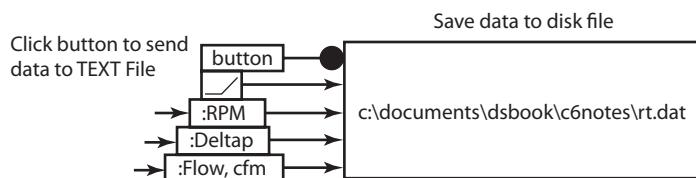
FIGURE 7-56 MAP FILE FOR DATA ACQUISITION EXAMPLE

Degree	% Open
-1	0
0	0
10	15.8
20	28.0
30	37.2
40	53.5
50	66.7
60	82.1
70	89.7
80	95.2
90	100
91	100

The first column contains the independent (input) variable information, and the second contains the dependent (output) variable information. VisSim map files also can be used for single input/multi-output and two input/multi-output applications. Map extrapolation is based on the endpoints of the map. You can force VisSim to not extrapolate the map by repeating the first and last rows with slightly out of range values for the independent variable. We have done this in Figure 7-56 by adding the -1° row and the 91° row.

The last issue that must be discussed is how we go about saving data to a disk from the VisSim diagram. In the lower-left corner of the diagram is a region entitled *Save Data to Disk File*. This region is shown in Figure 7-57.

FIGURE 7-57 SAVE DATA TO DISK PROCESS



The large rectangular block is a VisSim *Export* block whose function is to write the variables fed into it to a disk file (whose name is written on top of the *Export* block, in this case the name is `c:\document\dsbook\c6notes\rt.dat`) in a column format. We'll postpone discussing the "button" input for a moment and focus on the ramp (which is time in seconds) block and the `:RPM`, `:DeltaP`, and `:Flow, cfm` variables. Our VisSim diagram is executing every 0.25 s (*TimeStep* = .25) so each row of data will be written to disk at .25 sec intervals. In the first column will be the ramp value, which is time followed in columns 2, 3, and 4 by `:RPM`, `:DeltaP`, and `:Flow, cfm`, respectively. Frequently, we may not need or want to write data this often. You can control when data is exported using the optional Boolean input of the export block. This option is activated from

within the export block, so you right-click to get in the block and then select the *External Trigger* option. The blue circular Boolean input will appear on the export block, and you can use this to control your export of data. When the Boolean input is high (1), data is exported every *TimeStep* seconds, and when the Boolean goes low (0), no data is exported.

7.7 Summary

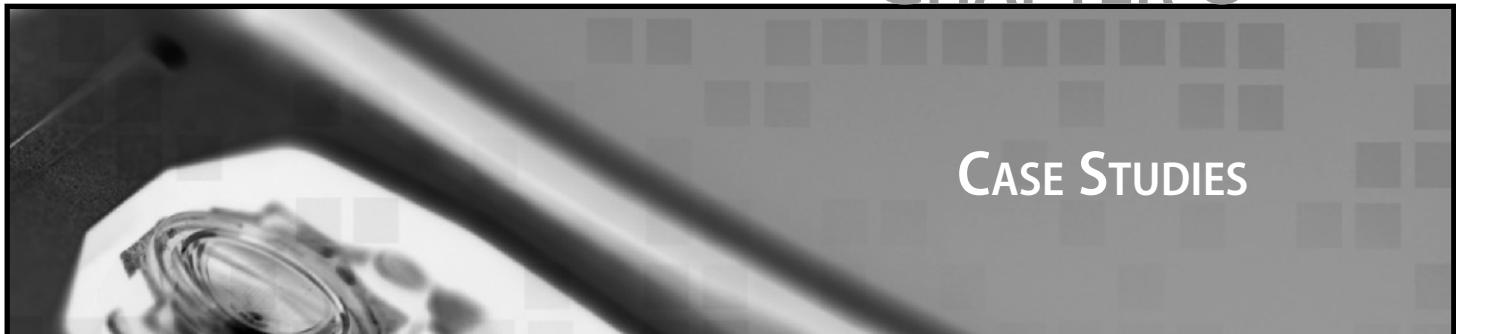
This chapter has presented a procedure for interfacing a computer with a real-world process using sensors, actuators, and a GPIO card. Because systems of this type are the basis for all mechatronics applications, several detailed examples were presented. These complete examples were possible to describe in this chapter because of the selected programming environment. This Windows-based environment used LabVIEW and VisSim software for programming the real-time application.

REFERENCES

- VisSim Users Guide*, Visual Solutions Inc., MA, 1995.
- Microsoft Windows Operating System Version 3.1 Users Guide*, Microsoft Corporation, 1993.
- Microsoft MS DOS Operating System Version 3.1 Users Guide*, Microsoft Corporation, 1993.
- Shetty, D., Campana, C., and Moslehpoor, S., “Standalone Surface Roughness Analyzer” *IEEE Journal of Instrumentation and Measurement*, March 2009, Vol. 58, No.3 pp 698–706.
- Shetty, D., Ramasamy, S., and Choi, S. “Non contact Visual Measurement System Integrating Labview with Matlab” *International Journal of Engineering Education*, Vol. 21, 2004.
- LabVIEW: Advanced Programming Techniques (II)* Rick Bitter, Motorola, Taqi Mohiuddin, Mindspeed Technologies, Matt Nawrocki, Motorola, Schaumburg, CRC Press, 2006.

CHAPTER 8

CASE STUDIES

- 
- 8.1 Comprehensive Case Studies
 - 8.1.1 Mass–Spring–Oscillation and Damping—an Example of Mechatronics Technology Demonstration
 - 8.1.2 Position Control of a Permanent Magnet DC Gear Motor (PM DC Gear Motor)
 - 8.1.3 Auto-Control System for Greenhouse Temperature
 - 8.2 Data Acquisition Case Studies
 - 8.2.1 Testing of Transportation Bridge Surface Materials
 - 8.2.2 Transducer Calibration System for Automotive Applications
 - 8.2.3 Strain Gauge Weighing System
 - 8.2.4 Solenoid Force–Displacement Calibration System
 - 8.2.5 Rotary Optical Encoder
 - 8.3 Data Acquisition and Control Case Studies
 - 8.3.1 Thermal Cycle Fatigue of a Ceramic Plate
 - 8.3.2 PH Control System
 - 8.3.3 De-Icing Temperature Control System
 - 8.3.4 Skip Control of a CD Player
 - 8.3.5 Simulation of Rocket Thrust Control in Laboratory
 - 8.3.6 Time Delay Blower
 - 8.4 Summary
 - References
 - Problems

This chapter presents a collection of case studies suitable for use in an educational laboratory or in an industrial setting. The case studies are divided into three sections. The first section contains three comprehensive case studies using the LabVIEW application software. The second section contains five case studies using the VisSim application software. The third section contains data acquisition and control case studies.

Each case study consists of a three part format which includes an *Overview*, *Parts List*, and the *Experiment*. The *Overview* states the objective of the case study and any other pertinent information. The *Parts List* includes all hardware necessary to duplicate the experiment (the PC and its software are not included). The *Experiment* section describes the real-time interface, development of the software application program, and results when available.

8.1 Comprehensive Case Studies

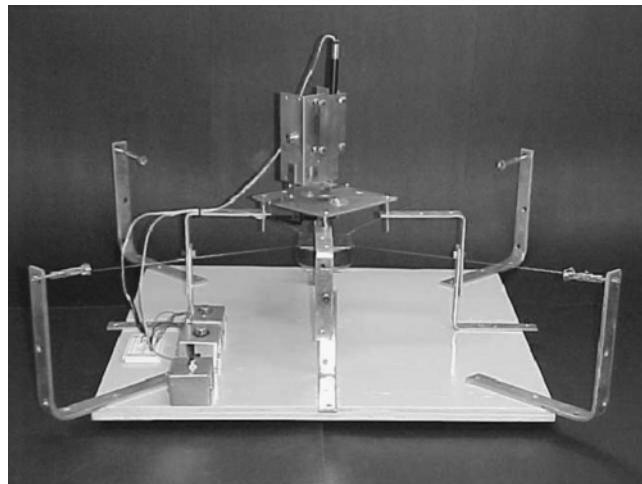
Three comprehensive case studies reinforce the mechatronics technology principles. The first example consists of a demonstration of a mass, spring, and damping system with simulation and control. The remaining two examples are that of sensing, monitoring, and control. The data acquisition system consists of four general components:

1. Sensors and power supplies
2. GPIO card, screw terminal panel, and connector cables
3. PC computer
4. Application software (i.e., LabVIEW, VisSim)

8.1.1 Mass–Spring–Oscillation and Damping—an Example of Mechatronics Technology Demonstration

Overview Many real-world systems can be modeled by the mass–spring–damper system and hence considering one such system the Mechatronics Technology Demonstrator (MTD) shown in Figure 8-1, it is discussed here as how to model and simulate a mass–spring–damper system, verify the behavior of the developed model with experimentally obtained data and develop a PID controller to control the position of the mass.

FIGURE 8-1 MASS–SPRING–OSCILLATION AND DAMPING, AN EXAMPLE OF A MECHATRONICS TECHNOLOGY DEMONSTRATION



Shetty and Kondo, University of Hartford.

MTD is a low cost technology demonstrator, developed and refined by the authors. It is a mass–spring–damper system with an electromagnetic force actuator and a non-contact position sensor. It is built from components available at most electronic, hardware, and home supply stores. It is suitable for studying the key elements of mechatronic systems including; mechanical system dynamics, sensors, actuators, computer interfacing, and application development. The MTD can be constructed in two configurations, vertical and horizontal. The vertical configuration as shown in Figure 8-1 offers greater motion control over shorter distances while the horizontal configuration provides just the opposite.

Regardless of the configuration, position of the mass in the MTD is measured using a position sensing detector (PSD) device. The PSD outputs a voltage proportional to the intensity of the light cast upon it. The light source, a laser similar to the type used for overhead presentations, is fastened to the base of the MTD and aimed at a mirror attached to the mass. The laser is adjusted until the reflected beam just hits the center of the PSD when the mass is motionless and in its normal position. As the mass moves around its normal position the reflection angle changes which, in turn, changes the area (intensity) of the light hitting the PSD and hence its voltage.

To provide force inputs to the mass for motion and/or active damping a voice-coil/magnetic actuator is used. The magnet is separated from the voice coil and attached directly to the mass. The voice coil is attached to the base of the MTD. Application of current to the coil results in either vertical or horizontal motion of the magnet depending on the orientation of the mounting.

Alignment of the magnet inside the voice coil is somewhat critical to reduce binding. Both the sensor and the voice coil actuator are connected to the PC based visual modeling and real time simulation application using a general purpose I/O card.

Parts List The hardware components required for this case study are presented in Table 8-1.

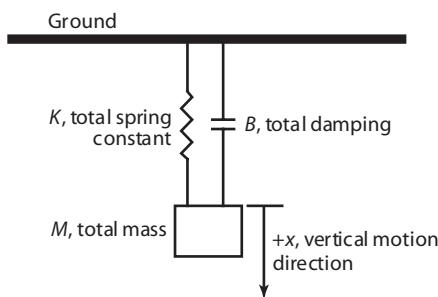
TABLE 8-1

1	Position Sensitive Detector (PSD)
2	Laser Pointer
3	Power Transistor
4	Power Supply
5	Braces and anchors
6	Laser Bracket
7	Aluminum Sensor housing

Experiment

Modeling and Simulation The physics based model for the mass-spring-damper system is based on the mechanical illustration shown in Figure 8-2.

FIGURE 8-2 MASS-SPRING-DAMPER MECHANICAL ILLUSTRATION FOR PHYSICS-BASED MODEL



From Newton's law of motion, we get

$$f = M\ddot{x} + B\dot{x} + Kx$$

Hence, the resulting transfer function of the model is

$$\frac{X(s)}{F(s)} = \frac{1}{Ms^2 + Bs + K}$$

where

x = displacement of the mass

f = force exerted on the mass

M = mass of the cart

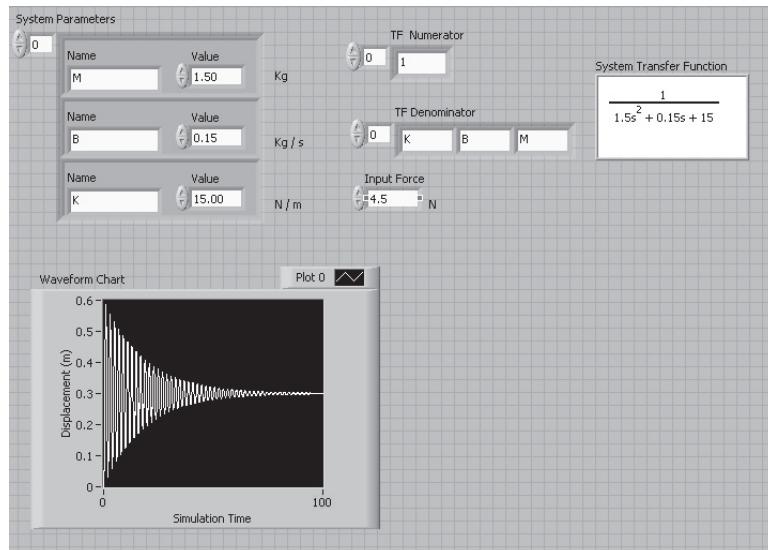
K = the equivalent spring constant

B = the equivalent damping constant

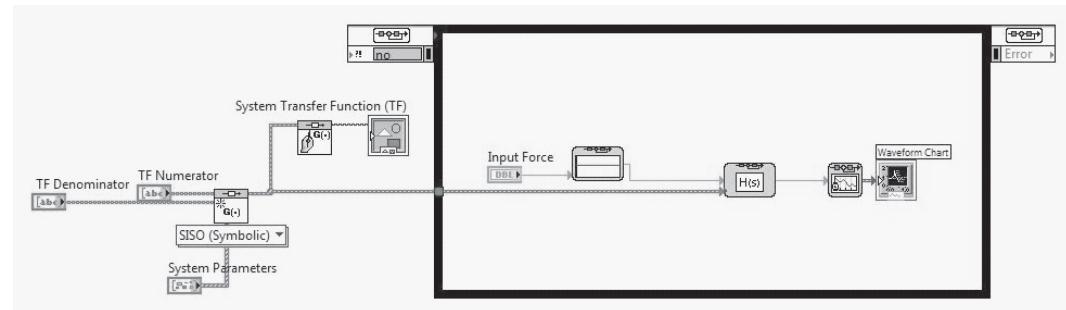
The response of the obtained physical model of the mass-spring-damper system was simulated using LabVIEW. Figure 8-3 shows the response of the open-loop system with following system parameters:

Mass:	$M = 1.5 \text{ kg}$
Spring coefficient:	$K = 15 \text{ N/m}$
Damping coefficient:	$B = 0.15 \text{ kg/s}$
Input force:	$f = 4.5 \text{ N}$

FIGURE 8-3 SIMULATION OF THE MASS-SPRING-DAMPER PHYSICS-BASED MODEL



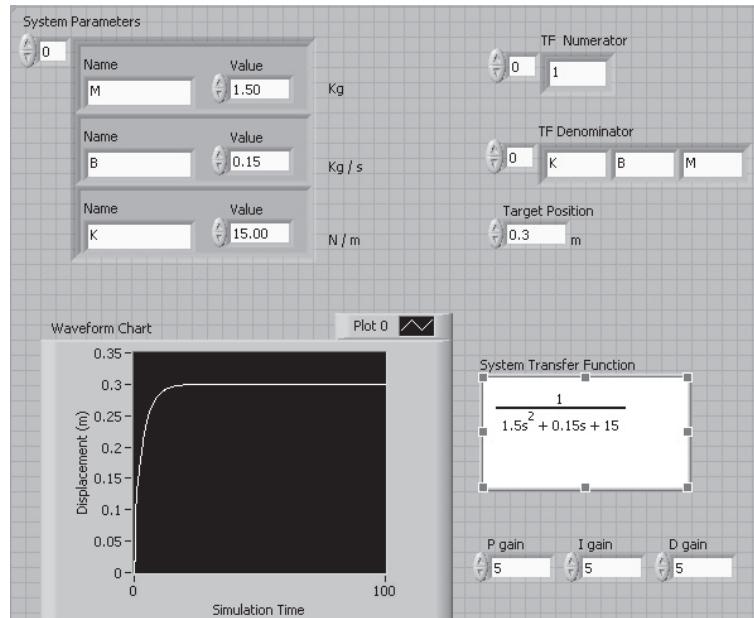
(a) Front Panel



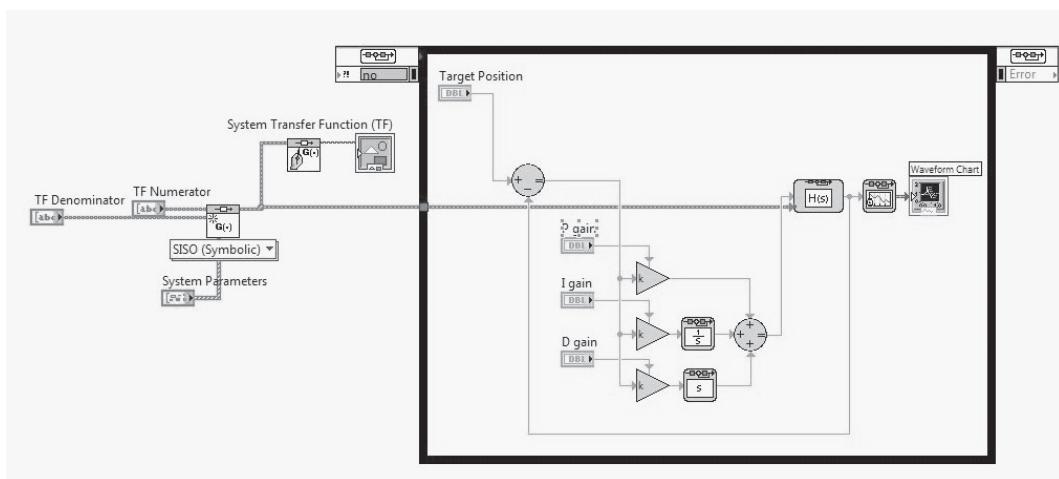
(b) Block Diagram

As seen from Figure 8-3(a), the mass oscillates between 0 and 0.6 m for an input force of 4.5 N for the initial 50 seconds and then settles down at 0.3 m. Though the system is stable there is lot of oscillation at the initial stage. Let us introduce a PID controller that will calculate the controlled input force for the system to place the mass at the target position. Figure 8-4(a) shows the response of the closed-loop system with following PID parameters.

FIGURE 8-4 SIMULATION OF THE CLOSED-LOOP SYSTEM



(a) Front Panel



(b) Block Diagram

Proportional gain = 0.5

Integral gain = 0.5

Derivative gain = 0.5

As seen from Figure 8-4(a), the mass reaches the target position in 15 seconds with the selected PID gains and the system parameters.

Real-Time Implementation of the MTD The real-time implementation of the MTD was done using the DAQmx vi, which provides shot response time without using much memory and is ideal for motion control applications. Figure 8-5 shows the block diagram of the implementation (*Hint:* For creating the input and output program refer to LabVIEW example problems: Help>Find Examples>Hardware Input and Output>DAQmx>Analog Measurements> Voltage, Help>Find Examples>Hardware Input and Output>DAQmx>Analog Generation>Voltage).

FIGURE 8-5 BLOCK DIAGRAM OF THE MTD REAL-TIME IMPLEMENTATION

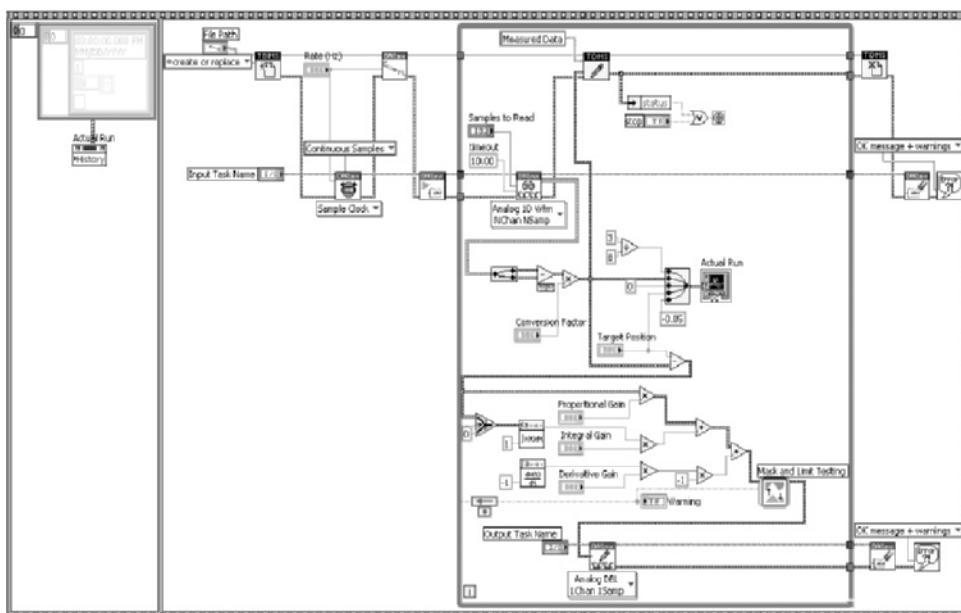
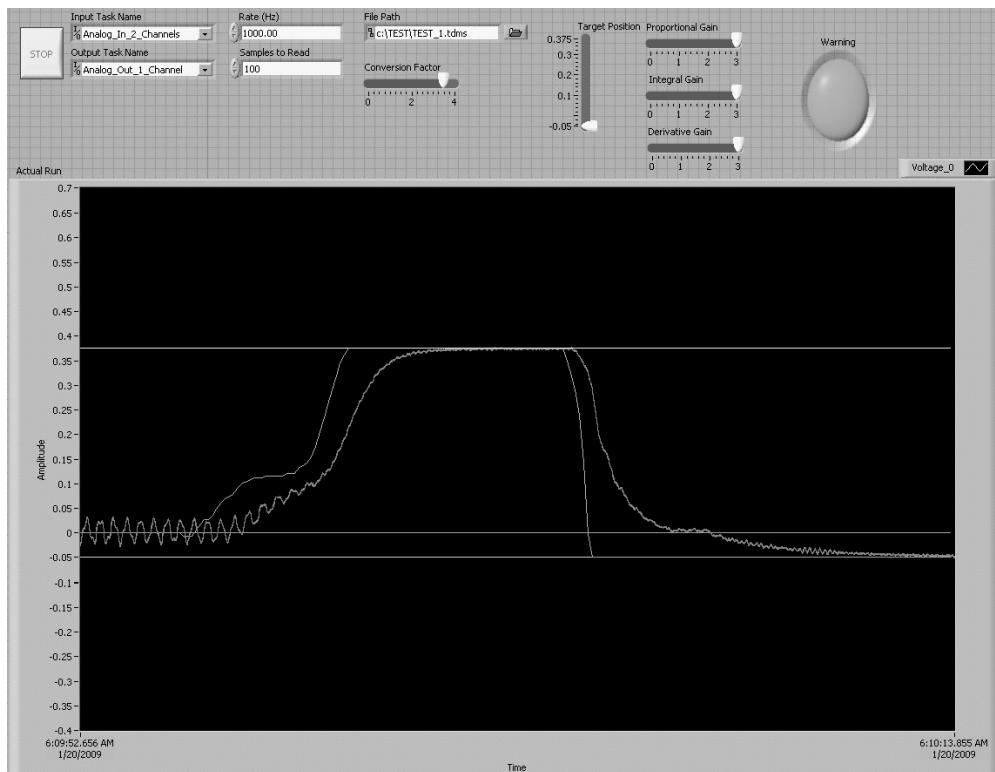


Figure 8-6 shows the front panel of the implementation where the output of the MTD (actual position of the mass colored red) is compared with the target path (desired position of the mass colored blue). (Note: Colors not shown here.)

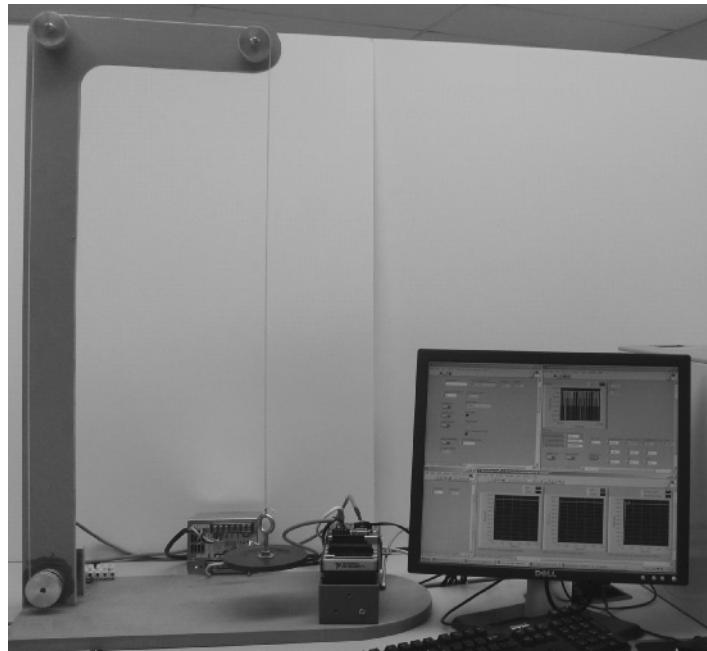
FIGURE 8-6 FRONT PANEL OF THE MTD REAL-TIME IMPLEMENTATION

8.1.2 Position Control of a Permanent Magnet DC Gear Motor (PM DC Gear Motor)

Overview Position control is a common task required in a wide variety of applications, such as those listed here.

1. **Transport Systems**—Conveyor belts, elevating platforms, building technology, Controlling gates, fans and pumps
2. **Machine Control**—Handling machines, feeding equipment, packaging machines,
3. **Material Handling**—Diverters, smart conveyors, gate and door
4. **Packaging**—Packaging machine, palletizers/de-palletizers

Figure 8-7 shows a PM DC gear motor position-control system, where the motor acts as the actuator lifting a mass and a Hall-effect sensor for measuring the rotational displacement of the motor shaft.

FIGURE 8-7 PM DC GEAR MOTOR POSITION-CONTROL SYSTEM

Thomas and Shetty.

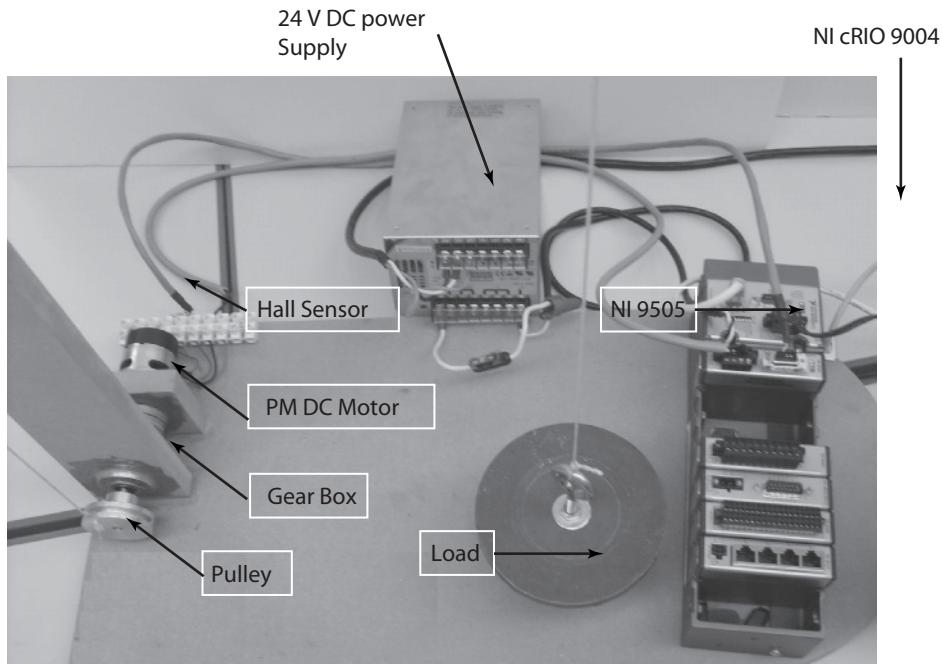
TABLE 8-2

1	24 V DC Power Supply
2	PM DC Gear Motor with inbuilt Hall Effect encoder
3	NI cRIO 9004
4	NI I/O module 9505
5	Misc. wiring, load, pulleys, string and mounting

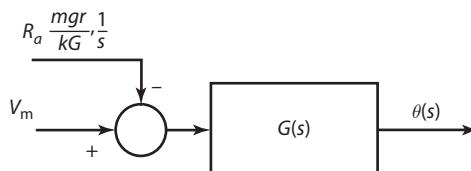
Parts List The hardware components required for this case study are presented in Table 8-2. and also depicted in Figure 8-8.

Experiment

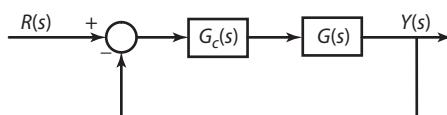
Implementation of the Dynamical System In Chapter 4, it was discussed how to get the transfer function for the PM DC gear motor or the open loop-system, which can be represented as shown in Figure 8-9.

FIGURE 8-8 HARDWARE DETAILS OF THE PM DC MOTOR POSITION CONTROL SYSTEM

Thomas and Shetty.

FIGURE 8-9 PM DC GEAR MOTOR SYSTEM OPEN-LOOP BLOCK DIAGRAM

In Chapter 6, it was shown how to design the PI controller for the position control of the PM DC gear motor with a negative unit feedback, as shown in Figure 8-10.

FIGURE 8-10 UNIT FEEDBACK SYSTEM

The previous discussions have focused on simulation of the dynamic systems. This example considers a real system with PI controller limited to ± 24 V. Figure 8-11 represents the block diagram and Figure 8-12 shows the front panel of this LabVIEW implementation.

FIGURE 8-11 BLOCK DIAGRAM FOR DYNAMICAL SYSTEM

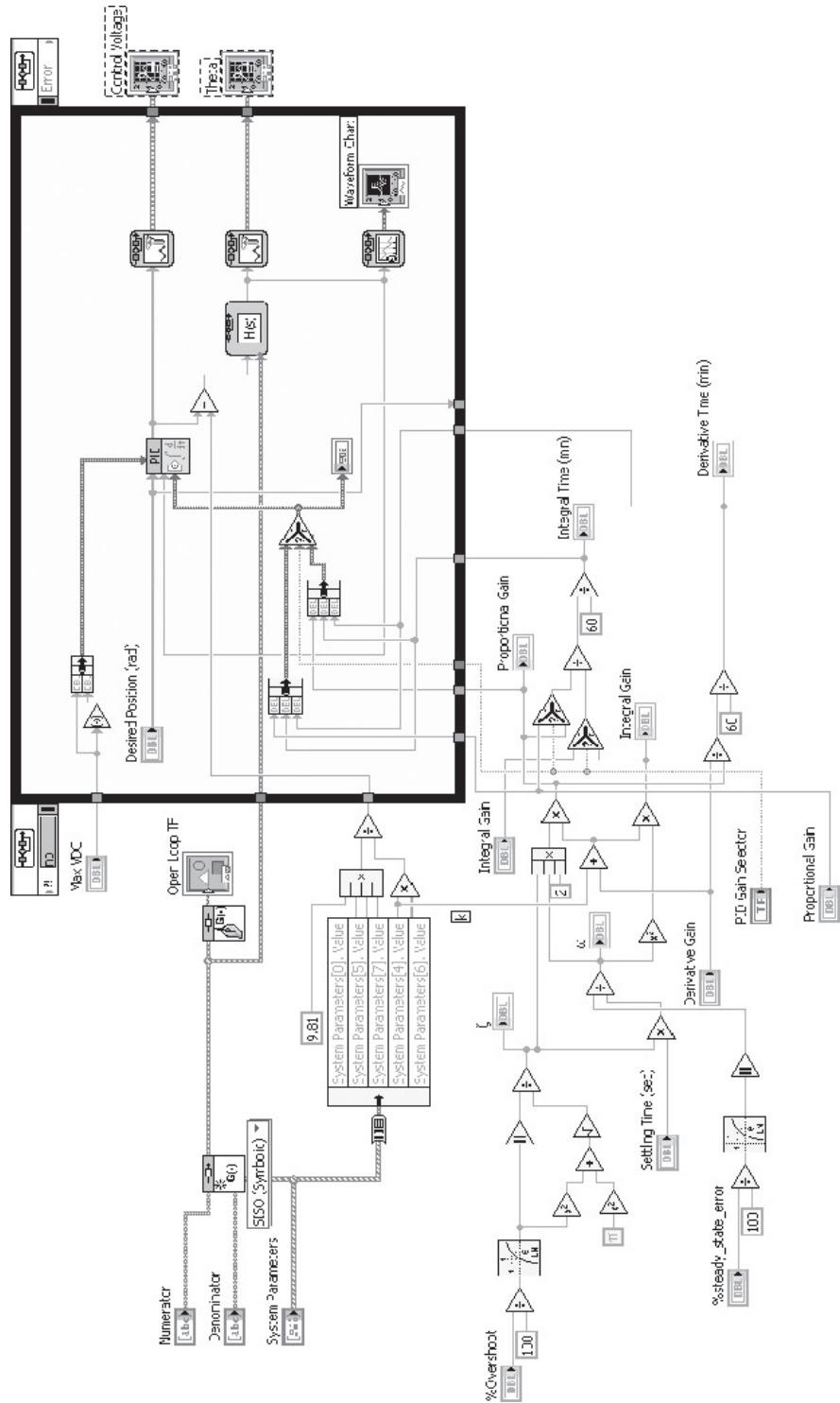


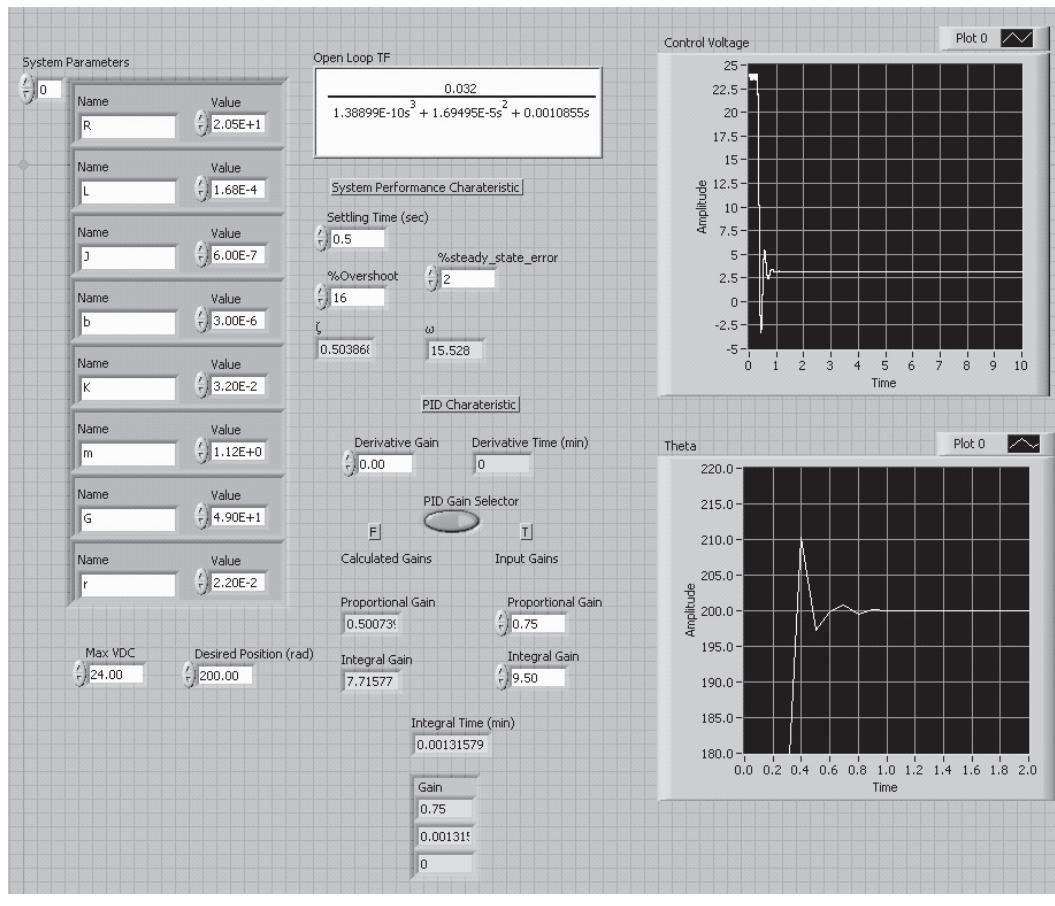
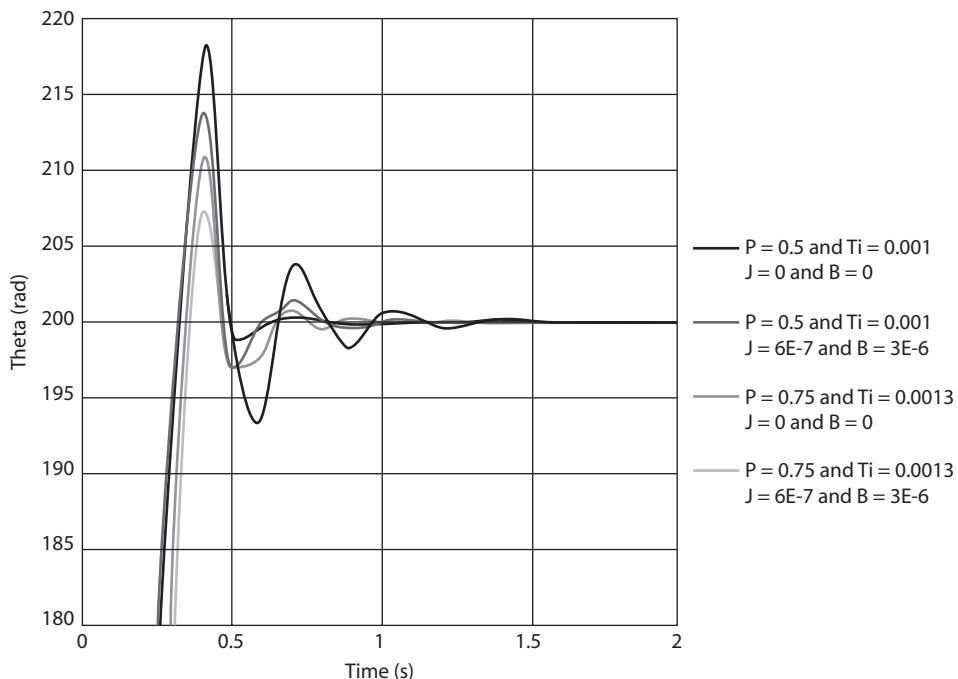
FIGURE 8-12 FRONT PANEL FOR IMPLEMENTATION OF THE DYNAMICAL SYSTEM

Figure 8-13 shows the comparison chart for the response of the dynamical system for the calculated PI values and the tuned PI values both when the inertia and damping losses in the system are neglected and also when small values of inertia and damping losses are considered. The unit for inertia ‘J’ is $\text{kg}\cdot\text{m}^2$ and for damping ‘B’ is $\text{kg}\cdot\text{m}^2/\text{sec}$. It is seen from Figure 8-13 that the response of the system with the calculated values of the PI gains are within the performance criteria when the inertia and damping losses in the system are neglected. However, it is also seen that for small values of the inertia and damping losses the response of the dynamical system does not meet the performance criteria. Hence, there is a need to tune the PI gains and with the tuned values of the PI ($P = 0.75$ and $I = 9.5$) the response of the dynamical system meets the performance criteria even when inertia and damping losses are considered.

In the next section we will see how to implement the real system and test the system response for the tuned values of the PI gains. The results of the analysis of the real system will be compared with the results of the dynamical system to check how close our mathematical model is to the real system.

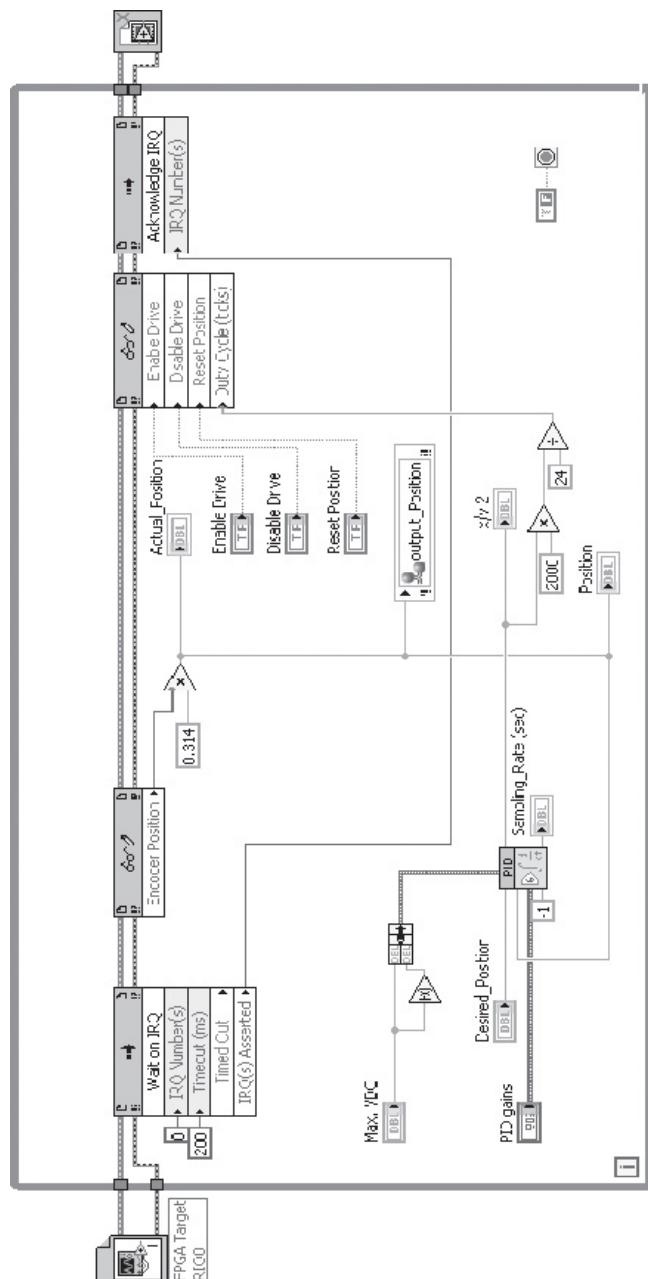
FIGURE 8-13 COMPARISON CHART OF THE RESPONSE OF THE DYNAMICAL SYSTEM

Implementation of the Real System The implementation of the real system involves signal conditioning of the Hall-sensor output signal and modulation of the power supply to the motor based on the PI controller output data. Both were covered in Chapter 7. It also involves creating the logic for the PI controller which will calculate the control voltage for the motor based on the desired motor shaft position and the measured shaft position. The logic for the controller was created for NI cRIO 9004 processor which is shown in Figure 8-14.

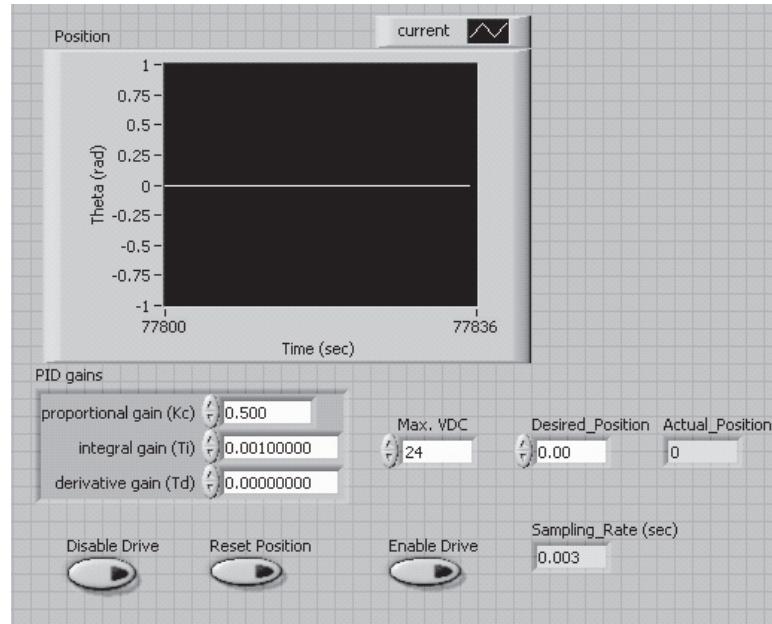
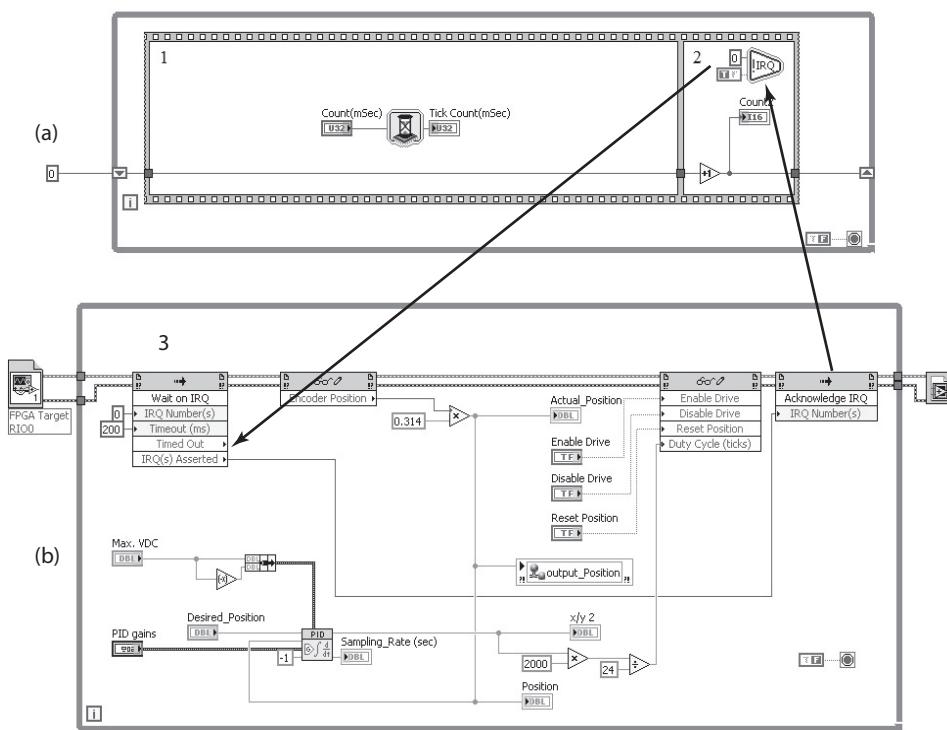
As seen from Figure 8-15 the encoder position which is counter output obtained from the FPGA logic for counting the motor shaft angular displacement is multiplied by a conversion factor 0.314 for converting counts to radians. This is because the inbuilt Hall sensor in use works with a 10-pole magnetic wheel attached coaxially to the motor shaft. Hence for one complete rotation of the motor shaft we get $10/2 = 5$ pulses and each pulse for quadrature encoder gives four counts. Therefore, every rotation of the motor shaft (i.e., 2π radian will give 20 counts). The measured angular position of the motor shaft is compared with the desired position, and based on the error, the PI controller will calculate the control voltage in the range \pm Max. VDC. This control voltage is multiplied by a conversion factor 2000/24 for converting voltage to duty cycle (ticks). Based on the duty cycle (ticks) the power supply to the motor is modulated.

Another factor to consider for our real time implementation is the need for synchronizing the PI output and the measured sensor output i.e. both should happen at the same time. Figure 8-16(a) shows the FPGA logic and Figure 8-16(b) shows the cRIO processor logic for synchronization. As seen from Figure 8-16, the loop 1 will wait for the time specified by the variable *Count* before

FIGURE 8-14 BLOCK DIAGRAM OF PI CONTROLLER



passing execution to the loop 2. In loop 2, we have the interrupt *IRQ*, which will pass the execution to the loop 3 and wait for an acknowledgement from that loop. In loop 3, PI calculation will be performed based on the value of the motor position that was send along with the interrupt and at the same time pass on the PI value from the previous calculation and an acknowledgement to the

FIGURE 8-15 FRONT PANEL FOR IMPLEMENTATION OF THE PI CONTROLLER**FIGURE 8-16** SYNCHRONIZATION OF THE INPUT AND OUTPUT DATA

interrupt in loop 2. Once the acknowledgment is received by the *IRQ*, execution will be passed to loop 1. Thus, the variable count will decide the loop rate (sampling rate) for the system.

The complete block diagram of the FPGA logic is given in Figure 8-17, and the front panel for the real-time implementation is shown in Figure 8-18.

FIGURE 8-17 BLOCK DIAGRAM COMPLETED

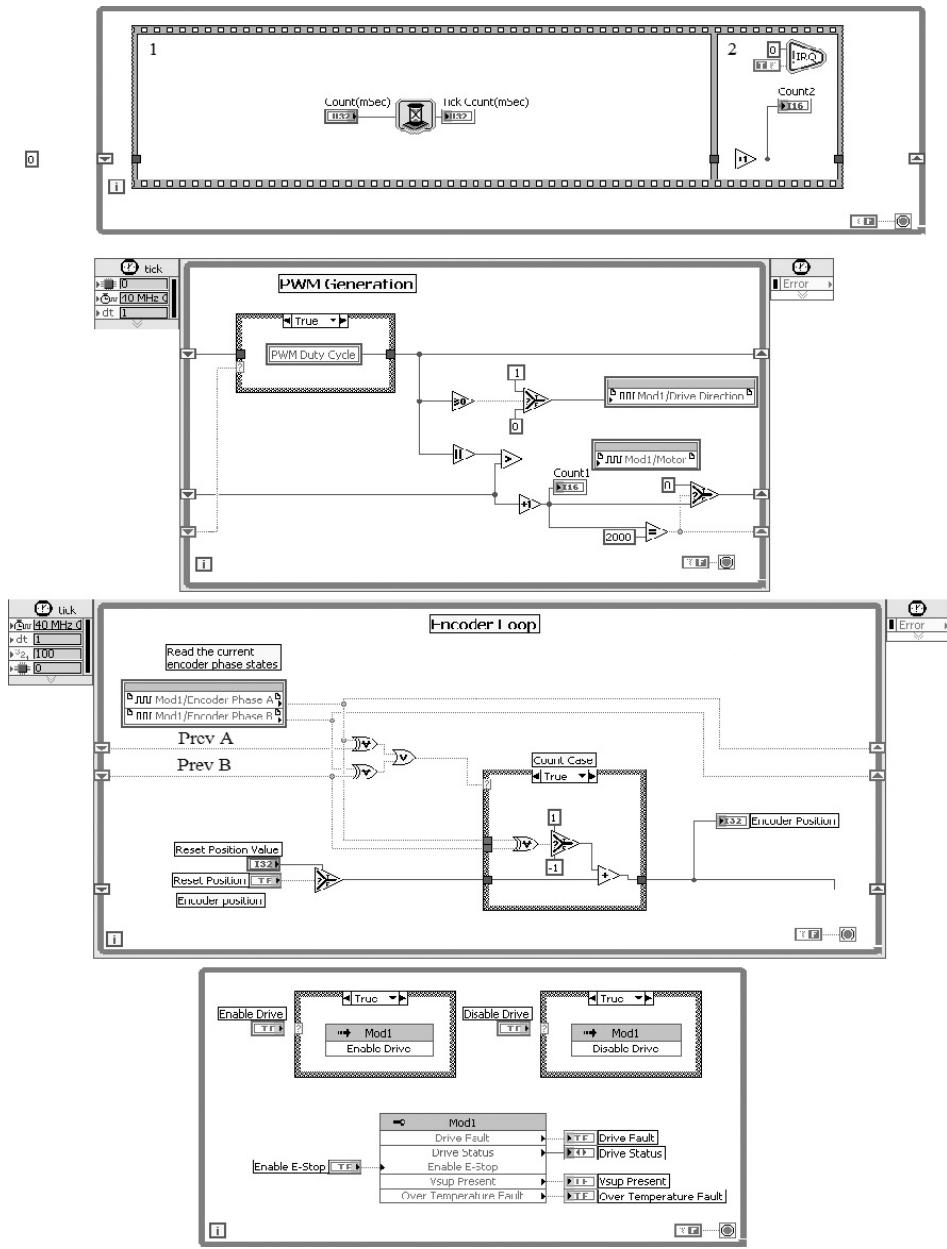


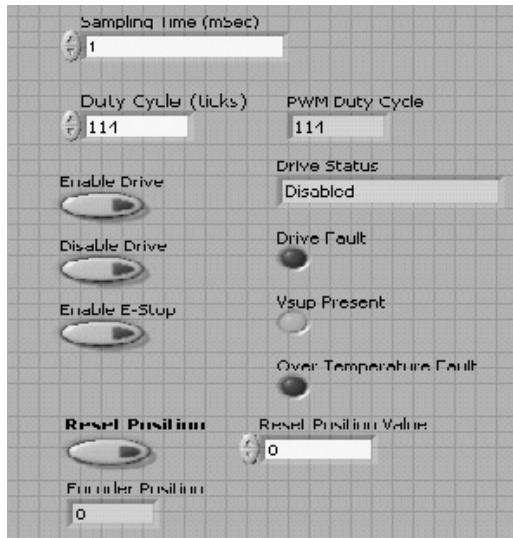
FIGURE 8-18 FRONT PANEL OF FPGA LOGIC FOR THE REAL-TIME IMPLEMENTATION

Figure 8-19 shows the comparison chart for the response of the real system for the calculated PI values and the tuned PI values.

As seen from Figure 8-19, the response of the real system for the calculated PI values has a damping ratio less than 16%. However, the settling time is 1.3 s. The response of the the real system for the tuned PI values meets all the performance criteria. Figure 8-20 shows the comparison chart for the response of both real system and dynamical system with small values of inertia and damping losses for the tuned values of the PI.

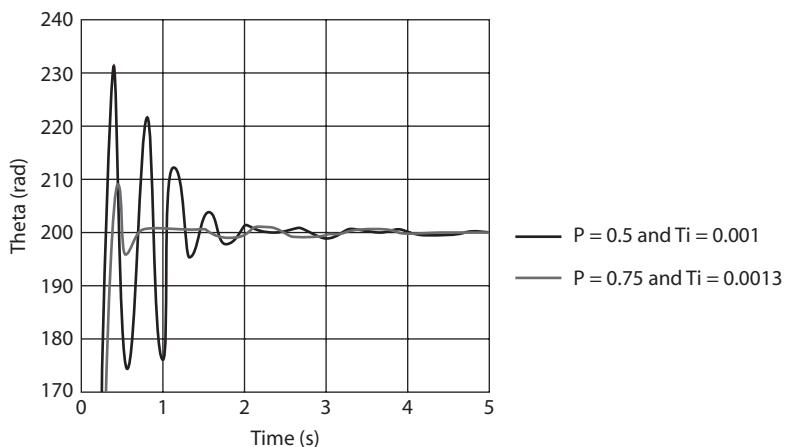
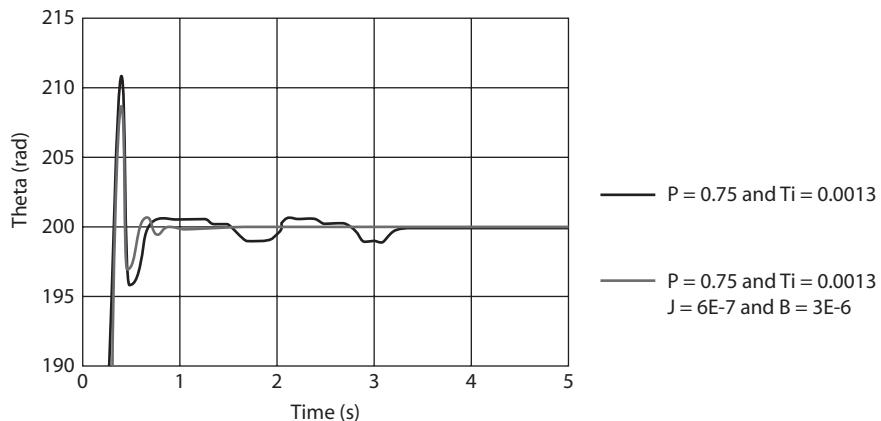
FIGURE 8-19 COMPARISON CHART FOR THE RESPONSE OF THE REAL SYSTEM

FIGURE 8-20 COMPARISON CHART FOR THE RESPONSE OF THE REAL SYSTEM AND DYNAMICAL SYSTEM



The comparison chart shows that the response of the dynamical system with small values of inertia and damping losses is very close to the response of the real system. The couple of the small oscillation in the response of the real system is due to the vibration in the string when the load is lifted.

8.1.3 Auto-Control System for Greenhouse Temperature

Overview The purpose of the project is to develop an automatic control system that will ensure that the temperature within the greenhouse (Figure 8-21) is always maintained at a specified value.

FIGURE 8-21



Campana, Dinan and Hamilton.

An electric light source is used as a heat source, a variable speed fan regulates the temperature within the greenhouse, and a thermocoupler monitors the temperature. A PID controller developed using LabVIEW is used as the automatic control for the system. The heat source is switched on and an internal temperature for the green house is selected. As the temperature rises and passes the desired temperature, the variable speed fan is switched on; this reduces the internal temperature. As the internal temperature approaches the desired temperature, the fan speed is reduced by use of a PID controller. A manual override for the system is included, and as an added precaution a Boolean logic circuit will switch off the heat source if the fan is unable to reduce the internal temperature of the greenhouse effectively.

Parts List

TABLE 8-3

1	60 watt Light Bulb
2	Omega K-type Thermocoupler
3	Computer Fan
4	Papst Oyster 2000

Background An effective greenhouse has a tightly controlled temperature. If the temperature is too warm or too cold, the plants will not survive. Therefore, it is important to tightly control the temperatures present inside the greenhouse. The purpose of this project is to use a model of a greenhouse and simulate the temperature inside with a light bulb. A variable speed fan will be used to control the temperature.

The research on temperature effects on plants indicates that plants produce maximum growth when exposed to a day temperature that is about 5 to 8°C higher than the night temperature.

The objectives of the project are to:

- Develop a system which is able to precisely control the internal temperature of a greenhouse with minimal overshoot.
- Develop the system in such a way that human intervention and operation of the system are possible but not frequently necessary.
- Monitor the internal temperature of the greenhouse and tightly control deviations from the chosen internal temperature of the greenhouse.
- Select an input voltage for the variable speed fan appropriate to the amount of cooling necessary.
- By use of Boolean logic, cut power to the heat source if the temperature within the greenhouse rises at a rate greater than the fan is able to regulate.

Experiment

Mechanical System Assembly and Validation A prototype of the greenhouse was built using Plexiglas, wood and medium density fiberboard. While constructing the prototype, careful provision had to be made to house the fan and heat source (bulb). This is shown in Figure 8-22.

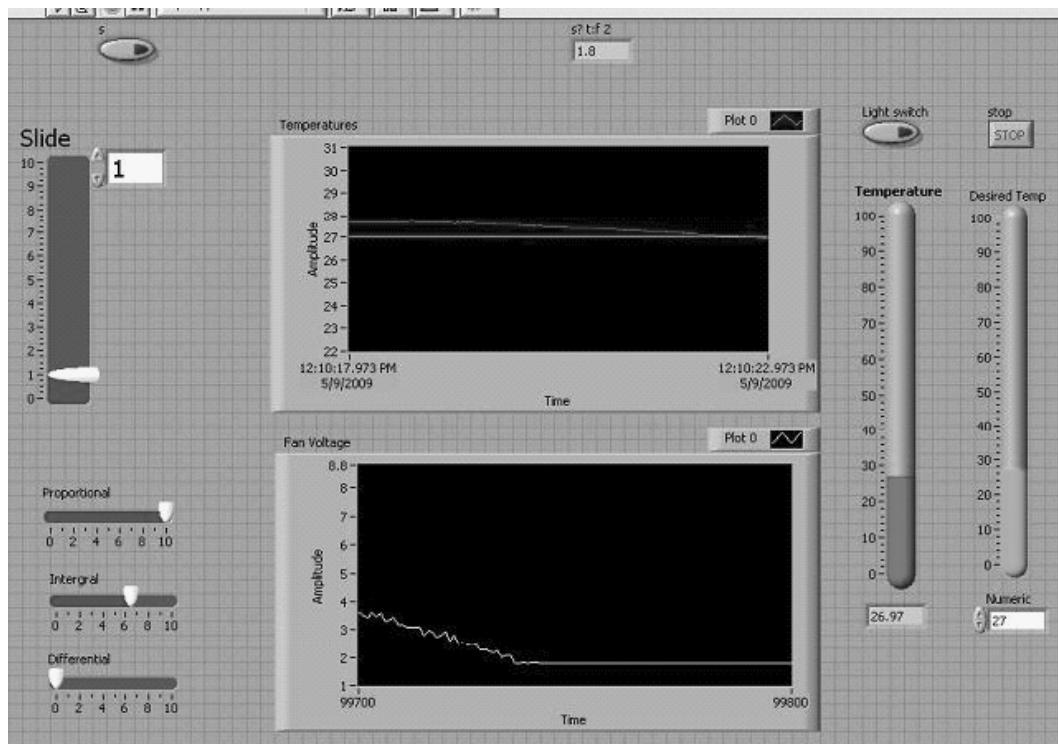
FIGURE 8-22 PHYSICAL MODEL OF GREENHOUSE



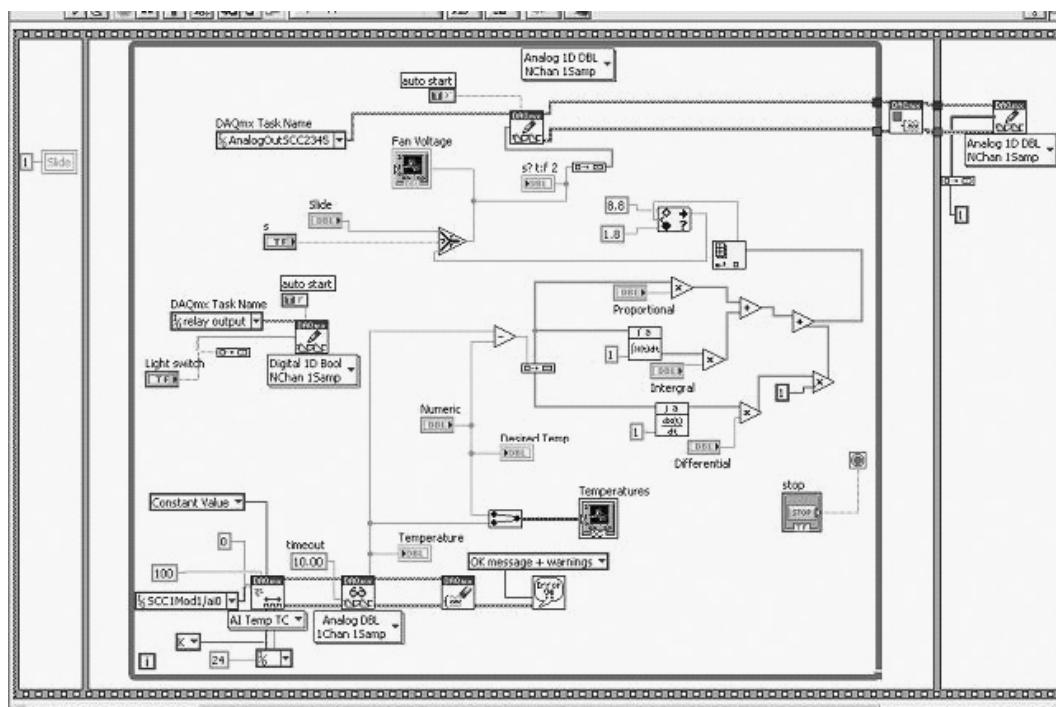
Campana, Dinan, and Hamilton.

The thermocouple was plugged into the system and calibrated. The thermocouple used is a K-type, and the tip of the thermocouple is the part that actually detects the heat. The tip of the thermocouple needs to be oriented in such a way that it gets direct heat from the light bulb and not reflections. The light bulb and fan are part of the experiment. The fan gets actuated so that when the temperature goes above the desired temperature, it provides cool air. Once the air temperature is reduced to the desired temperature, the fan turns off. This process repeats itself.

Results As is shown in Figure 8-23, the control system is able to track the internal temperature of the greenhouse and effectively operate the variable speed fan to keep the temperature at or near the desired temperature. Once the initial peak in temperature is reduced, the operating temperature of the system is maintained within 1°C. An image of the LabVIEW program front panel is shown in Figure 8-24. The measured temperatures were typical best-case scenarios with percent error for overshoot and steady-state operation being approximately 2.96% and 0.1%, respectively.

FIGURE 8-23 FRONT PANEL FOR TEMPERATURE IN GREENHOUSE

Analysis and Conclusion The results show that the experiment performed as per the desired specifications. The fan was able to cool the greenhouse to the temperature that was set initially. The lower graph in the screen shows the voltage input to the fan. If the temperature of the greenhouse does not reach a steady state, the temperature would rise above the level and the fan would turn on. Depending on the response of the fan, the fan would stay on long enough maintain the green house set temperature. The system is successful within acceptable error. The PID controller requires tuning for each application. As the volume of the greenhouse increases, there is an increase in the overall inertia of the system. In such cases, the volume of air is higher and it takes additional effort to control. The increased mass of air within the system will also reduce the responsiveness of the system to temperature changes. For more precise control of the system, it is necessary to add additional features to the PID controller. The controller is used to monitor the rate of change of internal temperature of the greenhouse and provide an estimate the time so as to reduce the overshoot.

FIGURE 8-24 BLOCK DIAGRAM FOR GREENHOUSE

8.2 Data Acquisition Case Studies

This section presents several experiments which illustrate various uses of data acquisition. They are all monitoring experiments and the input data is used for either display, calibration, or other processing purposes. Each of the data acquisition systems consists of four general components:

1. Sensors and power supplies
2. GPIO card, screw terminal panel, and connector cables
3. PC computer
4. Application software (i.e., windows, VisSim or LabVIEW)

No actuators are present in a data acquisition application. All case studies are supplied with a *Parts List*. Cost was an important issue during the part selection process in most of the case studies; therefore, most parts were selected from high volume equipment distributors.

8.2.1 Testing of Transportation Bridge Surface Materials

Overview The purpose of this project is to measure the strength of materials used for beams in transportation bridges. Wooden members treated with creosote have been used for decking in small bridges for many years. On some of these bridges, blacktop has been placed over the wood to create a permanent and smooth driving surface. The strength testing of these materials is difficult because their rough surface finish provides a poor bond for sensor mounting. This experiment investigates the use of strain gauges mounted on steel supports attached to the wooden beams to measure deflection.

Parts List The hardware components required for this case study are presented in Table 8-4. The computer software, Windows, and VisSim/RT, are not included in this list and are assumed to be installed on the application PC.

TABLE 8-4

1	2100 System Measurements Group, Inc. eight-channel strain gauge and amplifier system
2	DC power supply
3	Miscellaneous wiring and mounting
4	Advantech 812PG GPIO card and screw terminal

Experiment The sample beam used for testing was a composite beam constructed by bolting six individual boards together. Strain gauges were mounted on the bottom surface. Due to the rough surface characteristic of the composite beam, the strain gauges were bonded to “dogbone” shaped steel strips with epoxy. After the epoxy hardened, the wire leads were soldered to each gauge.

Amplification and balancing of the strain gauges was achieved by an eight-channel strain gauge conditioner. This device accepted strain gauge output and after amplification and conditioning (using internal bridge circuitry) produced a 0 to 600 mV output signal that was sent to an analog input channel of the GPIO card for use in the application program.

Once the data link was established to the application program, a calibration was performed to convert the analog input voltage to engineering units. For this calibration, a range of loads was applied to the top of the composite beam. The loads were applied to the center of the beam by an Instron testing machine and ranged from 0 to 90 kN. The readings in Table 8-5 were recorded.

TABLE 8-5 CALIBRATION RESULTS

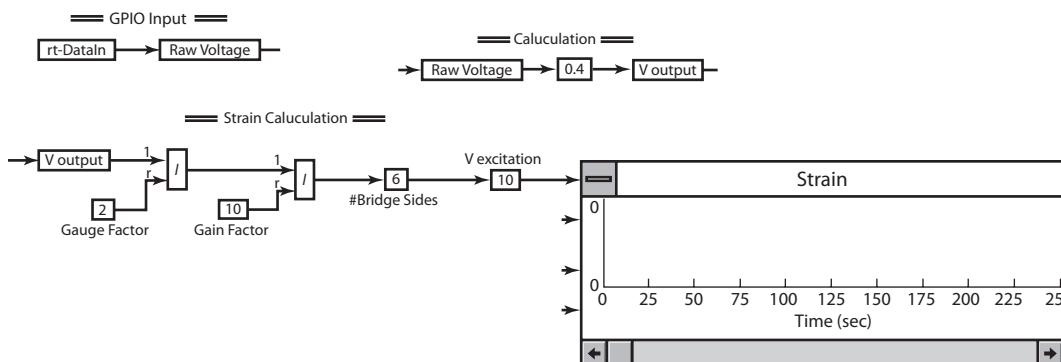
Strain Gauge Output, Volts	VisSim rt-DataIn Value
-.4	-.16
-.2	-.08
0	-.002
.2	.078
.4	.161

A scaling factor of .4 was chosen. Multiplying this scaling factor by the VisSim *rt-DataIn* value converts the value to that read by the strain gauge, V_{output} . The strain is then computed according to the following formula.

$$\epsilon = \frac{1}{\text{Gauge Factor}} \cdot \frac{1}{\text{Gain Factor}} \cdot (\# \text{BridgeSides}) \cdot (V_{\text{output}}) \cdot (V_{\text{excitation}})$$

The VisSim application program is presented in Figure 8-25.

FIGURE 8-25 TRANSPORTATION MATERIAL TESTING—APPLICATION DIAGRAM



The *rt-DataIn* block configuration was set for an analog channel in the ± 1 volt range. The 812 card was necessary for this application due to its programmable range for analog input channels. The *TimeStep* used was 0.25 seconds, and the time range was set to one hour. A 386 SX PC was used in this experiment.

The major problem encountered in this experiment was producing repeatable results. The metal pieces used to mount the strain gauges acted as antennas and received noise from the Instron test machine used to flex the beam. Shielding and proper grounding reduced this effect, but it was still a concern.

8.2.2 Transducer Calibration System for Automotive Applications

Overview The purpose of this project is to investigate a method for automatic calibration of the motion transducers used to sense exhaust valve motion on diesel engines in test cells. Current calibration methods are performed manually by measuring the valve motion with a dial indicator and entering the output voltage of the transducer which measures distance into the data collection system. This process is inaccurate, time consuming, and susceptible to operator error.

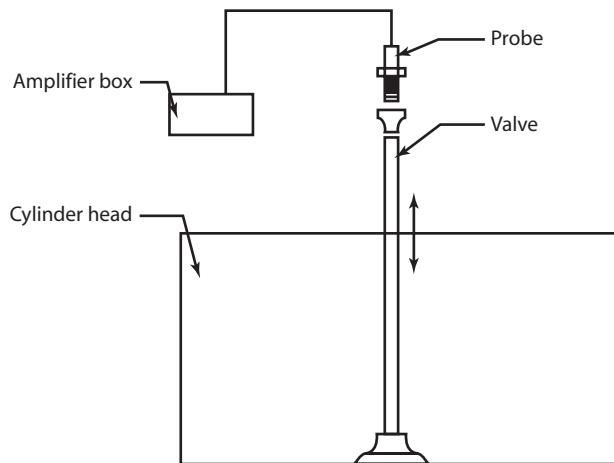
There are two objectives in this study. The first objective will establish the calibration curve for the test cell motion transducer. This transducer is a Bently Nevada 7200 Series 5-mm Proximity Transducer. The second objective will define requirements for using this system while the engine is being slowly rotated.

Parts List The hardware components required for this case study are presented in Table 8-6. The computer software, Windows and VisSim/RT, is not included in this list and is assumed to be installed on the application PC.

TABLE 8-6

1	Bently Nevada 7200 Series 5mm Proximity Transducer
2	DC power supply
3	Misc. wiring and mounting
4	Advantech 812PG GPIO card and screw terminal

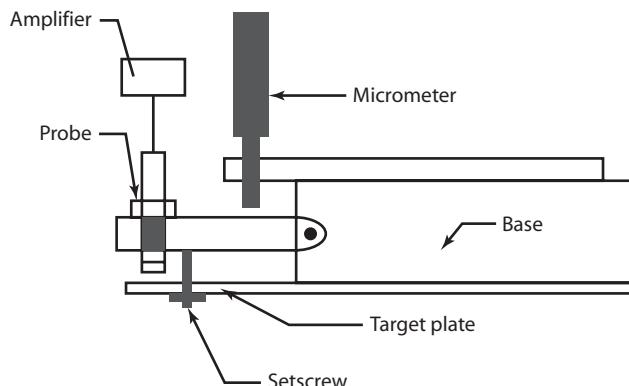
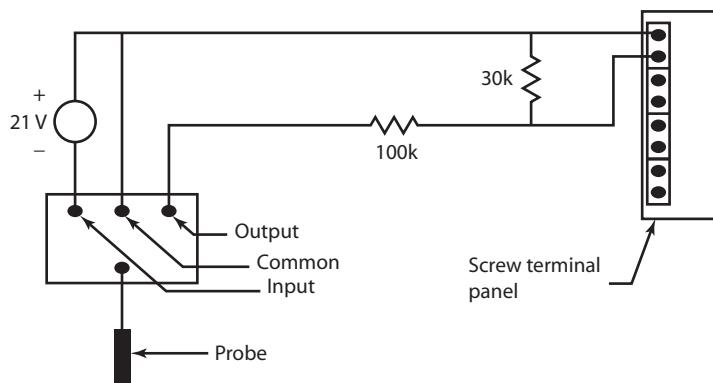
Experiment The first objective of this study is the calibration of a Bently Nevada 7200 Series 5-mm Proximity Transducer using a data acquisition system. This transducer operates on the eddy current principle. The amplifier box connected to the transducer produces a radio frequency signal that is radiated through the probe tip into the observed material setting up eddy currents. The loss of energy in the return signal is detected by the amplifier box and conditioned for linear display. Figure 8-26 illustrates a typical installation.

FIGURE 8-26 TYPICAL TRANSDUCER INSTALLATION

The transducer (probe) is mounted to a spacer that rests on top of the cylinder head. It is aimed at a target soldered to the top of the valve. A test fixture which models the transducer installation was built for use in this experiment. The test rig is presented in Figure 8-27.

The test rig consists of a pivoting bracket which holds the transducer (probe). The probe is aimed at the target plate. The probe vertical position is manually adjustable with a setscrew. A micrometer is used for vertical measurement. The amplifier output ranged from 0 to 22 V which was too large for input to the GPIO card, so a voltage divider was used to reduce the voltage to the 0 to 5 V range. This circuit and the connections with the probe and the GPIO card are shown in Figure 8-28.

Once the circuit was connected to the computer the application diagram was built and the calibration process was initiated by moving the micrometer in 0.25-mm increments and recording the *rt-DataIn* voltage signal in VisSim to a text file (see Chapter 6 Notepad file examples). The file was later edited to include the micrometer distance information, and the calibration process was complete.

FIGURE 8-27 TRANSDUCER TEST RIG**FIGURE 8-28 CIRCUIT DIAGRAM**

Noise problems did not affect this test due to the high voltage values and the use of the voltage divider circuit. Higher sampling rates needed for a slowly rotating engine could not be achieved with this hardware.

8.2.3 Strain Gauge Weighing System

Overview The purpose of this project is to apply the strain gauge to a weighing application. A scale, similar to the type used in your home, will be constructed in this experiment using four strain gauges, two metal beams, a GPIO card, PC, and software. To reduce costs a commercially available strain-gauge-based scale was purchased and disassembled, the parts were then used in this experiment.

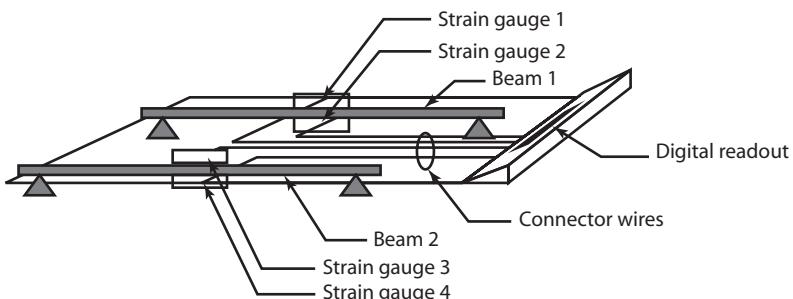
Parts List The hardware components required for this case study are presented in Table 8-7. The computer software, Windows and VisSim/RT, are not included in this list and are assumed to be installed on the application PC.

TABLE 8-7

1	Councilor Digescale D300 Electronic digital scale
2	DC power supply
3	Misc. wiring and mounting
4	Advantech 812PG GPIO card and screw terminal

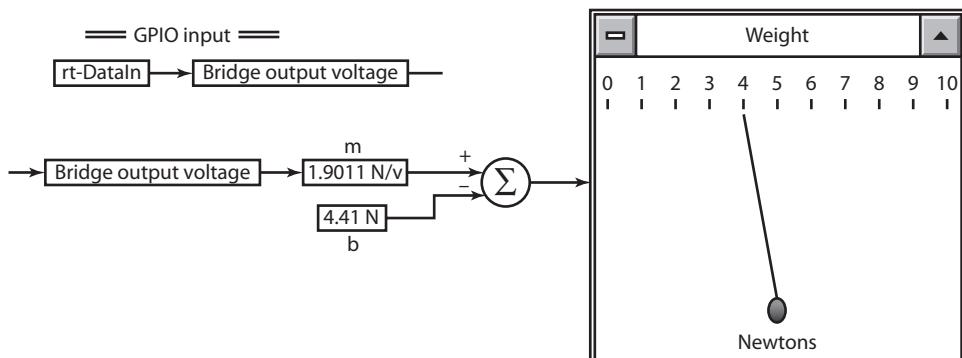
Experiment The strain gauge is a sensor which consists of a wire bonded to a mounting platform. As a tension or compression force is applied to the platform the wire length and resistance varies. To reduce noise, strain gauges are usually configured in a Wheatstone bridge circuit. The voltage output of such a circuit is in the millivolt range and can be related to the force (weight) applied to the strain gauges in the bridge. This relationship is usually determined experimentally and the resulting calibration is used thereafter.

This project is based on the Councilor Digescale D300 Electronic digital scale which utilizes a strain gauge for weight measurement. The scale was disassembled, and the strain gauge output was sent to the GPIO card as an analog voltage input. An application program was then written to convert the millivolt input signal to a weight value (based on a calibration relationship) and the resulting weight value was displayed on a dial meter in the application program. A diagram of the Councilor Digescale D300 Electronic digital scale is presented in Figure 8-29.

FIGURE 8-29 COUNCelor DIGISCALE D300 ELECTRONIC DIGITAL SCALE

The four strain gauges in the D300 scale are connected in a bridge configuration powered by a 9 volt battery. The bridge output voltage is then converted to a weight and displayed on the digital readout. In this experiment the connector wires from the four strain gauges were severed and configured as a Wheatstone bridge circuit powered by a 10 V power supply. The bridge output was then attached to an analog input channel to the 812PG card. The 812 card with software programmable gain was selected over the less expensive 711 card due to the millivolt range of the input signal. The 812 card has four software selectable gain ranges allowing ample resolution for millivolt input signals in the lowest gain range.

Initially the calibration table was constructed using another D300 scale as a reference. The calibration was linear of the form $y = m \cdot x + b$. The application program which includes this calibration is presented in Figure 8-30.

FIGURE 8-30 STRAIN GAUGE WEIGHING SYSTEM APPLICATION DIAGRAM

The VisSim TimeStep was selected as 0.1 s and the time range set to 1000 s. When sensing signals in the millivolt range, noise becomes a major problem. This application is no exception. Significant noise was present during the data acquisition process. In this situation, proper shielding must be verified, and hardware filtering may need to be implemented.

8.2.4 Solenoid Force–Displacement Calibration System

Overview The purpose of this project is to experimentally determine the force versus displacement characteristics of a solenoid used to actuate a hydraulic valve.

Parts List The hardware components required for this case study are presented in the Table 8-8. The computer software, Windows and VisSim/RT, are not included in this list and are assumed to be installed on the application PC.

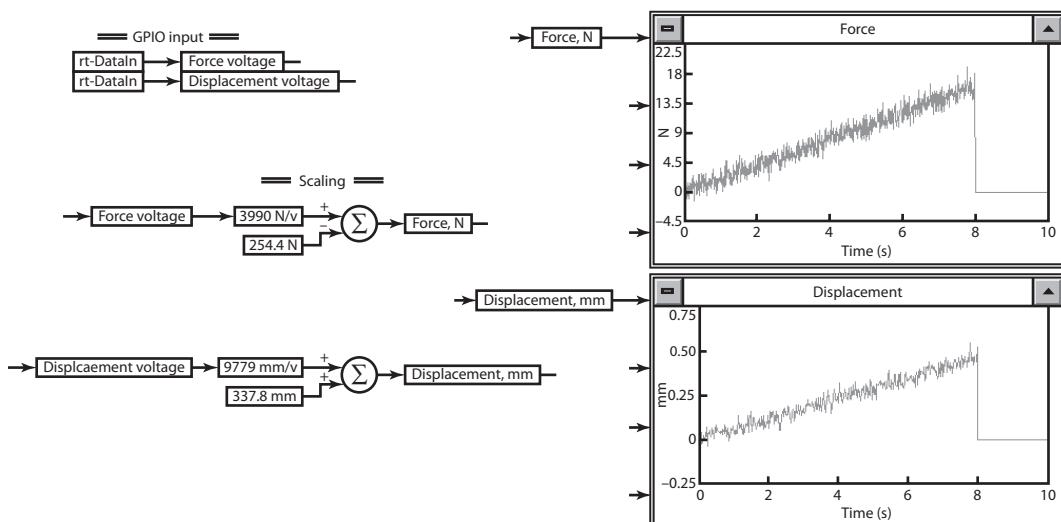
TABLE 8-8

1	Chatillon DRC100 digital force gauge
2	DC power supply
3	Misc. wiring and mounting
4	Advantech 711 GPIO card and screw terminal
5	Kaman KD2300 displacement measuring system

Experiment In this experiment, the solenoid is an actuation device for a three-port, two-way hydraulic valve. When the solenoid is activated, its two piece plunger assembly is pulled into the coil working against a spring attached to a ball seal. When the plunger pull-in force exceeds the spring force, the flow direction for the valve is reversed. To ensure proper operation, it is necessary to verify that the force exerted by the solenoid exceeds the spring force when flow reversal is desired. The position of the plunger is defined relative to the ball seal opening and is referred to as the “gap”. In this experiment, the pull-in force will be measured over a gap range of 0 to 0.5 mm.

Two analog inputs are supplied to the application program; force and displacement. Force is sensed using the force gauge attached to a pivoting assembly attached to the plunger. The displacement sensor measures displacement based on eddy currents induced in a conductive metal “sensor” created by impedance variations. The impedance variations are proportional to the distance separation between the solenoid coil and the conductive metal sensor. After the two input channels are read into the application program, they are converted to engineering units using experimentally determined linear relationships of the form $y = m \cdot x + b$. A diagram of the application program and test results for a sample 0.5 mm range are presented in Figure 8-31.

FIGURE 8-31 SOLENOID CALIBRATION SYSTEM APPLICATION PROGRAM



Initially, a two-point calibration was performed and the scaling constants for displacement and force were determined. The sample test shown was conducted by positioning the plunger inside the coil at a gap of 0 mm. With 20 V applied to the solenoid coil, the plunger position was manually moved (a rack and pinion assembly was used for this) to a gap of 0.5 mm. This application was executed using a *StepTime* of 0.01 s and a time range of 10 s.

8.2.5 Rotary Optical Encoder

Overview The purpose of this project is to validate recent results suggesting that stress shielding and implant micro motion are the main contributors to orthopedic hip implant failures. A rotary optical encoder is used as the sensor for measuring the micro rotations.

The incremental rotary optical encoder was selected for several reasons, including simplicity, accuracy, and reliability. Rotary optical encoders are widely used in machine tools, robotics, and other motion control systems. Encoders can provide information (such as relative displacement, direction, velocity, and position) depending on the method of decoding the output signal. Rotary encoders come in two types: absolute and incremental. Absolute encoders provide a coded binary word that corresponds to a unique position on the encoder disk. Incremental encoders provide a

simple pulse for each line encountered during rotation. Both types of encoders consist of a glass disk that is etched with patterns. A light source is projected towards the encoder disk, and a photoelectric sensor detects either the unique position for an absolute encoder or the alternating opaque and transparent lines for the incremental encoder.

Most incremental encoders have two output channels that are 90° out of phase. This configuration is called *quadrature*. Quadrature output provides four times the resolution of the encoder disk if a line count method were used. By sensing the two output signals and their relationship, four distinct conditions arise. Based on transitions between the conditions, the direction, count (displacement), and velocity values can be computed.

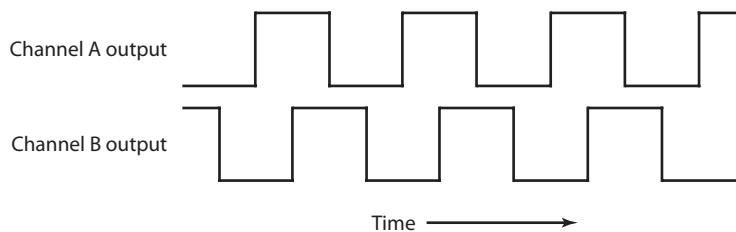
Parts List The hardware components required for this case study are presented in Table 8-9. The computer software, Windows and VisSim/RT, are not included in this list and are assumed to be installed on the application PC.

TABLE 8-9

1	1024 Line Hewlett Packard Incremental Rotary Optical Encoder
2	Hewlett Packard Photoelectric Detector
3	Misc. wiring and mounting
4	Advantech 711 GPIO card and screw terminal

Experiment The incremental rotary optical encoder has a disk radius of 11 mm. The two digital outputs (channel A and channel B) of the encoder were connected to two of the digital input channels of the 711 card. The general shape of the waveforms is shown in Figure 8-32.

FIGURE 8-32 ENCODER OUTPUT WAVEFORMS



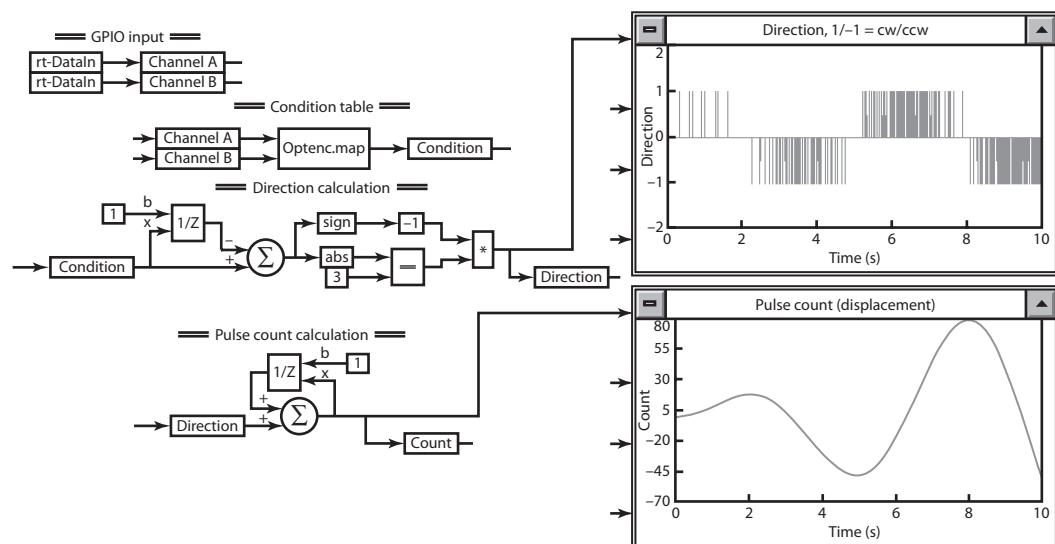
A high value on the waveform signal corresponds to a digital high voltage reading interpreted as a “1” in VisSim and a low value is interpreted as a 0 value. The four possible conditions are presented in Figure 8-33.

The VisSim application diagram performed several tasks. After reading the two digital input signals, the condition value was determined from an implementation of the condition table and then the count and direction were computed. The application diagram is execute with a *TimeStep*

FIGURE 8-33 CONDITION TABLE

Channel	Condition value			
	1	2	3	4
A	0	1	1	0
B	0	0	1	1

of 0.001 seconds for a total run time of 10 s. If a slower computer is used, the *TimeStep* must be increased to prevent overframing. The application diagram and results for a test 10 s run are presented in Figure 8-34.

FIGURE 8-34 OPTICAL ENCODER APPLICATION DIAGRAM

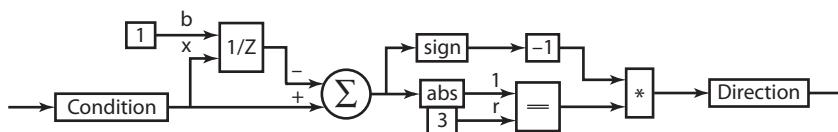
The condition table logic is implemented using a two dimensional table lookup routine called a *2D map* in VisSim. The map file is shown in Figure 8-35.

FIGURE 8-35 CONDITION TABLE MAP FILE

Notepad - OPTENC.MAP		
File	Edit	Search
0	1	
0	1	2
1	4	3

The x -axis of the map corresponds to the Channel A input, and the y -axis to the Channel B input. The direction calculation is based on the difference between consecutive condition values. If the condition value minus the last condition value (its past value) is -3 , then the direction is “1” or clockwise. If the difference is “ $+3$ ” then the direction is -1 or counterclockwise. The direction calculation is presented in Figure 8-36.

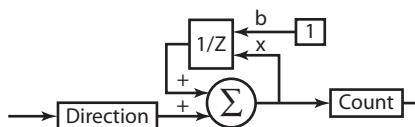
FIGURE 8-36 DIRECTION CALCULATION



The past value of the condition is computed using the unit delay block, $1/Z$, which is triggered to run every iteration as indicated by the 1 applied to its Boolean input. The absolute value of the delta condition value is then compared with 3. When it equals 3, the direction is -1 if the sign of the delta condition value is positive, and 1 if the sign is negative.

The pulse count, which corresponds to displacement, is computed by summing the direction sequence. A digital integrator is employed for this operation. The pulse count calculation is shown in Figure 8-37.

FIGURE 8-37 PULSE COUNT CALCULATION



The unit delay block, $1/Z$, holds the past value of the *Count* variable in this diagram, which is an accumulation of the pulses applied through the *Direction* variable.

8.3 Data Acquisition and Control Case Studies

This section presents several experiments which illustrate various uses of data acquisition and control (DAC) systems. In addition to the four components required for data acquisition systems, a DAC system uses actuators. As was the case in the data acquisition section of this chapter, cost is an important issue in actuator selection as well. Some of the actuators used in this section were purchased from equipment distributors such as Newark Electronics and OMEGA and others were extracted from consumer appliances.

8.3.1 Thermal Cycle Fatigue of a Ceramic Plate

Overview The objective of this experiment is to investigate the use of a DAC to evaluate fatigue failure in a ceramic plate due to thermal cycling. Current methods for evaluating fatigue failures are

imprecise. The plates are first heated by a 1600°C quartz lamp for a length of time based on historical information. Next, the plates are allowed to cool in ambient air, about 24°C for a length of time also based on historical information. The current method just outlined is an open-loop approach, which uses historical information to control the heating and cooling durations. This experiment applies a closed-loop approach to solve the same problem. Two actuators and one sensor are required for this approach. Plate heating is provided by actuating a heat lamp either ON or OFF and plate cooling by actuating a fan, which blows ambient air either ON or OFF. A single thermocouple is required to sense the plate temperature.

Parts List The hardware components required for this case study are presented in Table 8-10. The computer software, Windows and VisSim/RT, are not included in this list and are assumed to be installed on the application PC.

TABLE 8-10

1	250 watt heat lamp
2	2 speed fan
3	120 volt 15 amp double throw relay with 12 voltage trigger point
4	Aluminum test specimen
5	One type K thermocouple
6	Temperature display with millivolt output to computer (OMEGA)
7	DC power supply
8	Misc. wiring and mounting hardware
8	Advantech PCL 812 GPIO card and screw terminal

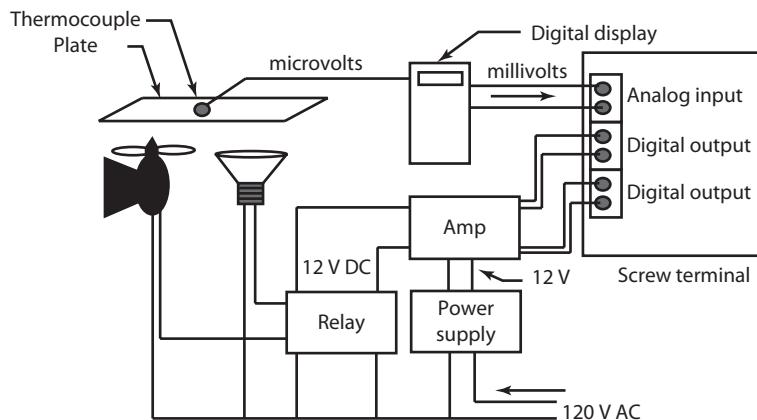
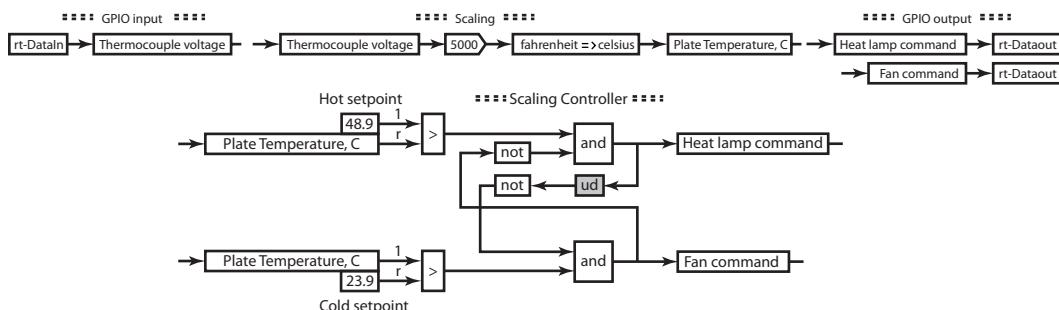
Experiment An application which cycles the temperature a plate in a controlled manner is developed. The control, which resides in the application program, is based on two temperature setpoints; ambient and hot. The hot setpoint used was 50°C. The control performs the following sequence of operations. Beginning with the plate temperature at ambient, the heat lamp is turned on and the plate temperature continuously monitored using the thermocouple until it reaches the hot setpoint value. The heat lamp is then turned OFF and the cooling fan turned ON. When the plate temperature reaches ambient, the fan is turned OFF, and the procedure is repeated.

A diagram of the system is shown in Figure 8-38. The screw terminal is attached to the GPIO card (in this experiment an Advantech 812 card was used due to its programmable gain range on analog inputs). The digital outputs are used to turn the fan and heat lamp ON and OFF. The VisSim application diagram for this system is presented in Figure 8-39.

The *rt-DataIn* block is configured to read an analog channel, and the two *rt-DataOut* blocks are configured as digital outputs. Within the application the thermocouple voltage is converted to engineering units of temperature (°C) for the range of operation using a conversion factor of 5000.

The controller is logic based and its operation is described in the following statements.

- The heat lamp is on when the plate temperature is less than the hot setpoint and the fan is off.
- The fan is on when the plate temperature is greater than the cold setpoint and the heat lamp is off.

FIGURE 8-38 THERMAL FATIGUE TEST RIG**FIGURE 8-39 APPLICATION DIAGRAM FOR THERMAL FATIGUE TEST RIG**

The controller is implemented to precisely satisfy these two statements with one caveat. To eliminate the implicit loop which results after implementing these statements a unit delay, referred to in the diagram as “ud”, must be included. If it were eliminated once could “trace” the implicit loop that exists between the two “and” blocks.

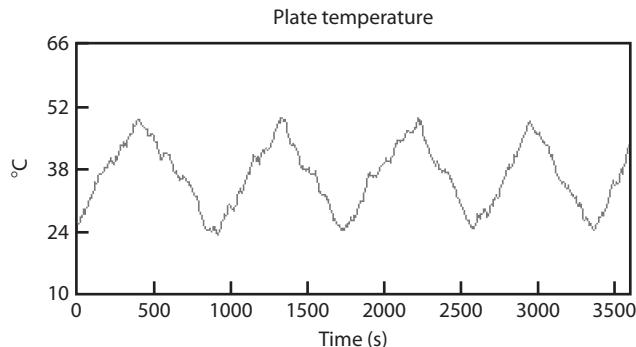
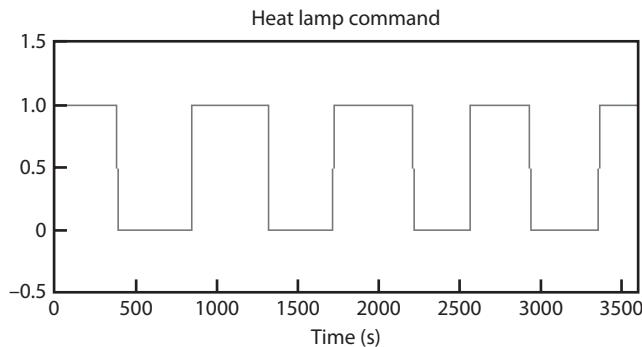
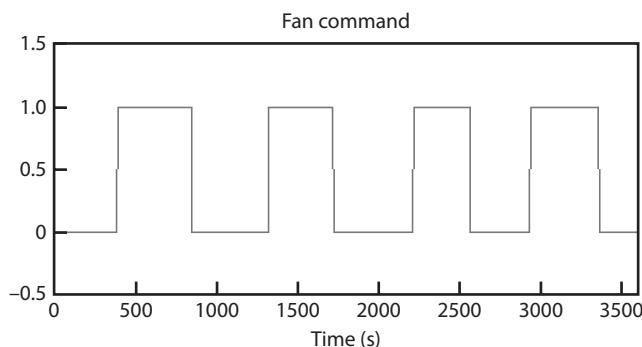
The plate temperature during a series of temperature cycles over a 60 min period are presented in Figure 8-40.

The fan and heat lamp command signals which caused the behavior are presented in Figures 8-41 and 8-42.

One important characteristic of the controller used in this experiment is its tolerance to measurement noise. Generally, more noise is present on sensor signals that operate at lower voltage ranges. The thermocouple operates at an extremely low voltage range, and frequently, its readings contain significant noise. Two methods are often employed to remove noise: external hardware filtering or desensitizing the processing algorithm. The latter approach has been used in this example.

8.3.2 PH Control System

Overview The objective of this experiment is to model an industrial pH-neutralization system. Such a system continuously monitors the pH level of a solution and makes adjustments as needed

FIGURE 8-40 PLATE TEMPERATURE**FIGURE 8-41 HEAT LAMP COMMAND****FIGURE 8-42 FAN COMMAND**

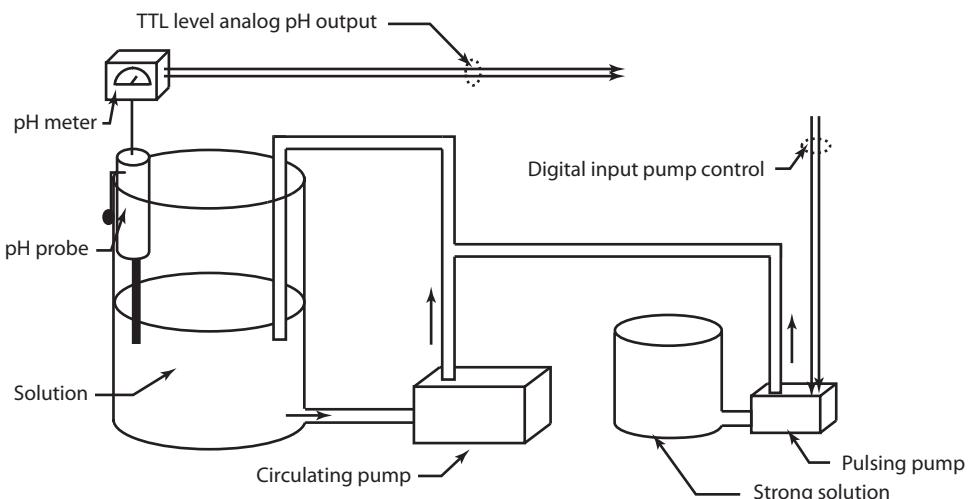
to maintain the pH level at a specified setpoint. The experiment requires one sensor for measuring pH and one actuator to control a metering device which adds a neutralizing agent to the solution.

Parts List The hardware components required for this case study are presented in Table 8-11. The computer software, Windows and VisSim/RT, are not included in this list and are assumed to be installed on the application PC.

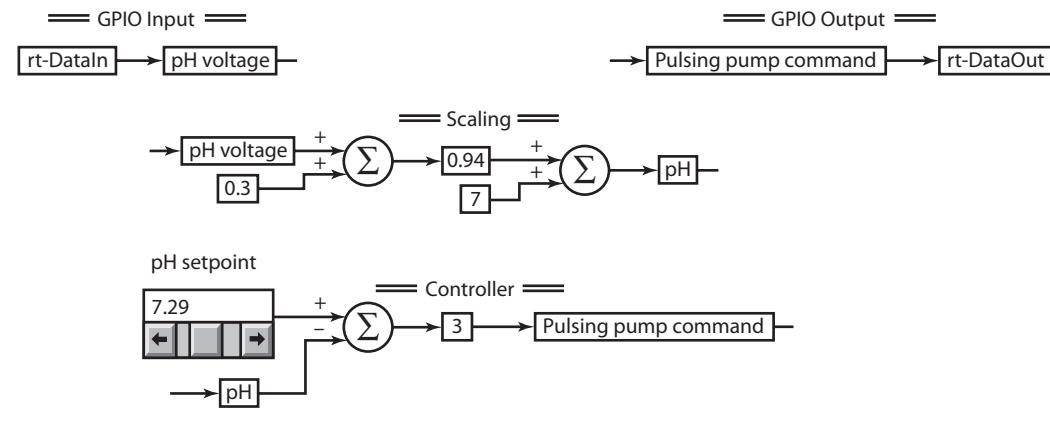
TABLE 8-11

1	DC Power Supply
2	Misc. plastic tubing and containers
3	Strong and weak pH solutions
4	pH probe (saturated potassium chloride KCL probe, OMEGA)
5	Pulsing pump capable of on/off control using TTL signals
6	Fixed speed circulating pump, uncontrolled
7	Misc. wiring and mounting hardware
8	Advantech PCL 812 GPIO card and screw terminal

Experiment A figure of the pH-control system is shown in Figure 8-43. A power supply for the pulsing pump is also present in the pH-system diagram but was excluded from the figure. The pH probe is partially submerged in the solution resevoir and communicates data to the pH meter, which displays the value on a digital readout and also creates a TTL level output signal corresponding to the displayed value. This analog voltage signal will be connected to the GPIO card to measure the pH. The circulating pump remains on for the duration of the experiment. Its function is to keep the solution resevoir mixed. The pulsing pump is an ON/OFF pump that can be remotely controlled with a digital output from a GPIO card.

FIGURE 8-43 PH SYSTEM

The VisSim diagram used for data acquisition and control of this system is shown in Figure 8-44. The 0.3 and 0.94 values used in the scaling calculation were needed to overcome the internal resistance of the pulsing pump. Since the pump behaves as an integrating device, the controller employed only a proportional term. The gain was experimentally determined as 3 after several tests. The *TimeStep* used was 1 s and the time range was 240 s.

FIGURE 8-44 PH SYSTEM DAC ALGORITHM

The pH system started with a solution pH of 7.3. At 20 s into the test 20 ml of pH 4.0 solution was poured into the solution reservoir to create a disturbance. This quickly dropped the pH of the reservoir solution to 6.5. The strong solution had a pH of 10.0 after a period of roughly two minutes the solution pH was returned to its original value.

8.3.3 De-Icing Temperature Control System

Overview This experiment investigates one of many new methods which may be suitable for de-icing the lifting surfaces of commercial aircraft. The approach utilizes the hot engine exhaust gas to melt ice. Since the exhaust is too hot to apply constantly to the aircraft surfaces, a system which controls flow is investigated in this experiment.

The wing was modeled using a small piece of sheet metal roughly 15 mm wide by 30 mm long. A thermocouple was mounted to the sheetmetal “wing” to monitor temperature. An electric hair dryer was used to model the hot engine exhaust airstream and aimed at the wing in the vicinity of the thermocouple. Power to the hairdryer was controlled by an ON/OFF switching circuit.

Parts List The hardware components required for this case study are presented in Table 8-12. The computer software, Windows and VisSim/RT, are not included in this list and are assumed to be installed on the application PC.

TABLE 8-12

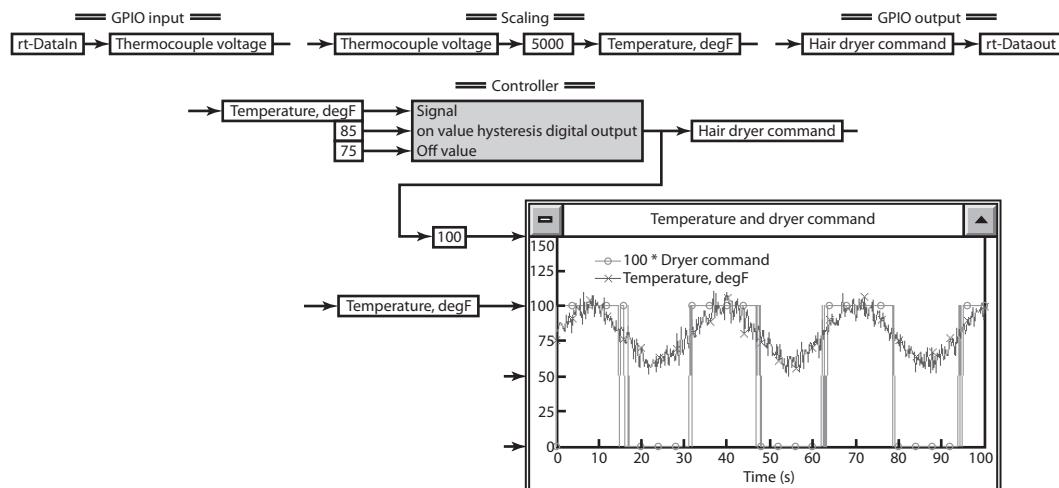
1	DC Power Supply
2	Thermocouple sensor (Ω)
3	12 volt Electric hairdryer
4	Misc. wiring and mounting and sheet metal
5	Advantech PCL 812 GPIO card and screw terminal

Experiment All components for this experiment were available except the transistor switch needed to turn the hairdryer on and off. This switching circuit was based on a 12-V DC single pole

double throw relay with 70 A contacts. The voltage from the GPIO card is sent to a transistor sensor that switches 12 V DC to the relay.

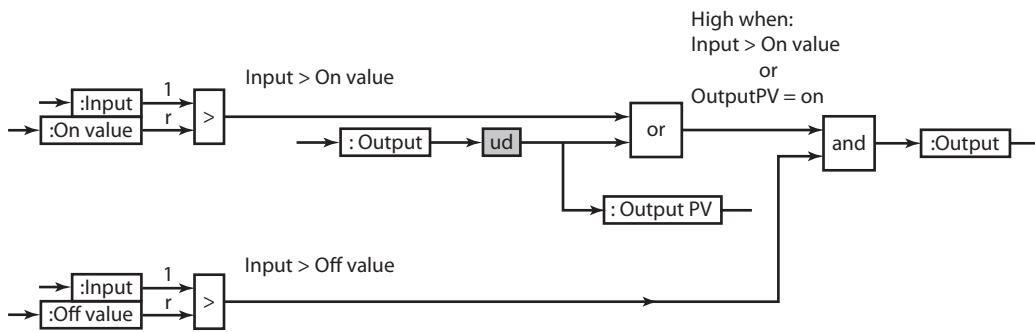
Since the hairdryer could be in either of two states, ON or OFF, a hysteresis-based controller was implemented. This type of controller (thermostat) is popular in many home heating systems. The application diagram is presented in Figure 8-45.

FIGURE 8-45 DE-ICING APPLICATION DIAGRAM



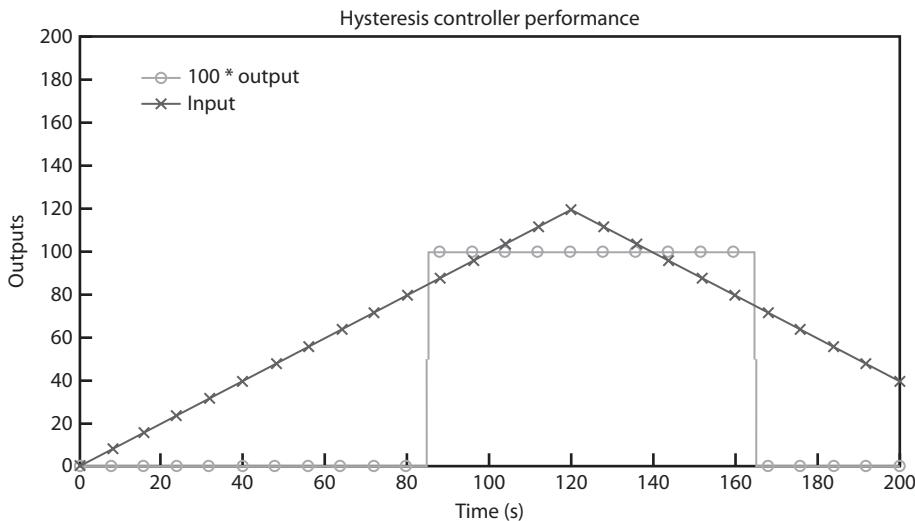
The hysteresis control block deserves some explanation. The contents of this “compound block” are shown in Figure 8-46. In this application, the *input* signal is the wing temperature, the *On Value* is set to 24°C and the *Off Value* is set to 29°C. The controller incorporates the following logic to turn the digital output (Output) on or off (1 or 0). *The Output is on if the (Input > On Value OR the Past Output is on) AND the Input > Off Value, otherwise the Output is off.* Lets walk through the implementation using *On Value* = 85, *Off Value* = 75, and apply a positive sloped input ramp signal to *Input*. Until the *Input* reaches the *On Value* (85), the *Output* stays off. When the *Input*

FIGURE 8-46 HYSTERESIS CONTROLLER



exceeds the *On Value* the *Output* turns ON, and remains ON. From this condition, reverse the slope of *Input* and make it negative, so it moves from its current value towards zero. The first condition to be executed is *Input > On Value* goes to false (0), however the *OutputPV* signal remains on (1), because the *Output* was initially on. When the *Input* falls below the *Off Value* the *Output* finally turns OFF. This behavior is illustrated in Figure 8-47.

FIGURE 8-47 BEHAVIOR OF THE HYSTERESIS CONTROLLER



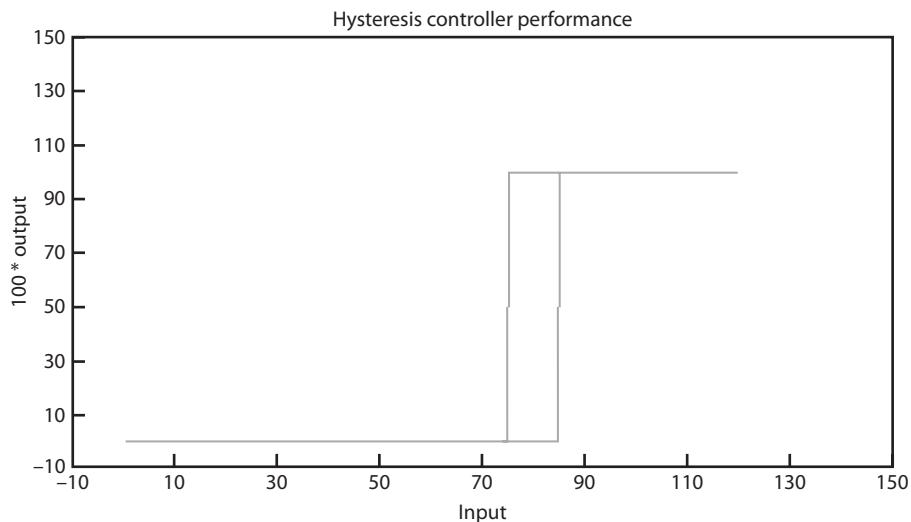
In this example, the *On Value* = 85, and the *Off Value* = 75. The *Output* turns on when the *Input* reaches 85 and turns off when the *Input* drops below 75. Another interesting way of viewing hysteresis is through an *x-y* plot of input versus output. This plot is shown in Figure 8-48.

8.3.4 Skip Control of a CD Player

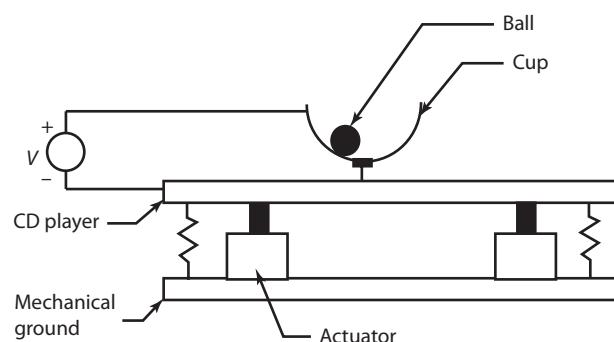
Overview Anyone who has ever used a portable CD player in their automobile has encountered the “skipping” effect that occurs due to jolting of the CD player while in use. One commonly employed remedy in use by many CD systems today is called nano-static ram. Although this is an effective method, it is expensive. This project explores an alternative approach which utilizes a “tilt” sensor (the type found in most pinball machines), an application algorithm, and electro-mechanical actuators.

Parts List The hardware components required for this case study are presented in Table 8-13. The computer software, Windows and VisSim/RT, are not included in this list and are assumed to be installed on the application PC.

Experiment A figure of the pinball tilt indicator and actuation system is presented in Figure 8-49. The tilt sensor operation is based on the metal ball completing a circuit. When the ball is not

FIGURE 8-48 HYSTERESIS INPUT - OUTPUT CHARACTERISTIC**TABLE 8-13**

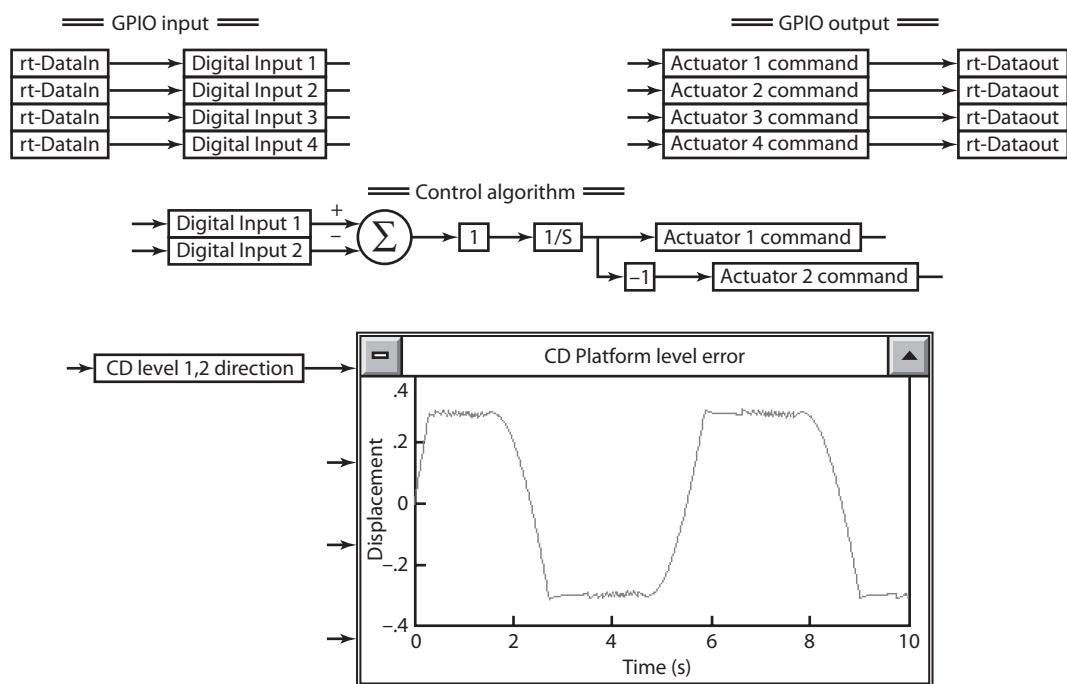
1	DC Power Supply
2	Pinball machine Tilt Sensor
3	4 Induction Magnet Actuators
4	Misc. wiring and mounting
5	Advantech PCL 812 GPIO card and screw terminal

FIGURE 8-49 PINBALL TILT SENSOR AND ACTUATORS, SIDE VIEW

centered in the bottom of the cup (level), the circuit is open. The figure shows a side view of the tilt indicator. The bottom contact in the cup is actually four separate contacts; two for fore-and-aft tilting and two for side-to-side tilting. The output of the tilt indicator is a binary (digital) signal where 1 means that all four contacts are active and 0 means that one or more contacts are inactive, which means that tilting is occurring. The binary output of the tilt sensor was modified to produce four separate binary signals corresponding to each of the four tilt directions (fore, aft, left, right). Four induction magnet actuators were mounted between the CD player and structural ground and each connected through the application program to the corresponding binary contact. Induction magnet actuators were selected due to their low power requirements, typically under 10 V; however, other low power actuators could also be used.

The application diagram for this experiment exercised control on one of the two axis. The application diagram is presented in Figure 8-50. The axes under control is designated as the 1, 2 axis. It could reflect the left, right motion or the fore, aft motion of the CD platform. The control algorithm is of type 1 (integral based). To prevent the actuators from limiting the commands to actuator 1 and 2 were slaved such that when actuator 1 was commanded to increase, actuator 2 was commanded to decrease. The integral gain 1 was arrived at after several tests. To test the performance of the algorithm, a 1 rad/sec sinusoidal disturbance was applied to the 1, 2 axis of the platform. The response of the platform is shown in the plot. Note the deadband of approximately 0.2 units due to the platform tilting with the ball still making contact on digital channels 1 and 2.

FIGURE 8-50 CD PLAYER LEVELING SYSTEM APPLICATION PROGRAM



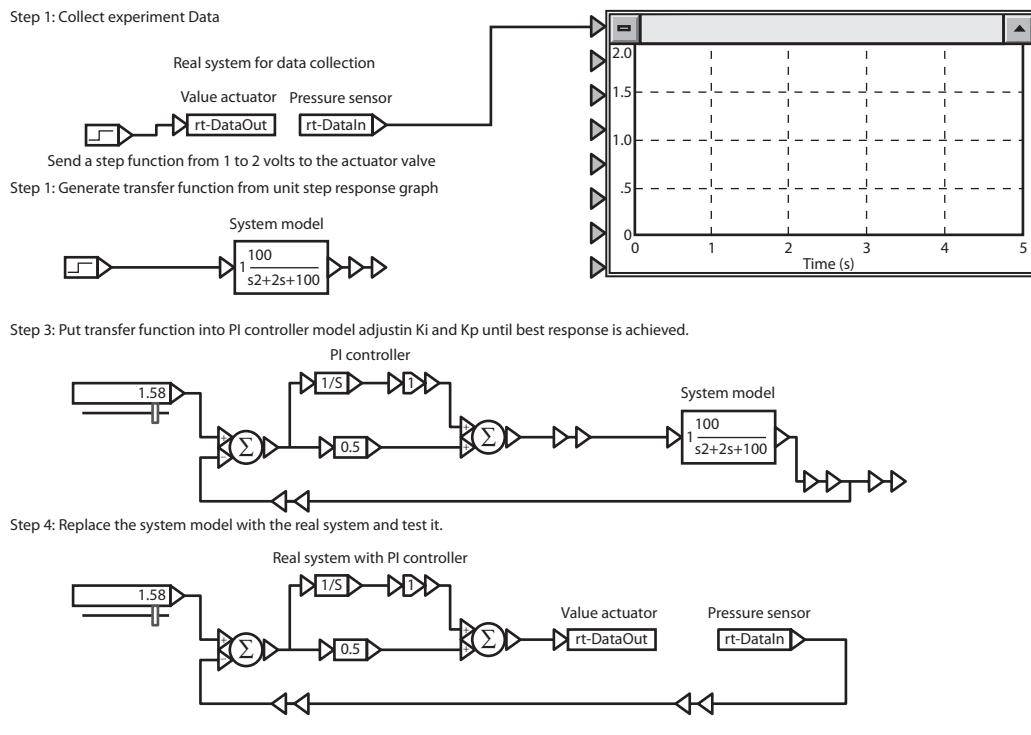
8.3.5 Simulation of Rocket Thrust Control in Laboratory

Overview The objective of this project was to simulate water thrust system in the laboratory, as used in rocket thrust control. Controlling the nozzle pressure to within a set tolerance of a given set point by modulating the water flow into the nozzle using a valve controlled by a rotary motor. A pressure sensor in the nozzle provides feedback for the proportional, integral (PI) control system.

Experiment The Thrust Control procedure employs a four step approach (Figure 8-51) to solving the problem of controlling the nozzle exhaust pressure.

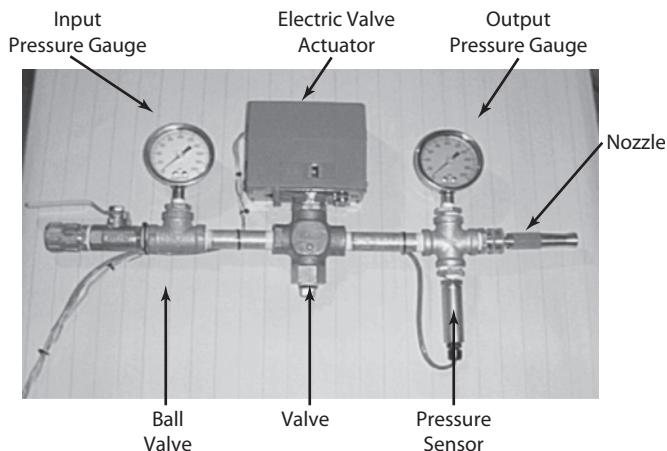
Figure 8-52 displays experimental set up and Figure 8-53 shows the simulation results.

FIGURE 8-51 SPACE SHUTTLE DISCOVERY STS-63 THRUST PROFILE TEST



8.3.6 Time Delay Blower

Overview The purpose of this case study is to monitor and control the temperature spread of air across a 3 m tube using mechatronics methodology of modeling and control using data acquisition

FIGURE 8-52 EXPERIMENTAL SETUP

Shetty, University of Hartford.

system. The system as shown in Figure 8-54 is designed to enable a user to put in a desired temperature and maintain the output temperature, taking into consideration changes on the ambient temperature.

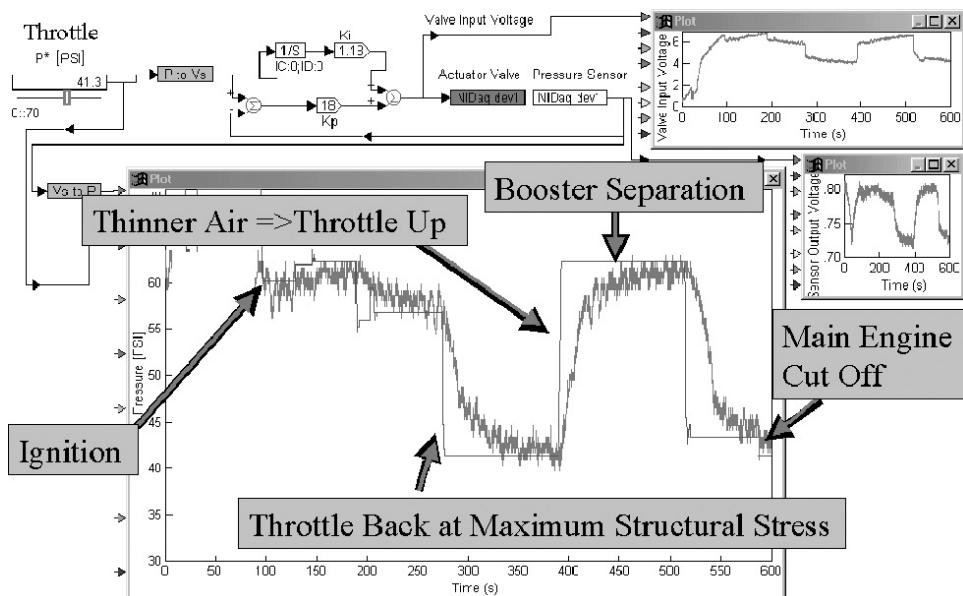
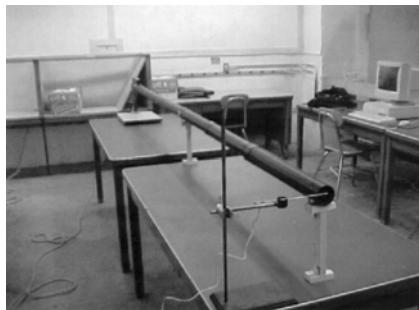
FIGURE 8-53 SIMULATION RESULTS

FIGURE 8-54 EXPERIMENTAL ARRANGEMENT OF TIME DELAY BLOWER

Shetty, University of Hartford.

Parts List

TABLE 8-14

1	Temperature probe
2	10feet section of pipe
3	Amplifier
4	Data acquisition card and screw terminal

Theory: Two heat-transfer relations are applicable to this example are the convection and the conduction. The blower produces heat in the form of hot air. The hot air travels through the tube (convection) until it reaches the end of the tube where the temperature probe reads the air temperature after its interaction with the ambient air. During the flow process, the air is also transferred from within the wall onto the outside air (conduction). These relationships between the material of the pipe and the heat transferred can be represented as follows.

- **Heat Transfer Models:**

$$q = h * A * (T_2 - T_1) \text{ where}$$

q = the amount of heat transfer in the process (kW)

h = the thermal conductivity of air ($\text{kW}/\text{m}^2*\text{K}$)

A = the cross sectional area where the heat passes through (m^2)

T_1 = the temperature at a position closer to the heat producer (K)

T_2 = the temperature at a position slight further than T_1 (K)

- **Conduction Process:**

$$q = k * A * (T_2 - T_1) / t \text{ where}$$

q = the amount of heat transfer in the process (kW)

k = the thermal conductivity of the material ($\text{kW}/\text{m}^2*\text{K}$)

A = the area where heat passes through (m^2)

T_1 = the temperature at a position closer to the heat producer (K)

T_2 = a temperature at a position slight further than T_1 (K)

t = the thickness of the material (m)

Experiment The heat blower is mounted on a stand, which consisted of a metal brace and a platform to support the brace. This was then placed at one end of the pipe. Two supports are then placed equidistant from the blower and the other end of the pipe to prevent the pipe from deforming due to weight. A temperature probe was then placed on a ring stand with a test tube clamp and placed approximately 1 in. from the opening of the pipe. The temperature probe is then connected to the input of the DA board. The VisSim model is constructed to systematically monitor the system. As the temperature read by the probe increase or decrease, depending on the value of the temperature selected to the temperature, the heat blower motor compensates the drop or the rise of the temperature.

This time delay is an important factor that has to be accounted for if the end temperature is to remain constant. With the help of the time delay transfer function, the output temperature can be controlled within 1° of the desired output temperature. This allows a proper calibration of the reading. A conversion from the voltage to the temperature is necessary for the user to analyze.

The model provides the operator the control of the system, and accurately displays what the state is doing and the conditions of the system during the real-time interface. After careful analysis of the system, a working VisSim and control system model were constructed and were able to produce a valid system, which will regulate the temperature of the 3 m pipe to within 1° of the desired temperature.

8.4 Summary

The case studies presented in this chapter address some of the common interfacing and control problems which arise when a mechatronics system is developed. These problems include sensor noise, overframing, and interfacing with high and low power electronics. The use of table look-up (or map) functions greatly simplifies implementation of complex nonlinear functions. The two-dimensional map function used in the rotary encoder case study to resolve the quadrature signals into condition numbers appears trivial.

Also introduced in this chapter is the notion of hysteresis as a mechanism for control. It is routinely found in many commercial devices including thermostats, deep well pump systems, and deicing systems. The essence of hysteresis is first introduced in the thermal cycling case study and more completely described in the de-icing temperature control case study.

REFERENCES

- Evans, Eva (2006). *Temperature Effect on Plants*. North Carolina State University. Online at [http://www.ces.ncsu.edu/depts/hort/consumer/weat her!tempeffect-plants.html](http://www.ces.ncsu.edu/depts/hort/consumer/weather!tempeffect-plants.html).
- Shetty, D., Campana, C., and Moslehpoor, S., “Standalone surface roughness analyzer,” *IEEE Journal of Instrumentation and Measurement*, March 2009, Vol. 58, No.3 pp 698–706.
- Shetty, D., Eppes, T., Campana, C., Filburn, T., and Nazaryan, N., “New approach to the inspection of cooling holes in aero-engines,” *Journal of Optics and Laser Engineering*, Volume 47, Issue 6, June 2009, Elsevier, 0143-8166, 2009.

Keshawarz, M. et. al and Shetty, D., “A Mechatronics Program as an alternative to separate programs in Electrical and Mechanical in Developing Countries,” AC 2009–1589., Proc. ASEE Conference, Texas, June 2009.

D. Shetty and L. Manzione, *Trends in Smart Manufacturing and Mechatronics*, Presented at the 2009 ASME International Manufacturing Science and Engineering Conference (*MSEC*), Conference Proceedings, Purdue University, October 2009.

PROBLEMS

1. Controlling Temperature of a Hot/Cold Reservoir

Design a system to control the water temperature of a mixing valve fed by two reservoirs. Output temperature is measured by a thermistor. The voltage of thermistor is utilized as an analog input to data acquisition cards. The simulation returns the temperature of water. The position of the valve is monitored by reading voltage across the potentiometer. The simulation and control program determines the position of the valve. The valve is controlled by 0 to 5 volt output. The thermistor is 0 to 2.5 volts.

2. Computer Monitored Automated Torque Wrench for Threaded Fasteners

Design an automated torque wrench which is used to tighten nuts to a specified value. The torque range is 3–12 N·m (+/– 2% error). Use a torque sensor with analog output, range 0–12 N·m is 0 to 5 volts. Actuation is with a relay control to a DC motor actuator.

3. Inverted Pendulum Control

Design a single inverted pendulum system driven by a DC motor and is controlled by simulation software. The input is the pendulum angle with respect to vertical read by a potentiometer on the pendulum axis. A transducer is used to control the car near the midpoint of the track.

4. Precision Position Sensing using Computer Interface

Design a positioning system based on an XY table with motion measured by encoders. The outputs of the encoders are fed to the interface. The table is actuated by using two stepper motors in the *x* and *y* directions. An encoder is coupled to each leadscrew and the position is fed back to the software to control the position of each stepper.

5. Active Vibration Control using Magnetostrictive Transducer

Magnetostrictive materials have a unique crystalline structure that produces high-strain amplitudes when placed in a magnetic field. Demonstrate the feasibility of using magnetostrictive actuators for the implementation of active vibration control. Investigate the ability of the actuators to achieve motion, frequencies, and amplitude, not possible through existing vibration control techniques.

APPENDIX 1

DATA ACQUISITION CARDS



The National Instruments Lab-PC+ is a low-cost multifunction I/O board for ISA computers. DAQPad-MIO-16XE-50 is available as an external unit that communicates with PC through the parallel ports. It supports analog, digital, and timing signals. It has 16 single-ended or eight differential analog inputs connected to multiplexor, a 16-bit A/D converter, a buffer, and a central DAQ-STC, which is a system-timing controller. It has eight bi-directional digital I/O lines connected to DAQ-STC. Actual connections to sensors and devices are made in a detachable screw terminal.

National Instruments Lab-PC-1200 100 kS/s, 12-Bit, 8 Anlg I/p Low Cost Multifunction I/O National Instruments DAQPad-MIO-16XE-50 20 kS/s, 16-Bit, 16 Anlg Inputs Multifunction I/O. The National Instruments Lab-PC-1200 and Lab-PC-1200AI are legacy low-cost multifunction I/O boards for ISA computers. They offer up to 100 kS/s, 12-bit performance on eight single-ended or four differential analog inputs; digital triggering; three 16-bit, 8 MHz counter/timers; two 12-bit analog outputs (Lab-PC-1200 only); and 24 digital I/O lines. Instead, you may want to consider using the PCI-6025E, a newer 12-bit PCI device with 16 analog inputs designed with our proven, Measurement Ready E Series architecture that samples at 200 kS/s on multiple channels.

The DAQPad-MIO-16XE-50 is a high-resolution, multifunction, portable DAQ system that communicates through the parallel port on IBM PC/XT/AT and compatible computers. The DAQPad-MIO-16XE-50 features 16 analog input (AI) channels that you can configure as 16 single-ended or eight differential inputs, a 16-bit successive-approximation ADC, two 12-bit DACs with voltage outputs, one constant current source for powering RTDs or thermistors, eight lines of TTL-compatible digital I/O (DIO), and two 24-bit counter/timers for timing I/O (TIO). The DAQPad-MIO-16XE-50 analog I/O circuitry is completely software-configurable and self-calibrated.

Appendix Table A1-1 presents some of the most popular graphical application software.

TABLE A1-1 POPULAR GRAPHICAL-BASED APPLICATION SOFTWARE

Name	Description
LabTech Notebook	General purpose DAC with analysis
Lab Windows	General purpose DAC with analysis
WorkBench PC	General purpose DAC
SnapMaster	General purpose DAC with analysis and display
Easiest	General purpose DAC with analysis
UnkelScope	High Speed DA
Snapshot	High Speed DA
Acquire	General purpose DAC

(Continued)

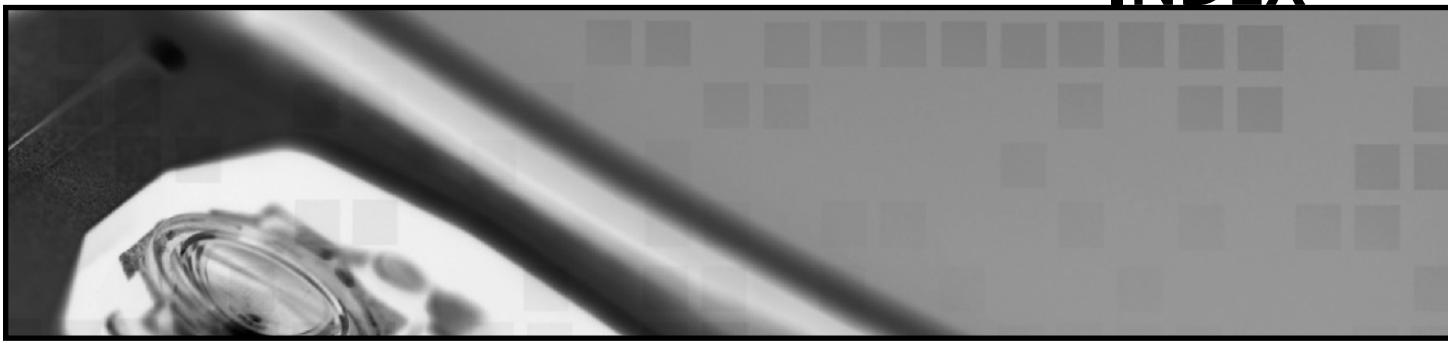
TABLE A1-1 (Continued)

LabView	General purpose DAC with analysis
Hyperception	High Speed DAC with analysis and display
Matrixx	High Speed DAC with analysis and display
Simulink	High Speed DAC with analysis and display
Visual Designer	General purpose DAC with analysis
Xanalog	High Speed DAC with analysis and display
VisSim	General purpose DAC with analysis and display

TABLE A1-2 POPULAR GPIO CARDS

Vendor	Board Name	Vendor	Board Name
Advantech <i>Analog and Digital I/O</i>	PCL-711	ComputerBoards <i>Analog Output (cont)</i>	CIO-DDA06
	PCL-711S		
	PCL-718		Digital I/O
	PCL-812		CIO-DIO24
	PCL-812PG		CIO-DIO24H
	PCL-818		CIO-DIO48 ¹
Data Translations <i>Analog and Digital I/O</i>	PCL-818PG	ComputerBoards <i>Analog Input</i>	CIO-DIO48H
	DTI-2811PGH		CIO-DIO96
	DTI-2811PGL		CIO-DIO192
Strawberry Tree <i>Analog and Digital I/O</i>	ACAO		CIO-PDIS08
	ACPC		
	ACJr		
Technology-80	M5312	ComputerBoards <i>PCMCIAs (laptop)</i>	CIO-DAS08
MetraByte <i>Analog and Digital I/O</i>	DAC-02		CIO-DAS08/AO
	DAC-16		CIO-DAS08/AOH
	DAS-08		CIO-DAS08/AOL
	DAS-08/AO		CIO-DAS08-PGH
	DAS-08/LT		CIO-DAS08-PGL
	DAS-08/PGA		CIO-DAS16
	DAS-16		CIO-DAS16/330
	DAS-16F		CIO-DAS16/330i
	DAS-1400		CIO-DAS16/F
	DAS-1600		CIO-DAS16/Jr
	DDA-06		CIO-DAS16/M1
	PDIS08		CIO-DAS48
ComputerBoards <i>Analog Output</i>	PIO-12		CIO-DAS48PGA
	PIO-24		CIO-DAS1601/12
	CIO-DAC02		CIO-DAS1602/12
	CIO-DAC08		CIO-DAS1602/16
	CIO-DAC08-I		

INDEX



A

- Absolute encoder, 162–163
- Acceleration measurement, 190–191
- Accuracy, 138, 349–353
 - sensor measurement of, 138
 - signal systems, 349–353
- Active sensors, 135
- Active vibration control, 192–193
- Actuators, 17, 30–31, 255–290
 - direct current (DC) motors, 255–261
 - electromagnetic, 17
 - fluid power, 17, 269–286
 - permanent magnet (PM), 257–261, 262–269
 - piezoelectric, 287–289
 - smart sensors and, 30–31
 - stepper motors, 262–269
 - unconventional types of, 17
- Alternating current (AC)
 - analysis, 14–15
- Alternating current (AC) motors, 262, 493–502
 - induction, 500–502
 - magnetic field, 493–494
 - operation of, 493–498
 - rotating field, 494–498
 - rotor movement, 498
 - synchronous, 498–500
- Amplification of signals, 393

- Amplifiers, 394–401
 - instrumentation, 400–401
 - operational (op-amps), 394–399
- Analog inputs, critical consideration of, 407
- Analog sensors, 134, 222–223
- Analog-to-digital converter (ADC), 403–409
 - data conversion using, 402–407
 - digital-to-analog converter (DAC) and, 408–409
 - flash converter, 408
 - successive approximation, 407–409
- Angle of rotation, stepper motors, 268
- Angular position measurement, 225–228
- Auto-control system, example of, 462–466
- Axial piston pumps, 278

B

- Backlash, sensor measurement of, 139
- Basic feedback systems (BFS), 46–47, 341–343, 357–359
 - block diagram, 46–47
 - G-equivalent form, 341–343
 - root locus and, 357–359
 - signal systems, 341–343, 357–359
- Bernoulli's equation, 198
- Binary logic, 297–302
- Binocular vision technique, 237–238
- Bit instructions, PLC, 314

- Block diagrams, 4, 7, 10–11, 43–75, 77–82, 83–90, 92–94, 109–116, 264–267
 analogy approach, 64–75
 basic feedback systems (BFS), 46–47
 block functions, 10–11, 43–44
 direct method, 51–64
 electrical systems, 77–82
 flow and initial conditions, 43–44
 flow variables (FV), 65–66
 fluid systems, 109–116
 impedance diagrams, 66–72
 manipulations, 44–50
 mechanical illustrations, conversion from, 59–64
 mechanical systems, 83–90, 92–94
 mechatronic design process, 4, 7
 modified analogy approach, 72–75
 potential variables (PV), 65–66
 signal wires, 10–11, 43–44
 simulation, 50
 stepper motors, 264–267
 summing junctions, 45–46
 transfer functions (ODE), conversion from, 51–59
 Bode plots, 370–378
 controls, 371–378
 lag compensator design, 372–373
 lead compensator design, 375–377
 proportional derivative (PD) compensator design, 377–378
 proportional integral (PI) compensator design, 373–375
 sketching procedure, 370
 Bonded strain gauges, 172–174
 Boolean algebra laws, 297
 Branches, 15
 Brushless direct current (DC) motors, 261
- C**
- Camera motion method, 236–237
 Capacitance transducers, 146, 154–162
 angular rotation and, 157–158
 area change in cylindrical shapes, 157
 area change of plates, 156
- dielectric constant variation and, 158–159
 differential arrangement and, 159–160
 distance change between plates, 155–156
 variation of inductance principle for, 146
- Capacitors, 76
 Channel inputs, 394
 Characteristic equation, 339
 Check valves, 280
 Circuits, 14–15, 285–286, 397–400
 alternating (AC) analysis, 14–15
 data conversion and, 397–400
 direct (DC) analysis, 14–15
 fluid power, 285–286
 op-amp, 397–399
 open, 15
 zero/span, 399–400
- Closed-loop transfer function (CLTF), 341, 358
- Code generator, 4, 8
 Code width, 407
 Coefficient of coupling, 150
 Communications instructions, PLC, 315
 Compensator designs, *see* Bode plots;
 Root locus
 Component variation, sensitivity analysis for, 139–144
 Compound relief (pilot operated) valves, 279
 Concurrent engineering, 5
 Condition monitoring, 18–21
 Constraints, 12
 Control instructions, PLC, 315
 Control structure, mechatronic hierarchy of, 24–25
 Control systems, 329–338, 388–392
 data acquisition and (DAC), 388–392
 functions, 331
 general purpose I/O (GPIO) card, 390–391
 input/output (I/O) process, 389–390
 installation of GPIO card and software for, 391–392
 Laplace transform ODE solution, 332–338
 signals, 331–332
 Controller design, 378–383

Controls, 329. *See also* Control systems
 Conversion module, 136
 Counter instructions, PLC, 314–315
 Counterbalance (back pressure) valves, 280
 Current source, 76–77

D

Damping ratio, 353–354
 Data acquisition (DAQ), 388–392, 409–444,
 466–489, 491–492
 cards, 390–391, 491–492
 case studies, 466–489
 control system and (DAC), 388–392,
 476–489
 general purpose I/O (GPIO) card, 390–391
 input/output (I/O) process, 389–390
 installation of GPIO card and software
 for, 391–392
 LabVIEW, 409–423
 software applications, 409–444
 VisSim, 423–444
 Data conversion, 394–409
 analog-to-digital converter (ADC),
 403–409
 instrumentation amplifiers, 400–401
 op-amp circuits, 397–399
 operational amplifiers (op-amps), 394–399
 process, 402–409
 zero/span circuits, 399–400
 Data display module, 136
 Data transmission module, 136
 Deflection sensors, 135
 Degrees of freedom, 344
 De-icing temperature control system,
 example of, 481–483
 DeMorgan’s theorem, 299–300
 Differential pressure, flow sensor
 measurement and, 198–201
 Digital hydraulic linear positioner, 273
 Digital I/O interfaces, 394
 Digital sensors, 134, 162–168, 224
 Digital-to-analog converter (DAC), 408–409
 Digital transducers, 162–168
 Direct current (DC) analysis, 14–15

Direct current (DC) motors, 255–261
 actuator applications, 255–256
 brushless, 261
 mathematical model of, 256–261
 permanent magnet (PM), 257–261
 Distance moved, stepper motors, 268
 Distance sensing, ultrasonic, 232
 Doppler shift technique, 231
 Drag-force flow meter, 203–204
 Drive equations, stepper motors, 265

E

Eddy current transducers, 216–219
 Electric timers, 286
 Electrical systems, 14–16, 75–82, 95–102
 alternating current (AC) analysis, 15
 block diagrams for, 77–82
 current source, 76–77
 direct current (DC) analysis, 15
 impedance characteristics, 76–77
 Kirchhoff’s laws for, 15
 mechanical system coupling, 95–102
 modeling and simulation of, 75–82
 power, 16
 voltage source, 76–77
 Electrical to mechanical coupling, 95–97
 Electromagnetic actuators, 17
 Electromagnetic flow meter, 208–210
 e-manufacturing, 34–35
 Encapsulation, 7
 Encoders, 162–168, 334
 Energy-input devices, *see* Pumps
 Energy-modulation devices, *see* Valves
 Energy-output devices (motors), 283–285
 Equal effects method, 141
 Equivalent circuits, 187–188
 Error analysis, 140
 Evidence-based diagnostics, 22
 Examine if Closed (XIC) instruction,
 PLC, 314
 Examine if Open (XIO) instruction,
 PLC, 314
 Excitation of signals, 393
 External gear pumps, 276

F

- Faraday's law, 97–100, 148
 Feedback, 271–272
 Fiber-optic range sensor devices, 241–264
 Fiber-optic temperature sensing, 214
 Filtering of signals, 393
 Fixed-volume, pressure-compensated, flow-control valves, 283
 Flash converter, 408
 Flow-divider valves, 283
 Flow measurement, 195–210, 228
 differential pressure and, 198–201
 drag-force flow meter, 203–204
 electromagnetic flow meter, 208–210
 Hall effect and, 228
 hot wire anemometers, 207
 laser Doppler effect for, 206
 liquid flow, 196–197, 228
 rotor torque mass flow meter, 205
 sensors for, 195–210
 solid flow, 196
 turbine flow meter, 204–205
 ultrasonic flow transducers (flow meters), 201–203
 Flow variables (FV), 65–66
 Fluid cylinders, 283
 Fluid-displacement motors, 284
 Fluid motors, 283–284
 Fluid power actuators, 17, 269–286
 applications of, 269–270
 circuits, 285–286
 control modes of, 285–286
 control systems in, 270–273
 digital hydraulic linear positioner, 273
 energy-input devices (pumps), 274–278
 energy-modulation devices (valves), 278–283
 energy-output devices (motors), 283–285
 feedback, 271–272
 servo valve for, 272
 spool valves, 270–271
 switches for, 286
 Fluid systems, 102–116
 block diagrams for, 109–116
 impedance characteristics of, 105, 108

- modeling and simulation of, 12–116
 properties of, 102–15
 restriction in, 105–108
 Force sensing resistor (FSR), 181–182

G

- G-equivalent form of basic feedback system (BFS), 341–343
 Gauge factor, 170–171
 Gear motors, 284
 General purpose I/O (GPIO) card, 390–391

H

- Hall effect, 218–246
 analog output sensors, 222–223
 angular position measurement, 225–228
 digital output sensors, 224
 flow measurement, 228
 liquid level measurement, 228
 open-collector output encoder, 224–225
 optical sensors, 233–246
 pneumatic transducers, 228–231
 position (magnetic) sensing, 224–225
 principle, 218–220
 pull-up resistor, 224–225
 range sensing techniques, 233–346
 rotational transducers, 221
 sensor construction, 222
 transducer applications, 224–246
 Hardware-in-the-loop simulation, 8–10, 17
 Hot wire anemometers, 207

I

- Impedance diagrams, 66–72, 76–77, 82, 91, 105–108
 block diagram modeling, 66–72
 complex systems, 71–72
 electrical systems, 76–77
 fluid systems, 105, 108
 mechanical systems, 82, 9
 parallel combinations, 69–70

phasors, 66–67
 series combinations, 69–70
 Incremental encoder, 162–163
 Inductance transducers, 146, 148–151
 Inductors, 76
 Inertias for rotational systems, 95
 Information systems, 10–13
 Input and output devices, PLC, 315
 Input branch, 316
 Input/output (I/O) process, DA, 389–390
 Instruction set overview, PLC, 314
 Instrumentation amplifiers, 400–401
 Interferometrics, temperature sensing using, 214–216
 Internal gear pumps, 276
 Isolation of signals, 393

K

Karnaugh map (K-map) methods, 302–309
 four-variable, 306–309
 three-variable, 304–306
 two-variable, 303–304
 Kirchhoff’s laws, 15

L

LabVIEW, 409–423
 applications, 415–423
 virtual instruments (VI) environment, 409–415
 Ladder logic diagrams, PLC, 315–318
 Ladder programming, 312–313
 Lag compensator design, 362–364, 372–373
 Laplace transform ODE solution, 332–338
 Laser Doppler effect for flow measurement, 206
 Laser interferometric transducers, 240–242
 Latching relays, 318
 Lead compensator design, 365–367, 375–377
 Life cycle design, 6–7
 Light stripe method, 236
 Limit switches, 286
 Linear (reflection type) encoder, 163–164, 166
 Linear relationships of electrical–mechanical coupling, 99–100

Linear variable differential transformer (LVDT), 151–152
 Linearity, sensor measurement of, 139
 Linearization of nonlinear systems, 343–346
 Linearization of signals, 394
 Liquid flow measurement, 196–197, 228
 Logic diagrams, 300–302
 Loop transfer function (LTF), 341, 358
 Lorentz’s law, 95–97, 99–100

M

Magnetoresistive transducers, 193–195
 Magnitude condition, root locus, 358–359
 Mass–spring–oscillation and damping example, 447–452
 Measurement parameters (modules) in sensors, 135–137, 139
 Mechanical illustrations, block diagram conversion from, 59–64
 Mechanical systems, 13–14, 82–90, 95–102
 block diagrams for, 83–90, 92–94
 electrical system coupling, 95–102
 impedance characteristics, 82, 91
 inertias for, 95
 modeling and simulation of, 82–90
 Newton’s laws for, 13–14, 82, 90–91
 rotational, 90–95
 translational, 82–90
 Mechanical to electrical coupling, 97–99
 Mechatronics, 1–40, 291–328, 446–490
 actuators, 17, 30–31
 applications of, 36–38
 block diagrams, 4, 7, 10–11
 case studies, 446–490
 comprehensive case studies, 446–466
 concurrent engineering and, 5
 condition monitoring, 18–21
 control structure, 24–25
 data acquisition (AC) case studies, 466–476
 data acquisition and control (DAC) case studies, 476–489
 design process, 6–10
 electrical systems, 14–16
 e-manufacturing, 34–35
 embedded smart sensors, 30–31

- Mechatronics, (*Continued*)
 hardware-in-the-loop simulation, 8–10, 17
 information systems, 10–13
 integrated design in, 4–5
 interactive modeling, 27–28
 mechanical systems, 13–14
 model-based manufacturing, 23–27
 multidiscipline methodology of, 1–3
 on-line quality monitoring, 22
 open architecture systems, 23–27
 optomechatronics, 33–34
 physical system, 10
 programmable logic controllers (PLC), 309–323
 rapid prototyping, 31–33
 real-time interfacing, 8–10, 17
 sensors, 16–17, 30–31
 system control, 291–328
 trade off evaluation, 28–30
 virtual machine prototyping, 28
- Mechatronics Technology Demonstrator (MTD), 447–452
- Message Read/Write (MSG) instruction, PLC, 315
- Modeling, 8, 10–11, 23–27, 27–28, 41–130
 block diagrams, 10–11, 43–75
 electrical systems, 75–82
 electrical-mechanical coupling, 95–102
 fluid systems, 102–116
 interactive, 27–28
 mechanical illustrations and, 59–64
 mechanical rotational systems, 90–95
 mechanical translational systems, 82–90
 operator notation, 42–43
 physical systems, 41–130
 simulation and, 8, 10–11, 50
 transfer functions, 42–43, 51–59
- Modules (measurement parameters) in sensors, 135–137, 139
- Moiré fringe transducers, 164, 167–168
- Motion and position measurement, sensors used for, 144–162
- Motor equations, stepper motors, 265–267
- Motors, 257–262, 265–267, 283–285, 452–462, 493–502
 alternating current (AC), 262, 493–502
- direct current (DC), 257–261, 452–462
 fixed-displacement fluid, 284
 fluid, 283–285
 gear, 284
 permanent magnet (PM) actuators, 257–262, 265–267, 452–462
 piston, 284–285
 stepper, 265–267
 vane, 284
 variable-displacement fluid, 284
- Multiplexing of signals, 393
- Mutual inductance, changes in, 150–151
- N**
- Needle valves, 282
 Newton's laws, 13–14, 82
 Nodes, 15
 Noise, 407
 Non-latching relays, 318
 Null sensors, 135
 Number systems in mechatronics, 291–297
- O**
- Offset voltage, resistance transducers and, 175–180
 On-line quality monitoring, 22
 Op-amp circuits, 397–399
 Open architecture mechatronic systems, 23–27
 Open-collector output encoder, 224–225
 Operational amplifiers (op-amps), 394–399
 Operator notation, 42–43
 Optical sensors, 233–246. *See also Range sensors*
 Optimization procedure, 11–13
 Optomechatronics, 33–34
 Ordinary differential equations (ODE), 51–59, 332–338
 block diagram model conversion from, 51–59
 Laplace transform solution of, 332–338
 Output Energize (OTE) instruction, PLC, 314
 Output equation (OE), 52–53

P

Pade approximation, 346–348
 Parallelogram law, 14
 Parameters, 10, 18–19, 135–139, 329
 condition monitoring, 18–19
 constant variables as, 329
 errors and uncertainties in, 139
 measurement (modules), 135–137, 139
 modeling fixed-values, 10
 quality, 137–139
 sensors, 135–137, 139
 Passive sensors, 135
 Percent overshoot (P.O.), 330
 Permanent magnet (PM) actuators, 257–261,
 262–269, 452–462
 direct current (DC) motors, 257–261,
 452–462
 position control, example of, 452–462
 stepper motors, 262–269
 Permittivity constant, 154–155
 PH control system, example of, 478–481
 Phase condition, root locus, 358–359
 Piezoelectric actuators, 287–289
 Piezoelectric film (tactile sensor), 181
 Piezoelectric transducers, 146, 183–192
 acceleration measurement by, 190–191
 analogy equations for, 188–190
 charge generation principle, 146
 equivalent circuit of, 187–188
 output, 185–187
 piezoelectric effect of, 183–185
 velocity measurement by, 191–192
 Piston motors, 284–285
 Piston-type pump mechanisms, 277–278
 Pitot tube, 199–200
 Planetary pumps, 276
 Pneumatic transducers, 228–231
 Pole placement method, 378–383
 Pole-zero (PZ) plot, 354–355
 Position control of PM DC motor,
 example of, 452–462
 Position (magnetic) sensing, 224–225
 Position sensor detectors (PSD), 239–241
 Position valves, 281
 Positioning system, stepper motors, 267–269

Positive-displacement pumps, 274–275
 Potential variables (PV), 65–66
 Potentiometers, 146–148
 Power, electrical systems, 16
 Pressure-control valves, 278–282
 Pressure switch valves, 280
 Pressure switches, 286
 Primary transducers, 145
 Principle of transmissibility, 14
 Probing signals, 346
 Programmable logic controllers (PLC),
 309–323
 applications of, 309–310
 architecture, 310–312
 bit instructions, 314
 communications instructions, 315
 control instructions, 315
 features of, 313–323
 input and output devices, 315
 instruction set overview, 314
 ladder logic diagrams, 315–318
 ladder programming, 312–313
 relays, 318–319
 sequence instructions, 315
 timer and counter instructions, 314–315
 Proportional derivative (PD) compensator
 design, 377–378
 Proportional integral (PI) compensator design,
 364–365, 373–375
 Pull-up resistor, 224–225
 Pulse technique, 231
 Pulses, stepper motors, 268
 Pumps, 274–278
 axial piston, 278
 design feature classification of, 275–276
 external gear, 276
 internal gear, 276
 non-positive-displacement, 275
 piston-type mechanisms, 277–278
 planetary, 276
 positive-displacement, 274–275
 rotary, 276, 277–278
 screw, 276–277
 vane-type mechanisms, 277
 Push-button switches, 286
 Pyrometers, 216

Q

Quality parameters in sensors, 135–139
 Quantization process, 403

R

Radiative temperature sensing, 213–216
 Range of signals, 406
 Range sensors, 233–346
 binocular vision technique, 237–238
 camera motion method, 236–237
 fiber-optic devices for, 241–264
 laser interferometric transducers, 240–242
 light stripe method, 236
 optical, 233–246
 position sensor detectors (PSD), 239–241
 spot projection method, 235–236
 time-of-flight method, 237
 triangulation principle, 234–235
 Rapid prototyping, 31–33
 Rate feedback compensator design, 367–370
 Real-time interfacing, 8–10, 17, 387–445
 application software for, 409–444
 data acquisition, 388–392
 data conversion, 394–409
 hardware-in-the-loop simulation, 8–10, 17
 signal conditioning and, 387–445
 Reduced instruction set computers (RISC), 314
 Regulator valves, 280
 Relays, PLC, 318–319
 Relief valves, 279
 Repeatability, sensor measurement of, 139
 Resistance temperature detector (RTD)
 wire, 210–211
 Resistance transducers, 146–148, 168–183, 224
 Hall effect and, 224
 offset voltage and, 175–180
 potentiometers, 146–148
 pull-up, 224
 resistive sensitivity, 168–171
 strain gauge arrangements, 171–183
 tactile sensors, 180–183
 Wheatstone bridge circuit arrangement,
 174–175, 176, 182–183

Resistive transducers, *see* Resistance transducers

Resistors, 76
 Resolution, 138, 404–406
 sensor measurement of, 138
 signal, 404–406
 Rise time, 330
 Rocket thrust simulation, example of, 486
 Root locus, 357–370
 basic feedback system (BFS) and,
 357–359
 controls, 362–370
 equation, 359
 lag compensator design, 362–364
 lead compensator design, 365–367
 magnitude condition, 358–359
 phase condition, 358–359
 proportional integral (PI) compensator
 design, 364–365
 rate feedback compensator design,
 367–370
 sketching procedure, 359–362

Rotameter, 200–201
 Rotary encoders, 163–164, 167
 Rotary optical encoder, example of, 473–476
 Rotary piston pumps, 277–278
 Rotary pumps, 276
 Rotary variable differential transformer
 (RVDT), 152–154
 Rotational speed, stepper motors, 268–269
 Rotor torque mass flow meter, 205

S

Sampling rate, 404
 Scanning process, 311
 Screw pumps, 276–277
 Secondary transducers, 145
 Selector switches, 286
 Semiconductor strain gages, 180
 Sensing module, 136
 Sensitivity, 137–144, 356–357
 component variation, 139–144
 equal effects method, 141
 error analysis, 140

- sensor measurement of, 137–138
- signal systems, 356–357
- square root of sum of squares (RSS)
 - method, 141
- Sensors, 16–19, 30–31, 131–254. *See also*
 - Transducers
 - active, 135
 - actuators and, 17, 30–31
 - analog, 134, 222–223
 - applications of, 216–246
 - component variation, 139–144
 - condition monitoring use of, 18–19
 - deflection, 135
 - digital, 134, 162–168, 224
 - embedded smart systems, 30–31
 - encoders, 162–168, 224
 - flow measurement, 195–210
 - measurement parameters (modules) in, 135–137, 139
 - mechatronic requirements of, 16–17
 - motion and position measurement using, 144–162
 - null, 135
 - optical, 233–246
 - passive, 135
 - quality parameters in, 137–139
 - range sensing techniques, 233–246
 - resistance (force and torque) and, 168–171, 224
 - sensitivity analysis, 139–144
 - strain gauges, 171–183
 - tactile, 180–183
 - temperature measurement, 210–216
 - thermistors, 211–212
 - transducers, 132–139, 145–246
 - vibration–acceleration, 183–195
 - Sequence instructions, PLC, 315
 - Sequence valves, 280
 - Sequential engineering, 6
 - Service communications (SC) instruction, PLC, 315
 - Servo valve, 272
 - Settling time, 330, 407
 - Signal wires, 10–11, 43–44
 - Signals, 10, 43–49, 132, 176, 329–386, 387–445
 - accuracy, 349–353
 - application software for, 409–444
 - behavior, 330–332
 - block diagrams and, 43–49
 - conditioning, 387–445
 - control systems, 329–332, 388–392
 - data acquisition (DAQ), 388–392, 409–444
 - data conversion, 394–409
 - probing, 346
 - real-time interfacing, 387–445
 - sensitivity, 356–357
 - sensor detection of, 132
 - software applications, 409–444
 - stability, 349
 - steady-state error, 330, 349–350
 - strain gauges, enhancement in, 176
 - systems, controls, and, 329–386
 - time delays, 346–348
 - transducers and, 392–394
 - transient region, 330
 - transient response, 353–356
 - Simple relief (direct acting) valves, 279
 - Simulation, 8, 10–11, 50
 - Single input–single output (SISO) system
 - functions, 42–43, 338
 - Sketching procedures, 359–362, 370
 - Bode plots, 370
 - root locus, 359–362
 - Skip control of CDs, example of, 483–485
 - Smart sensors, 30–31
 - Software, DAQ applications, 409–444
 - LabVIEW, 409–423
 - VisSim, 423–444
 - Solenoid force–displacement calibration system, example of, 472–473
 - Solid flow measurement, 196
 - Spool valves, 270–271
 - Spot projection method, 235–236
 - Square root of sum of squares (RSS)
 - method, 141
 - Stability, signal systems, 349
 - State equation (SE), 51–52
 - Steady-state error, 330, 349–350
 - Step angle, stepper motors, 268

- Stepper motors, 262–269
 angle of rotation, 268
 applications of, 262
 block diagrams for, 264–267
 distance moved, 268
 drive equations, 264–265
 modeling approach, 263–264
 motor equations, 265–267
 number of pulses, 268
 permanent magnet (PM), 262–269
 positioning system use of, 267–269
 rotational speed, 268–269
 step angle, 268
 variable reluctance (VR), 262
- Strain gauges, 171–183, 470–472
 acceleration sensing using, 179–180
 axial tensile strain in, 176–177
 bonded, 172–174, 182–183
 cantilever deflection in, 177
 circumferential strains in, 178
 offset voltage and, 175–180
 resistance transducers and, 171–180
 semiconductor, 180
 signal enhancement, 176
 tactile sensors and, 182–183
 temperature effects in, 178–179
 transverse compressive strain in, 176–177
 unbonded, 171–172
 weighing system, example of, 470–472
- Wheatstone bridge circuit arrangement, 174–175, 176, 182–183
- Stress sensing, ultrasonic, 232–233
- Successive approximation type ADC, 407–409
- Summing junctions, 45–46
- Supervisory control and data acquisition (SCADA) systems, 311–312
- Switches, fluid power circuits, 286
- System control, 291–328
 binary logic, 297–302
 Boolean algebra laws, 297
 Karnaugh map (K-map) methods, 302–309
 logic diagrams, 300–302
 number systems, 291–297
- programmable logic controllers (PLC), 309–323
 truth tables, 298–302
- Systems, 42–43, 311–312, 329–386. *See also* Electrical systems; Fluid systems; Mechanical systems
 accuracy, 349–353
 basic feedback (BFS), 341–342, 357–359
 Bode plots, 370–378
 control, 329–338
 controller design, 378–383
 defined, 329
 G-equivalent form, 341–343
 linearization of nonlinear, 343–346
 Pade approximation, 346–348
 pole placement method and, 378–383
 root locus, 357–370
 sensitivity, 356–357
 signals and, 329–386
 single input–single output (SISO), 42–43, 338
 stability, 349
 supervisory control and data acquisition (SCADA), 311–312
 time delays, 346–348
 transfer function form, 338–341
 transient response, 353–356

T

- Tactile sensors, 180–183
- Telescoping property, 346
- Temperature, 20–21, 178–179, 213–216
 condition monitoring of, 20–21
 fiber-optic sensing, 214
 interferometrics used for sensing, 214–216
 radiative sensing, 213–216
 strain gauges, effects in, 178–179
- Temperature measurement, 210–216
 fiber optic, 214
 interferometrics for, 214–216
 radiative, 213–216
 RTD wire, 210–211
 sensors for, 211–212
 thermistors, 211–212, 215
 thermocouples, 212–213

- Thermal cycle fatigue, example of, 476–478
 Thermistors, 211–212, 215
 Thermocouples, 212–213
 Time delay blower, example of, 486–489
 Time delays, signal systems, 346–348
 Time-of-flight method, 237
 Timer Done Bit (DN) instruction, PLC, 314
 Timer Enable Bit (EN) instruction, PLC,
 314–315
 Timer Off Delay (TOF) instruction, PLC, 315
 Timer On Delay (TON) instruction, PLC, 314
 Timer Timing Bit (TT) instruction, PLC, 315
 Trade off evaluation, 28–30
 Transducers, 132–139, 145–246, 392–394,
 468–470
 angular position measurement, 225–228
 capacitance, 146, 154–162
 calibration, example of, 468–470
 digital, 162–168
 eddy current, 216–219
 encoders and, 162–168
 Faraday’s law, 148
 flow measurement, 195–210, 228
 Hall effect, 224–246
 inductance, 146, 148–151
 laser interferometric, 240–242
 linear variable differential transformer
 (LVDT), 151–152
 liquid level measurement, 228
 magnetoresistive, 193–195
 measurement parameters (modules) in,
 135–137, 139
 moiré fringe, 164, 167–168
 optical range sensing, 233–246
 piezoelectric, 146, 183–192
 pneumatic, 228–231
 potentiometers, 146–148
 primary, 145
 principle of transduction classification of,
 146
 quality parameters in, 137–139
 resistance, 146–148, 168–183
 rotary variable differential transformer
 (RVDT), 152–154
 secondary, 145
 sensor classification, 132–135
 signal conditioning and, 392–394
 strain gauges, 171–183
 temperature measurement, 210–216
 thermistors, 211–212
 ultrasonic, 201–203, 231–233
 vibration–acceleration measurement and
 control, 183–195
 Transfer functions, 42–43, 51–59,
 338–341
 block diagram conversion from, 51–59
 linear systems form, 338–341
 ODE conversion, 51–59
 operator notation and, 42–43
 Transient region, 330
 Transient response, signal systems,
 353–356
 Transportation bridge surface testing, example
 of, 467–468
 Triangulation principle, 234–235
 Triggers, 407
 Truth tables, 298–302
 Turbine flow meter, 204–205
- ## U
- Ultrasonic Doppler flow meter, 202–203
 Ultrasonic transducers, 201–203, 231–233
 distance sensing, 232
 Doppler shift technique, 231
 flow sensing, 201–203
 pulse technique, 231
 stress sensing, 232–233
 Unbonded strain gauges, 171–172
 Unloading valves, 279–280
- ## V
- Valves, 278–283
 digital, 281–282
 directional control from, 280–281
 pressure-control, 278–282
 volume-control, 282–283
 Vane motors, 284
 Vane-type pump mechanisms, 277
 Variable-displacement motors, 284
 Variable manipulation module, 136

Variable reluctance (VR) stepper motor, 262
Variable-volume, pressure-compensated, flow-control valves, 283
Vibration, condition monitoring of 20–21
Vibration–acceleration sensors, 183–195
 acceleration measurement, 190–191
 active vibration control, 192–193
 magnetoresistive transducers, 193–195
 piezoelectric transducers, 183–192
 velocity measurement, 191–192
Virtual instruments (VI), 409–415
Virtual machine prototyping, 28
VisSim, 423–444
 application, 430–444
 environment, 423–430
 real-time operation configuration, 426–430

Voltage source, 76–77
Volume-control valves, 282–283

W

Wear, condition monitoring of, 20–21
Wheatstone bridge circuit arrangement, 174–175, 176, 182–183

Y

Young's modulus, 185

Z

Zero/span circuits, 399–400

PRINCIPAL UNITS USED IN MECHANICS

Quantity	International System (SI)			U.S. Customary System (USCS)		
	Unit	Symbol	Formula	Unit	Symbol	Formula
Acceleration (angular)	radian per second squared		rad/s^2	radian per second squared		rad/s^2
Acceleration (linear)	meter per second squared		m/s^2	foot per second squared		ft/s^2
Area	square meter		m^2	square foot		ft^2
Density (mass) (Specific mass)	kilogram per cubic meter		kg/m^3	slug per cubic foot		slug/ft^3
Density (weight) (Specific weight)	newton per cubic meter		N/m^3	pound per cubic foot	pcf	lb/ft^3
Energy; work	joule	J	$\text{N}\cdot\text{m}$	foot-pound		ft-lb
Force	newton	N	$\text{kg}\cdot\text{m/s}^2$	pound	lb	(base unit)
Force per unit length (Intensity of force)	newton per meter		N/m	pound per foot		lb/ft
Frequency	hertz	Hz	s^{-1}	hertz	Hz	s^{-1}
Length	meter	m	(base unit)	foot	ft	(base unit)
Mass	kilogram	kg	(base unit)	slug		$\text{lb}\cdot\text{s}^2/\text{ft}$
Moment of a force; torque	newton meter		$\text{N}\cdot\text{m}$	pound-foot		lb-ft
Moment of inertia (area)	meter to fourth power		m^4	inch to fourth power		in.^4
Moment of inertia (mass)	kilogram meter squared		$\text{kg}\cdot\text{m}^2$	slug foot squared		$\text{slug}\cdot\text{ft}^2$
Power	watt	W	J/s $(\text{N}\cdot\text{m/s})$	foot-pound per second		ft-lb/s
Pressure	pascal	Pa	N/m^2	pound per square foot	psf	lb/ft^2
Section modulus	meter to third power		m^3	inch to third power		in.^3
Stress	pascal	Pa	N/m^2	pound per square inch	psi	lb/in.^2
Time	second	s	(base unit)	second	s	(base unit)
Velocity (angular)	radian per second		rad/s	radian per second		rad/s
Velocity (linear)	meter per second		m/s	foot per second	fps	ft/s
Volume (liquids)	liter	L	10^{-3} m^3	gallon	gal.	231 in.^3
Volume (solids)	cubic meter		m^3	cubic foot	cf	ft^3