

Sensors and Control Systems in Manufacturing

Second Edition

SABRIE SOLOMAN

Sensors and Control Systems in Manufacturing

www.TechnicalBooksPDF.com

About the Author



Sabrie Soloman, Ph.D., Sc.D., MBA, PE, is the Founder, Chairman, and CEO of American SensoRx, Inc., and SensoRx, Chongqing, China, Ltd. He is a professor and founder of advanced manufacturing technology at Columbia University. Dr. Soloman is considered an international authority on advanced manufacturing technology and automation in the microelectronics, automotive, pharmaceuticals, and food industries. He has been and continues to be instrumental in developing and implementing several industrial modernization programs through the United Nations to various European and African countries. Dr. Soloman is the first to introduce and implement nano-MEMS technology in sensors manufacturing and applications. He introduced and implemented unmanned flexible synchronous/asynchronous systems to the microelectronic and pharmaceutical industries and was the first to incorporate advanced vision technology to a wide array of robot manipulators. Dr. Soloman invented the SpectRx NIR sensing technology for the medical and pharmaceuticals industry, as he correlated tablet hardness and disintegration with tablet NIR spectrum.

www.TechnicalBooksPDF.com

Sensors and Control Systems in Manufacturing

Sabrie Soloman
Chairman and CEO, American SensoRx, Inc., USA
Professor, Founder, Advanced Manufacturing Technology, Columbia University, USA
Chairman and CEO, SensoRx, Chongqing, Ltd., China

Second Edition



New York Chicago San Francisco
Lisbon London Madrid Mexico City
Milan New Delhi San Juan
Seoul Singapore Sydney Toronto

www.TechnicalBooksPDF.com

The McGraw-Hill Companies

Copyright © 2010, 1994 by The McGraw-Hill Companies, Inc. All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

ISBN: 978-0-07-160573-1

MHID: 0-07-160573-8

The material in this eBook also appears in the print version of this title: ISBN: 978-0-07-160572-4, MHID: 0-07-160572-X.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol afterevery occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. To contact a representative please e-mail us at bulksales@mcgraw-hill.com.

Information contained in this work has been obtained by The McGraw-Hill Companies, Inc. ("McGraw-Hill") from sources believed to be reliable. However, neither McGraw-Hill nor its authors guarantee the accuracy or completeness of any information published herein, and neither McGraw-Hill nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that McGraw-Hill and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

TERMS OF USE

This is a copyrighted work and The McGraw-Hill Companies, Inc. ("McGraw-Hill") and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill's prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED "AS IS." McGRAW-HILL AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

www.TechnicalBooksPDF.com

humbly dedicate it to them …!

The desert blooming flowers wisely gather the falling droplets of morning dew—guarding them zealously in the hidden corners of their flourishing petals; so as my cherished family Nasier, Anne, Elvira, and Stephan, each has given me the rare and treasured portion of their souls to flourish my mind, enthuse my heart, and blossom my knowledge. It is them who should gather the credit of such book—I humbly dedicate it to them.

www.TechnicalBooksPDF.com

This page intentionally left blank

www.TechnicalBooksPDF.com

Contents

Foreword xxv
Preface xxvii
Acknowledgments xxxi

1 The Role of Sensors in the 21st Century 1
1.0 Introduction 1
1.1 MEMS Microscopy System 5
1.2 Establishing an Automation Program 5
1.3 Photo Sensing Fluorescence in Genome Sequencing 7
1.4 A New Sensing Tool for Decoding the Genome 8
1.5 Mapping RNA Protein Folding Energy Through Bio-Sensors 8
1.6 Nano-Sensor Manufacturing Challenges 9
1.7 Nano-Crystals Enable Scalable Memory Technologies 9
1.8 Chromaticity—Color Rendering Index (CRI) 10
1.9 The LED Color Chart 15
1.10 The Color Rendering Index (CRI) 15
1.10.1 The Foot-Candle 15
1.10.2 Full Spectrum 15
1.10.3 Intensity 15
1.10.4 Kelvin Color Temperature 19
1.11 LEDs—Light-Emitting Diodes 19
1.11.1 LED Bars 19
1.11.2 LED Clusters or Arrays 19
1.11.3 LED Drivers 19
1.11.4 LED Lighting 20
1.11.5 LED Strips 20
1.11.6 Low Voltage 20
1.11.7 Lumen Maintenance 20
1.11.8 Lumens 20
1.11.9 Lux 20
1.11.10 LED mA 20
1.11.11 Max Rated Temperature 21
1.11.12 mcd 21

www.TechnicalBooksPDF.com

1.11.13 Nanometers (nm) 22
1.11.14 Operating Life 22
1.11.15 Printed Circuit Boards 22
1.11.16 Power Supply 22
1.11.17 PWM 22
1.11.18 RGB 22
1.11.19 SMD/SMT 23
1.11.20 SSL 23
1.12 Task Lighting-Lamps 23
1.12.1 UV-A 23
1.12.2 UV-B 23
1.12.3 UV-C 23
1.12.4 View Angle Degree 24
1.12.5 Voltage 24
1.12.6 Voltage Regulator 24
1.12.7 Volts 24
1.12.8 Waterproof 24
1.12.9 Watts 24
1.12.10 Watts per LED 25
1.12.11 Weatherproof 25
1.12.12 White 25
1.13 The Basics on LEDs 25
1.13.1 Monochromatic LEDs 26
1.13.2 LED White Light 27
1.13.3 LED Intensity 28
1.13.4 Visibility 28
1.13.5 Operating Life 29
1.13.6 Voltage/Design Currents 30
1.13.7 Heat Dissipation 30
1.13.8 EMI Countermeasures 30
1.14 Non-Phosphor White LEDs at a Viewing Angle of 30° 30
1.14.1 Non-Phosphor White LEDs 31
1.15 Luminous Intensity (Candlepower) 32
1.15.1 Luminous Flux 33
1.15.2 Illumination 34
1.15.3 Luminance (Brightness) 34
1.15.4 Foot-Candle to Lux Conversion . . . 35
1.15.5 The Inverse Square Law 36
1.15.6 The Cosine Law 36
1.15.7 The Difference Between Lumen and Watt 37
1.15.8 Calculating Beam Angles 37
1.15.9 Calculating Lumens Output—Luminous Flux 38

www.TechnicalBooksPDF.com

1.15.10 Calculating Center and Edge Points—Two Points 38
1.15.11 Calculating Center, Middle, and Edge Points—Three Points 38
1.15.12 Calculating Four Calculation Points 39
1.15.13 Calculating Five Calculation Points 39
1.15.14 Calculating Eight Points 39
1.16 LED and Spectralon 39
1.16.1 Spectralon® Material—Light Reflectance Calibration 39
1.16.2 Spectralon® Features 40
Further Reading 40

2 Classification and Types of Sensors 43
2.0 Introduction 43
2.1 Classification of Control Processes 48
2.2 Open- and Closed-Loop Control Systems 50
2.3 Understanding Photoelectric Sensors 51
2.3.1 Principles of Operation 51
2.3.2 Manufacturing Applications of Photodetectors 52
2.4 Detection Methods 58
2.4.1 The Through-Beam Detection Method 58
2.4.2 The Reflex Detection Method 58
2.4.3 The Proximity Detection Method 59
2.5 Proximity Sensors 61
2.5.1 Typical Applications of Inductive Proximity Sensors 62
2.5.2 Typical Applications of Capacitive Proximity Sensors 64
2.6 Understanding Inductive Proximity Sensors 64
2.6.1 Principles of Operation 64
2.6.2 Inductive Proximity Sensing Range 67
2.6.3 Sensing Distance 68
2.6.4 Target Material and Size 69
2.6.5 Target Shape 71
2.6.6 Variation Between Devices 72
2.6.7 Surrounding Conditions 73
2.7 Understanding Capacitive Proximity Sensors 75

www.TechnicalBooksPDF.com

2.7.1 Principles of Operation 75
2.7.2 Features of Capacitive Sensors 77
2.7.3 Sensing Range 77
2.7.4 Target Material and Size 78
2.7.5 Surrounding Conditions 79
2.8 Understanding Limit Switches 79
2.9 Inductive and Capacitive Sensors in Manufacturing 80
2.9.1 Relays 81
2.9.2 Triac Devices 82
2.9.3 Transistor DC Switches 84
2.9.4 Inductive and Capacitive Control/Output Circuits 85
2.9.5 Accessories for Sensor Circuits 87
2.9.6 Inductive and Capacitive Switching Logic 88
2.9.7 Inductive and Capacitive Sensor Response Time—Speed of Operation 93
2.10 Understanding Microwave Sensing Applications 98
2.10.1 Characteristics of Microwave Sensors 98
2.10.2 Principles of Operation 99
2.10.3 Detecting Motion with Microwave Sensors 100
2.10.4 Detecting Presence with Microwave Sensors 104
2.10.5 Measuring Velocity with Microwave Sensors 105
2.10.6 Detecting Direction of Motion with Microwave Sensors 105
2.10.7 Detecting Range with Microwave Sensors 106
2.10.8 Microwave Technology Advancement 109
2.11 Confocal Microscopy Sensors 110
2.11.1 Confocal Profiling Characterization Systems 110
2.11.2 Driving Inspection Requirements 111
2.11.3 Probe-Mark Inspection Requirements 113
2.11.4 Multi-Chip Packages 113

www.TechnicalBooksPDF.com

2.12 Understanding Laser Sensors 113
2.12.1 Properties of Laser Light 114
2.12.2 Essential Laser Components 114
2.12.3 Semiconductor Displacement Laser Sensors 119
2.12.4 Industrial Applications of Laser Sensors 119
Further Reading 132

3 Fiber Optics in Sensors and Control Systems 133
3.0 Introduction 133
3.1 Photoelectric Sensors—Long-Distance Detection 133
3.1.1 Light-Emitting Diodes 134
3.1.2 Through-Beam Sensors 136
3.1.3 Reflex Photoelectric Controls 137
3.1.4 Polarized Reflex Detection 138
3.1.5 Proximity (Diffuse-Reflection) Detection 139
3.1.6 Automated Guided Vehicle System 140
3.2 Fiber Optics 141
3.2.1 Individual Fiber Optics 143
3.2.2 Bifurcated Fiber Optics 143
3.3 Optical Fiber Parameters 145
3.3.1 Excess Gain 145
3.3.2 Background Suppression 147
3.3.3 Contrast 148
3.3.4 Polarization 148
3.4 Inductive Proximity Sensors—Noncontact Metal Detection 148
3.5 Limit Switches—Traditional Reliability 150
3.6 Factors Affecting the Selection of Position Sensors 151
3.7 Wavelengths of Commonly Used Light-Emitting Diodes 152
3.8 Sensor Alignment Techniques 152
3.8.1 Opposed Sensing Mode 152
3.8.2 Retroreflective Sensing Mode 152
3.8.3 Proximity (Diffuse) Sensing Mode . . . 153
3.8.4 Divergent Sensing Mode 153
3.8.5 Convergent Sensing Mode 154
3.8.6 Mechanical Convergence 154
3.9 Fiber Optics in Industrial Communication and Control 155

www.TechnicalBooksPDF.com

3.10 Principles of Fiber Optics in Communications 155
3.11 Fiber-Optic Information Link 157
3.12 Configurations of Fiber Optics 157
3.12.1 Optical Power Budget 158
3.12.2 Digital Links—Pulsed 159
3.12.3 Digital Links—Carrier-Based 160
3.12.4 Analog Links 161
3.12.5 Video Links 161
3.12.6 Data Bus Networks 162
3.13 Configurations of Fiber Optics for Sensors 165
3.13.1 Fiber-Optic Bundles 165
3.13.2 Bundle Design Considerations 167
3.13.3 Fiber Pairs for Remote Sensing 168
3.13.4 Fiber-Optic Liquid Level Sensing 169
3.14 Flexibility of Fiber Optics 170
3.14.1 Fiber-Optic Terminations 170
3.15 The Testing of Fiber Optics 172
3.16 Testing Light Sources 172
3.16.1 Power Meters 173
3.17 Dual Laser Test Sets 175
3.17.1 Test Sets/Talk Sets 176
3.17.2 Attenuators 177
3.17.3 Fault Finders 177
3.17.4 Fiber Identifiers 178
3.18 Networking with Electrooptic Links 179
3.18.1 Hybrid Wire/Fiber Networks 180
3.18.2 Daisy Chain Networks 181
3.18.3 Active Star Networks 181
3.18.4 Hybrid Fiber Networks 182
3.18.5 Fiber-Optic Sensory Links for Minicell Controllers 184
3.19 Versatility of Fiber Optics in Industrial Applications 184
3.19.1 High-Clad Fiber-Optic Cables 185
3.19.2 Fiber-Optic Ammeter 189
Further Reading 189

4 Networking of Sensors and Control Systems in Manufacturing 191
4.1 Introduction 191
4.2 The Number of Products in a Flexible System 193

www.TechnicalBooksPDF.com

4.3 Sensors Tracking the Mean Time between Operator Interventions 194
4.4 Sensors Tracking the Mean Time of Intervention 194
4.5 Sensors Tracking Yield 194
4.6 Sensors Tracking the Mean Processing Time 194
4.7 Network of Sensors Detecting Machinery Faults 196
4.7.1 Diagnostic Systems 196
4.7.2 Resonance and Vibration Analysis 197
4.7.3 Sensing Motor Current for Signature Analysis 197
4.7.4 Acoustics 198
4.7.5 Temperature 198
4.7.6 Sensors for Diagnostic Systems . . . 198
4.7.7 Quantifying the Quality of a Workpiece 199
4.7.8 Evaluation of an Existing Flexible Manufacturing Cell Using a Sensing Network 199
4.8 Understanding Computer Communications and Sensors' Role 207
4.8.1 Application Layer Communication 209
4.8.2 Presentation Layer Communication 209
4.8.3 Session Layer Communication 210
4.8.4 Transport Layer Communication 210
4.8.5 Network Layer Communication . . . 210
4.8.6 Data Link Layer Communication by Fiber Optics or Coaxial Cable 210
4.8.7 Physical Layer Communication . . . 210
4.8.8 Adding and Removing Information in Computer Networks Based on Open System Interconnect (OSI) 211
4.9 Understanding Networks in Manufacturing 213
4.9.1 RS-232-Based Networks 213
4.9.2 Ethernet 215
4.9.3 Transmission Control Protocol (TCP)/Internet Protocol (IP) 216

www.TechnicalBooksPDF.com

4.10 Manufacturing Automation Protocol 216
4.10.1 Broadband System for MAP Protocol 217
4.10.2 Carrier-Band System for MAP Protocol 219
4.10.3 Bridges MAP Protocol 219
4.10.4 Token Systems for MAP Protocol 220
4.11 Multiple-Ring Digital Communication Network—AbNET 221
4.12 The Universal Memory Network 223
4.13 Satellite Sensor Network 225
Further Reading 226

5 The Role of Sensors and Control Technology in Computer-Integrated Manufacturing 229
5.1 Introduction 229
5.2 The CIM Plan 230
5.2.1 The CIM Plan in Manufacturing 231
5.2.2 The CIM Plan in Engineering and Research 231
5.2.3 The CIM Plan in Production Planning 231
5.2.4 The CIM Plan in Physical Distribution 231
5.2.5 The CIM Plan for Business Management 231
5.2.6 The CIM Plan for the Enterprise 232
5.3 The Manufacturing Enterprise Model 232
5.3.1 Marketing 234
5.3.2 Engineering and Research 235
5.3.3 Production Planning 238
5.3.4 Plant Operations 239
5.3.5 Physical Distribution 242
5.3.6 Business Management 244
5.4 Design of CIM with Sensors and Control Systems 246
5.4.1 Components of CIM with Sensors and Control Systems 247
5.4.2 CIM with Sensors and Control Systems at the Plant Level 248
5.5 Decision Support System for CIM with Sensors and Control Systems 252

www.TechnicalBooksPDF.com

5.5.1 Computer-Integrated Manufacturing Database (CIM DB) 252
5.5.2 Structure of Multiobjective Support Decision Systems 253
5.6 Analysis and Design of CIM with Sensors and Control Systems 254
5.6.1 Structured Analysis and Design Technique (SADT) 255
5.6.2 A Multiobjective Approach for Selection of Sensors in Manufacturing 256
5.7 Data Acquisition for Sensors and Control Systems in CIM Environments 257
5.7.1 Real-World Phenomena 257
5.7.2 Sensors and Actuators 257
5.7.3 Signal Conditioning 259
5.7.4 Data Acquisition for Sensors and Control Hardware 259
5.7.5 Computer System 261
5.7.6 Communication Interfaces 261
5.7.7 Software 262
5.8 Developing CIM Strategy with Emphasis on Sensors' Role in Manufacturing 262
5.8.1 CIM and Building Blocks 263
5.8.2 CIM Communications 264
5.8.3 Plant Floor Communications 264
5.8.4 Managing Data in the CIM Environment 266
5.8.5 CIM Environment Presentation . . . 267
5.8.6 The Requirement for Integration . . . 269
Further Reading 273

6 Advanced Sensor Technology in Precision Manufacturing Applications 275
6.1 Identification of Manufactured Components 275
6.1.1 Bar-Code Identification Systems 275
6.1.2 Transponders 277
6.1.3 Electromagnetic Identification of Manufactured Components 277
6.1.4 Surface Acoustic Waves 278
6.1.5 Optical Character Recognition 278
6.2 Digital Encoder Sensors 278
6.2.1 Position Encoder Sensors in Manufacturing 280

www.TechnicalBooksPDF.com

6.3 Fuzzy Logic for Optoelectronic Color Sensors in Manufacturing 282
6.3.1 Sensing Principles 282
6.3.2 Color Theory 283
6.3.3 Units of Color Measurement 285
6.3.4 Color Comparators and True Color Measuring Instruments 286
6.3.5 Color Sensor Algorithms 287
6.4 Optimum Detectors in Light Sensors 288
6.4.1 Available Light Sensing Options 288
6.5 Photodiodes 289
6.6 Phototransistors and Photodarlingtons 291
6.7 Photoconductive Sensors 292
6.7.1 Integrated Circuits 293
6.7.2 Hybrids 293
6.8 Sensor Electronic Assemblies 293
6.9 Other Sensors 294
6.10 Selecting a Sensor 294
6.11 Infrared Light Domain 294
6.12 Visible Light Region 295
6.13 Nanowatts Power 295
6.14 Considering High-Performance Photodiodes 295
6.15 Hybrids of Photomultipliers Options 295
6.16 Design Considerations in Fuzzy Logic Color Sensors 295
6.16.1 Fuzzy Logic Controller Flowcharts 296
6.17 Sensors Detecting Faults in Dynamic Machine Parts (Bearings) 298
6.18 Sensors for Vibration Measurement of a Structure 300
6.19 Optoelectronic Sensor Tracking Targets on a Structure 302
6.20 Optoelectronic Feedback Signals for Servomotors through Fiber Optics 303
6.21 Acoustooptical/Electronic Sensors for Synthetic-Aperture Radar Utilizing Vision Technology 305
6.22 The Use of Optoelectronic/Vision Associative Memory for High-Precision Image Display and Measurement 307
6.23 Sensors for Hand-Eye Coordination of Microrobotic Motion Utilizing Vision Technology 309

www.TechnicalBooksPDF.com

6.24 Force and Optical Sensors Controlling Robotic Grippers for Agriculture and Manufacturing Applications 310
6.25 Ultrasonic Stress Sensor Measuring Dynamic Changes in Materials 312
6.26 Predictive Monitoring Sensors Serving the CIM Strategy 314
6.27 Reflective Strip Imaging Camera Sensor—Measuring a 180°-Wide Angle 315
6.28 Optical Sensor Quantifying Acidity of Solutions 317
6.29 Sensors for Biomedical Technology 319
6.29.1 Sensor for Detecting Minute Quantities of Biological Materials 319
6.29.2 Sensors for Early Detection and Treatment of Lung Tumors 320
6.29.3 Ultrasensitive Sensors for Single-Molecule Detection 321
Further Reading 323

7 Industrial Sensors and Control 325
7.1 Introduction 325
7.2 Sensors in Manufacturing 328
7.3 Temperature Sensors in Process Control 329
7.3.1 Semiconductor Absorption Sensors 330
7.3.2 Semiconductor Temperature Detector Using Photoluminescence 331
7.3.3 Temperature Detectors Using Point-Contact Sensors in Process Manufacturing Plants 334
7.3.4 Noncontact Sensors—Pyrometers 334
7.4 Pressure Sensors 337
7.4.1 Piezoelectric Crystals 337
7.4.2 Strain Gauges 338
7.5 Fiber-Optic Pressure Sensors 338
7.6 Nano-Positioning Capacitive Metrology Sensors 340
7.6.1 Nano-Capacitive Positioning Sensors 340
7.7 Electrode Geometry, Sensor Surface Flatness, and Finish 343
7.8 Special Design Eliminates Cable Influences 343

www.TechnicalBooksPDF.com

7.9 Materials Achieving Greater Accuracy 343
7.10 Mounting, Calibration, and Measuring Ranges 344
7.11 Parallelism of Measuring Surfaces 345
7.12 Electronics Support 345
7.13 Sensor Installation 348
7.13.1 Factory Calibration for Improved Linearity 348
7.14 Using the Integrated Linearization System (ILS) for Highest Accuracy 348
7.14.1 Multichannel Measurements 348
7.15 Displacement Sensors for Robotic Applications 348
7.16 Process Control Sensors Measuring and Monitoring Liquid Flow 350
7.16.1 Flow Sensors Detecting Small Air Bubbles for Process Control in Manufacturing 351
7.16.2 Liquid Level Sensors in Manufacturing Process Control for Petroleum and Chemical Plants 353
7.17 Sensory MEMS Enable Certain Molecules to Signal Breast Cancer's Spread 355
7.18 On-Line Measuring and Monitoring of Gas by Spectroscopy 356
7.19 Using Avalanche Photodiodes to Improve System Performance 359
7.20 Structures of Avalanche Photodiodes—APD Structures 360
7.21 Critical Performance Parameters 362
7.22 Selecting an APD 363
7.22.1 Types of APDs 364
7.22.2 Understanding the Specifications Responsivity and Gain 364
7.23 Dark Current and Noise Current 365
7.23.1 Excess Noise Factor 365
7.24 The Geiger Mode 367
7.25 Crack Detection Sensors for Commercial, Military, and Space Industry Use 367
7.26 Control of the Input/Output Speed of Continuous Web Fabrication Using Laser Doppler Velocity Sensor 369
7.27 Ultrasonic/Laser Nondestructive Evaluation Sensor 370

www.TechnicalBooksPDF.com

7.28 Process Control Sensors for Acceleration ... 371
7.29 An Endoscope as Image Transmission Sensor ... 372
7.30 Sensor Network Architectures in Manufacturing ... 373
7.31 Power Line Fault-Detection Systems for Power Generation and Distribution Industries ... 375
Further Reading ... 376

8 Sensors in Flexible Manufacturing Systems 379
8.1 Introduction ... 379
8.2 The Role of Sensors in FMS ... 380
8.2.1 Current Available Sensor Technology for FMS ... 381
8.3 Robot Control through Vision Sensors ... 383
8.3.1 Image Transformation ... 383
8.3.2 Robot Vision and Human Vision ... 384
8.3.3 Robot Vision and Visual Tasks ... 384
8.3.4 Robot Visual Sensing Tasks ... 386
8.3.5 Robots Utilizing Vision Systems to Recognize Objects ... 386
8.4 Robot Vision Locating Position ... 388
8.5 Robot Guidance with Vision System ... 389
8.5.1 Robot Vision Performing Inspection Tasks ... 389
8.5.2 Components of Robot Vision ... 390
8.6 End Effector Camera Sensor for Edge Detection and Extraction ... 391
8.6.1 Shape and Size ... 392
8.6.2 Position and Orientation ... 392
8.6.3 Multiple Objects ... 393
8.7 End Effector Camera Sensor Detecting Partially Visible Objects ... 395
8.7.1 Runtime Phase ... 400
8.8 Ultrasonic End Effector ... 400
8.9 End Effector Sound-Vision Recognition Sensors ... 400
8.9.1 Standoff ... 402
8.9.2 Large Surface Measurements ... 402
8.9.3 Sensitivity of Measurements ... 404
8.9.4 Small Surfaces ... 404
8.9.5 Positioning ... 405
8.10 End Effector Linear Variable-Displacement Transformer Sensor ... 407

www.TechnicalBooksPDF.com

8.10.1 Extreme Environments 409
8.10.2 Cryogenic Manufacturing Applications 410
8.10.3 Measurement at High Temperatures in Manufacturing 411
8.11 Robot Control through Sensors 411
8.12 Multisensor-Controlled Robot Assembly 412
8.12.1 Control Computer 415
8.12.2 Vision Sensor Modules 416
8.12.3 Software Structure 416
8.12.4 Vision Sensor Software 417
8.13 History of Industrial Robotics 417
8.13.1 Early Modern Developments 418
8.13.2 Modern Developments 418
8.14 The Invention of Industrial Robots 419
8.14.1 Technical Description 421
8.14.2 Robot Programming and Interfaces 423
8.14.3 End Effectors 424
8.14.4 Movement and Singularities 425
8.15 Robot Programming 426
8.16 The 2006 World Robot Market—Total Worldwide Sales 428
8.17 Measurements of Robot Density Based on the Total Number of Persons Employed 434
8.17.1 Robot Densities—One Robot per Ten Workers in the Motor Vehicle Industry 435
8.17.2 Installations of Advanced Multipurpose Industrial Robots by Type 435
8.17.3 Distribution of Service Robots for Professional Use 435
8.17.4 Distribution of Service Robots for Personal and Private Use 435
8.18 Projections for the Period 2007–2010 436
Further Reading 437

9 Communications 439
9.1 Introduction 439
9.2 Single-Board Computer 439
9.3 Sensors for Input Control 440
9.4 Microcomputer Interactive Development System 444

www.TechnicalBooksPDF.com

9.5 The Personal Computer as a Single-Board Computer 445
9.5.1 Role of Sensors in Programmable Logic Controllers—PLC 446
9.5.2 Central Control Units 448
9.5.3 Process Computer 450
9.6 The NC Controller 450
9.6.1 Manufacturing Procedure and Control 451
9.6.2 The Machining Program 452
9.6.3 Absolute Control 457
9.6.4 NC Software 458
9.6.5 Operation of an NC System 462
9.6.6 Computer Numerical Control System 466
9.7 Industrial Handling 469
9.8 Packaging Technology 473
9.9 Linear Indexing for Manufacturing Applications 475
9.10 Synchronous Indexing for Manufacturing Applications 477
9.11 Parallel Data Transmission 478
9.12 Serial Data Transmission 480
9.13 The Collection and Generation of Process Signals in Decentralized Manufacturing Systems 483
Further Reading 486

10 SpectRx NIR Technology 487
10.1 An Introduction to SpectRx NIR Technology 487
10.2 Abbreviations—Nomenclature 487
10.3 Glossary 490
10.4 Formula and Mathematical Relations 491
10.5 The SpectRx FFT-NIR Technology Advantage 492
10.5.1 Software Philosophy 493
10.5.2 Data Flow Outlines 494
10.6 Radiometric and Spectral Theory 495
10.7 Sensitivity 495
10.7.1 Object View Spectral Radiation 495
10.7.2 Transmission of the Spectroradiometer 497
10.7.3 Throughput 497
10.7.4 Power at the Detector 497

www.TechnicalBooksPDF.com

10.7.5 Detector Current 498
10.7.6 Sources of Noise 498
10.7.7 Photon Noise 498
10.7.8 Johnson Noise (InSb Detectors) . . . 499
10.7.9 Quantization Noise 503
10.7.10 Dark Noise 503
10.7.11 Jitter Noise 504
10.7.12 Scanning Instability Noise 504
10.7.13 Instrument Efficiency 504
10.7.14 NESR 505
10.7.15 Signal-to-Noise Ratio 506
10.8 Instrument Line Shape and Spectral Resolution 506
10.8.1 Spectral Resolution 507
10.8.2 Apodization 508
10.9 Radiometric Calibration 509
10.9.1 Calibration Theory 510
10.9.2 Multiple Point Calibrations 512
10.9.3 Linear Multiple Point Calibrations 512
10.9.4 Nonlinear Multiple Point Calibrations 513
10.10 Radiometric Accuracy 515
10.10.1 Calibration Source Errors 515
10.10.2 Calibration Drift 517
10.10.3 Intrinsic Linearity 519
10.10.4 Detector Nonlinearity Errors 520
10.10.5 Channel Spectrum Error 520
10.10.6 Spectral Aliasing Error 522
Further Reading 522

11 Economic and Social Interests in the Workplace 523
11.1 Manufacturing Operation Control Through Financial Planning 523
11.1.1 Developing a Plan 524
11.1.2 Sales Budget 524
11.1.3 Product Budget 524
11.1.4 Selling and Administrative Budget 525
11.1.5 Capital Budget 526
11.1.6 Financial Budget 526
11.1.7 Income Statement and Balance Sheet Projections 526
11.1.8 Planning for Profit 527

www.TechnicalBooksPDF.com

11.1.9 Controlling Operation 527
11.1.10 Cost and Profit Analysis 528
11.2 Information Assimilation and Decision Making 529
11.2.1 Breakeven Analysis 529
11.2.2 Breakeven Chart and Formula 529
11.2.3 Utilization of Breakeven Analysis 531
11.2.4 Incremental Analysis 532
11.2.5 The Large New Production Order 532
11.2.6 Make versus Buy 533
11.2.7 The Importance of the Basic Data 534
11.3 Communication 535
11.3.1 The Need for Timeliness 536
11.3.2 The Reporting Strategy 537
11.3.3 Responsibility Centers 538
11.3.4 Controllable and Uncontrollable Costs 538
11.3.5 Construction of the Report 540
11.3.6 Analyzing Cost and Profit Data 540
11.3.7 Communication of Business Financial Status 541
11.4 Mathematical Methods for Planning and Control 542
11.4.1 Dealing with Uncertainty 542
11.4.2 Objective Probability 542
11.4.3 Subjective Probability 543
11.4.4 Capital Budget 544
11.4.5 Cost of the Fixed Asset 544
11.4.6 The Expected Net Cash Flow 544
11.4.7 Opportunity Cost 544
11.4.8 The Present Value 544
11.4.9 Inventory Analysis 547
11.4.10 Linear Programming 548
11.4.11 The Profit Function 549
11.4.12 The Constraints 549
11.4.13 Project Management 551
11.4.14 Simulation 555
11.5 Where Do Sensors and Control Systems Take Us? 556
Further Reading 563

Index 565

www.TechnicalBooksPDF.com

This page intentionally left blank

www.TechnicalBooksPDF.com

Foreword

The ongoing revolution in manufacturing, well into its third decade, has now (albeit somewhat belatedly) become recognized and joined by government, industry, and academe in this country. It is based on a number of concepts which have made their way into the professional jargon and have been brought to the public's awareness by technical and business writers: concurrent engineering, flexible manufacturing, just-in-time inventory, agile manufacturing, automation, and quality engineering. Each of these are ingredients that contribute to the ultimate goal, which, simply stated, is to achieve the highest quality products at the lowest possible cost, and to do so in a timely fashion. A tall order, but one on which depends the welfare of a host of individual companies and, even more importantly, the economic health of entire countries, with political and social implications beyond overstatement.

A principal ingredient in the process, perhaps the most important one, is the achievement and implementation of error-free production, and at the same time, a guarantor of quality and a minimizer of waste of materials and labor. At first impression, the term "error-free" will sound like a pious ideal, to be striven for but impossible to attain. A moment of reflection will persuade, however, that the aim need not be a philosophical abstraction. In the final analysis, it is the end product *alone* that must fall within the range of prescribed tolerances, not each of the many steps in the production process. That is to say, given within the context of computer-integrated manufacturing a sufficient array of monitors distributed throughout the workspace—i.e., sensors (and appropriate means to feed back and respond to, in real time, the information gathered by them) and control systems which can identify, rectify, or remove defects in the course of production—every item that reaches the end of the production line will be, ipso facto, acceptable.

Professor and Chairman Sabrie Soloman labored hard to bring forth the book before the reader, which contains not only an exposition of sensors and controls, but a host of invaluable asides, comments, and extended discourses on key topics of modern manufacturing.

www.TechnicalBooksPDF.com

The author is an active practitioner of advanced manufacturing techniques and a highly regarded teacher of the subject. In this volume, he brings error-free production a step nearer to realization; the world of manufacturing engineering owes him a debt of gratitude.

H. Deresiewicz
Chairman Department of Mechanical Engineering
Columbia University

www.TechnicalBooksPDF.com

Preface

Rising costs, shorter lead times, complex customer specifications, competition from across the street—and around the world.

Business today faces an ever-increasing number of challenges. The manufacturers that develop more effective and efficient forms of production, development, and marketing will be the ones who meet these challenges.

The use of advanced sensors and control technology makes a fundamental commitment to manufacturing solutions based on simple and affordable integration. With this technology, one can integrate manufacturing processes, react to rapidly changing production conditions, help personnel to react more effectively to complex qualitative decisions, and lower the cost of and improve product quality throughout the manufacturing enterprise.

The first step in achieving such flexibility is to establish an information system that can be reshaped whenever necessary, thus enabling it to respond to the changing requirements of the enterprise—and the environment. This reshaping must be accomplished with minimal cost and disruption to the ongoing operation.

Sensors and control technology will play a key role in achieving flexibility in the information system. However, this technology alone can not shorten lead time, reduce inventories, and minimize excess capacity to the extent required by today's manufacturing operation. This can be accomplished only by integrating various sensors with appropriate control means throughout the manufacturing operation within *computer integrated manufacturing* (CIM) strategy. The result is that individual manufacturing processes will be able to flow, communicate, and respond together as a unified cell, well structured for their functions.

In order to develop a sensory and control information system that will achieve these objectives, the enterprise must start with a specific long-range architectural strategy, one providing a foundation that accommodates today's needs as well as taking those of tomorrow's—including the support of new manufacturing processes, incorporating

www.TechnicalBooksPDF.com

new data functions, and establishing new data bases and distributed channels—into account. The tools for this control and integration are available today.

The United States leads the world in inventing new products; however, many of these new products, ultimately, are manufactured by other countries. The inability of the United States manufacturers to compete globally cannot be blamed on low-cost labor in other countries; more than one-half the trade deficit comes from foreign industries that pay higher wages. The inability to apply affordable manufacturing systems for automation can be a contributing factor to this dilemma.

An *affordable manufacturing system* is, simply, a system that contains a variety of reliable parts, harmoniously joined together to generate a specific motion that will achieve a particular manufacturing operation, directed and controlled by simple and effective sensors and control systems. Modern manufacturing technology is prevalent in the design and control of engineering systems and also in applying the sensory and control technology in production systems to situations such as product fabrication and assembly.

When considering the design aspect of a system the following must be taken into account: dynamics, kinematics, statics, and even styling of parts. All of these play a vital role in forming optimum manufacturing parameters in system design. Motion generation and control through various sensors provide a review of manufacturing engineering concepts from a system's point of view, directed toward the manufacturing engineering problems.

Manufacturing engineering plays a key role in translating new product specifications from design engineering into process plans which are then used to manufacture the product. As the product is being designed, manufacturing technical evaluators work with design engineers to ascertain if sensors and control systems for process monitoring and control can be integrated into the design of a system, and at what cost. Tolerances, materials, clearances, appropriate handling of parts, acceptable types and positioning of sensors, and product assembly times are particularly important factors in this evaluation because they directly affect productivity, the guidelines of which are essential to this analysis.

Cost estimates are equally important. If a new process is needed because, for example, existing processes are deemed too expensive or incapable of producing the desired product, process engineers would be asked to either develop a new economical process or to change product design. Choosing the best alternative could be very difficult because such a decision is based on many conflicting objectives, for example, customer specifications, cost, feasibility, timeliness, market share, parts availability, standard parts, availability of part tooling, sensors locations, and the like. A well-designed system incorporating sensors and control technology for the generation of motion for a

www.TechnicalBooksPDF.com

manufacturing system plays a fundamental role in the new manufacturing thinking.

Advanced sensory and control technology is more than an implementation of new technologies. It is a long-range strategy that allows components of the entire manufacturing operation to work together to achieve the qualitative and quantitative goals of business. It must have top management commitment and may entail changing the mind-set of people in the organization and managing that change. However, the rewards are great since successful implementation of this technology is, in large measure, responsible for the success of computer-integrated manufacturing strategy today.

Sabrie Soloman

www.TechnicalBooksPDF.com

This page intentionally left blank

www.TechnicalBooksPDF.com

Acknowledgments

Mere thanks is insufficient to Harleen Chopra, project manager; Michael McGee, copyeditor; and Vijay Sadyal, compositor for their relentless efforts and unfaltering diligence to bring the work of this book to the high standard of world-class scientific and technical publications.

Furthermore, Mr. Taisuke Soda, Senior Editor, McGraw-Hill Professional, New York, has my sincere gratitude for his insight and distinctive wisdom in guiding me through the composition of this book.

Also, this book was made possible by the efforts of my colleagues and friends in various universities and industries, and by the encouragement and assistance of my staff at Columbia University, New York, American SensoRx, Inc., USA, and SensoRx, Chongqing, Ltd. China.

www.TechnicalBooksPDF.com

This page intentionally left blank

www.TechnicalBooksPDF.com

Sensors and Control Systems in Manufacturing

www.TechnicalBooksPDF.com

This page intentionally left blank

www.TechnicalBooksPDF.com

CHAPTER 1

The Role of Sensors in the 21st Century

1.0 Introduction

Integrated sensors and control systems are the way of the future. In times of disaster, even the most isolated outposts can be linked directly into the public telephone network by portable versions of satellite earth stations called *very small aperture terminals* (*VSAT*s). They play a vital role in relief efforts such as those for the eruption of Mount Pinatubo in the Philippines, the massive oil spill in Valdez, Alaska, the 90,000-acre fire in the Idaho forest, and Hurricane Andrew's destruction in south Florida and the coast of Louisiana.

VSATs are unique types of sensors and control systems. They can be shipped and assembled quickly and facilitate communications by using more powerful antennas that are much smaller than conventional satellite dishes. These types of sensors and control systems provide excellent alternatives to complicated conventional communication systems, which in disasters often experience serious degradation because of damage or overload.

Multispectral sensors and control systems will play an expanding role to help offset the increasing congestion on America's roads by creating "smart" highways. At a moment's notice, they can gather data to help police, tow trucks, and ambulances respond to emergency crises. Understanding flow patterns and traffic composition would also help traffic engineers plan future traffic control strategies. The result of less congestion would be billions of vehicle hours saved each year.

Figure 1.1 shows the Magellan spacecraft as it was close to completing its third cycle of mapping the surface of the planet Venus. The key to gathering data is the development of a synthetic aperture radar as a sensor and information-gathering control system, the sole

www.TechnicalBooksPDF.com

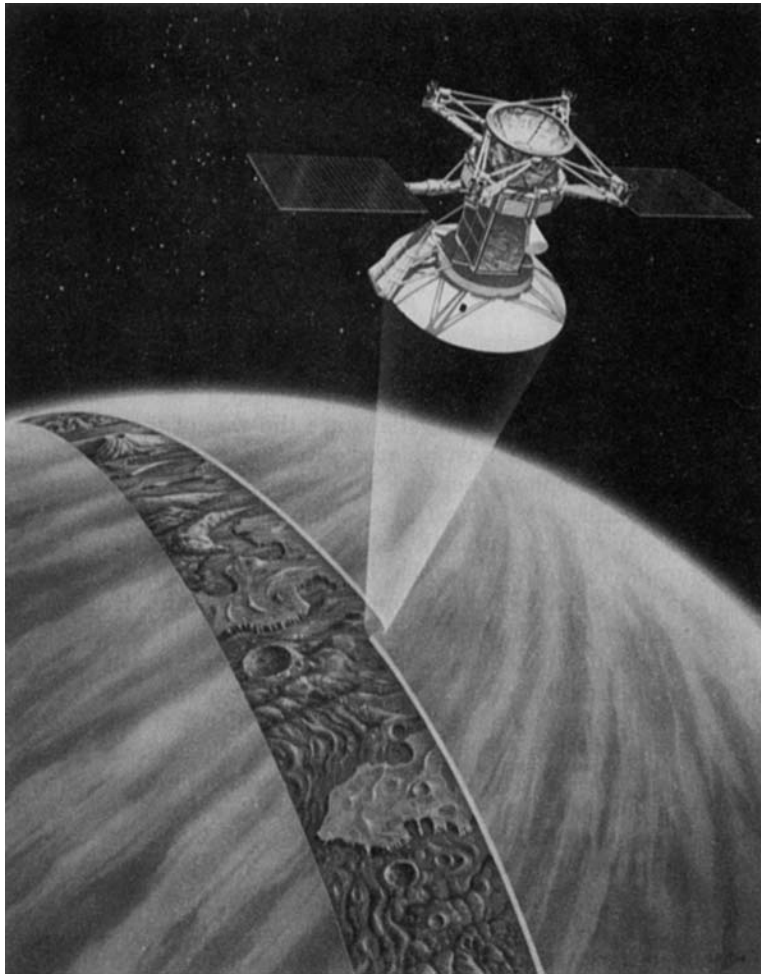


Figure 1.1 The Magellan spacecraft. (*Courtesy Hughes Corp.*)

scientific instrument aboard Magellan. Even before the first cycle ended, in mid-1991, Magellan had mapped 84 percent of Venus' surface, returning more digital data than all previous U.S. planetary missions combined, with resolutions ten times better than those provided by earlier missions. To optimize radar performance, a unique and simple computer software program was developed, capable of handling nearly 950 commands per cycle. Each cycle takes one Venusian day, which is the equivalent of 243 Earth days.

If only simple adequate sensors had been in place, such a disproportional disaster like the 2004 Indian Ocean tsunami (Fig. 1.2), could have been avoided or reduced.

www.TechnicalBooksPDF.com



Figure 1.2 Indian Ocean Tsunami

The Indian Ocean tsunami shook the world. The magnitude 9.3 earthquake that spawned its catastrophic waves was the second most powerful on record. The waves traveled with such force that, seven hours after the earthquake, they killed almost 300 people on Somalia's coast. However, 15 minutes after the earthquake released its energy into the ocean, tsunami waves hit Indonesia's Aceh Province with unimaginable force.

Survivors describe hearing sounds like a great rumbling thunder or a loud wind. They talk about seeing the ocean recede and then, moments later, having an "oily black wave" bearing down on their villages. For most, the landfall and immediate aftermath are too horrific to describe.

Although thousands also perished in places like India, Sri Lanka, and Thailand, Aceh Province was by far the hardest hit by the tsunami: nearly 170,000 of the 230,000 estimated deaths were along the Acehnese coast, in tiny fishing villages and sprawling cities. At least 500,000 were instantly homeless. Unfortunately, Aceh was ill-equipped to deal with such devastation. If only simple adequate sensors had been in place, such a disproportional disaster could have been avoided or its devastation reduced.

For more than a decade, Aceh had suffered from the constant violence of a civil war. The province was essentially shut off from the rest of Indonesia and, as a result, the rest of the world. There was a sense of isolation and distrust even among neighboring villages.

The tsunami changed everything, immediately. In some sense, the slate was wiped clean; the past mattered little. Help was needed from any available source, whether former foe or friend.

www.TechnicalBooksPDF.com

Mercy Corps was one of the first international relief and development organizations to enter and begin an emergency response in the once-closed province. Within a few days of the tsunami, we had almost 30 veteran staff on the ground in Aceh, distributing critical provisions and relief supplies for more than 100,000 survivors.

With assistance from international humanitarian organizations, communities pulled together like never before to reopen schools and hospitals, clean debris from streets, and care for survivors. Mercy Corps helped encourage this renaissance of community through an extensive cash-for-work program: tsunami survivors were paid a daily living wage for their work, which infused flagging local economy with much-needed cash. It was a new beginning for Aceh in the wake of the world's worst recent disaster. It was an opportunity for peace; indeed, an agreement was finalized between the government and separatist forces just a few months later.

With collaboration from communities, local organizations and government, Mercy Corps helped survivors quickly move on with their lives.

"One of our greatest strengths is moving seamlessly from relief and recovery to long-term development of communities affected by disaster or crisis," stated Mark Ferdig, Mercy Corps' Program Director in Aceh. "We are laying the groundwork for economic, social, and technical sensors warning development. The goal is to give communities the tools they need to succeed long after such disasters."

As a result, more than 400 small businesses reopened or were created through an innovative Financial Access program, including:

17 local banks supported by Mercy Corps funds and training

21,000 farmers involved in organic gardening and other agriculture programs

879 community groups formed to address ongoing issues and ensure the sustainability of programs

104 health clinics and hospitals rehabilitated or supported with critical equipment and supplies

Three years after the Indian Ocean tsunami—and more than a decade after violent conflict engulfed its population—Mercy Corps remains to help Aceh rebuild and become better.

The results are readily apparent in rebuilt villages and bright green crop fields.

As they near the end of a long post-disaster reconstruction, the Acehnese (people) are rising: more confident, capable, and collaborative than ever before.

www.TechnicalBooksPDF.com

1.1 MEMS Microscopy System

In the 17th century, scientist Anton van Leeuwenhoek peered at microbes through a crude optical microscope, for the quest of higher performance. Just as Leeuwenhoek's microscope impacted the science of his day, today's microscopy systems affect the way engineers develop and test semiconductors and materials, as well as technologies like microelectromechanical systems (MEMS).

Both destructive and nondestructive microscopy tools now complement—and surpass—the scanning electron microscope (SEM), an iconic fixture of the well-equipped laboratory. Newer-generation instruments, some reasonably priced and user-friendly, are moving once-outsourced imaging and analysis tasks back into the manufacturing environment.

Manufacturing organizations in the United States are under intense competitive pressure. Major changes are being experienced with respect to resources, markets, manufacturing processes, and product strategies. As a result of international competition, only the most productive and cost-effective industries will survive.

Today's sensors and control systems have explosively expanded beyond their traditional production base into far-ranging commercial ventures. They will play an important part in the survival of innovative industries. Their role in information assimilation, and control of operations to maintain an error-free production environment, will help enterprises stay effective on their competitive course.

1.2 Establishing an Automation Program

Manufacturers and vendors have learned the hard way that technology alone does not solve problems. A prime example is the gap between the information and the control worlds, which caused production planners to set their goals according to dubious assumptions concerning plant-floor activities, and plant supervisors then could not isolate production problems until well after they had arisen.

The problem of creating effective automation for an error-free production environment has drawn a long list of solutions. Some are as old as the term *computer-integrated manufacturing* (*CIM*) itself. However, in many cases, the problem turned out not to be technology, but the ability to integrate equipment, information, and people.

The debate over the value of computer-integrated manufacturing technology has been put to rest, although executives at every level in almost every industry are still questioning the cost of implementing CIM solutions. Recent economic belt tightening has forced industry to justify every capital expense, and CIM has drawn fire from budget-bound business people in all fields.

www.TechnicalBooksPDF.com

Too often, the implementations of CIM have created a compatibility nightmare in today's multivendor factory-floor environments. Too many end users have been forced to discard previous automation investments and/or spend huge sums on new equipment, hardware, software, and networks in order to effectively link together data from distinctly dissimilar sources. The expense of compatible equipment and the associated labor cost for elaborate networking are often prohibitive.

The claims of CIM open systems are often misleading. This is largely due to proprietary concerns, a limited-access database, and operating system compatibility restrictions. The systems fail to provide the transparent integration of process data and plant business information that makes CIM work.

In order to solve this problem, it is necessary to establish a clearly defined automation program. A common approach is to limit the problem description to a workable scope, eliminating those features not amenable to consideration. The problem is examined in terms of a simpler workable model. A solution can then be based on model predictions.

The danger associated with this strategy is obvious: If the simplified model is not a good approximation of the actual problem, the solution will be inappropriate and may even worsen the problem.

Robust automation programs can be a valuable asset in deciding how to solve production issues. Advances in sensor technology have provided the means to make rapid large-scale improvements in problem solving and have contributed in essential ways to today's manufacturing technology.

The infrastructure of an automation program must be closely linked with the use and implementation of sensors and control systems, within the framework of the organization. The problem becomes more difficult whenever it is extended to include the organizational setting. Organization theory is based on a fragmented and partially developed body of knowledge, and can provide only limited guidance in the formation of problem models. Managers commonly use their experience and instinct in dealing with complex production problems that include organizational aspects. As a result, creating a competitive manufacturing enterprise—one involving advanced automation technology utilizing sensors and control systems and organizational aspects—is a task that requires an understanding of both how to establish an automation program and how to integrate it with a dynamic organization.

In order to meet the goals of integrated sensory and control systems, an automated manufacturing system must be built from compatible and intelligent subsystems. Ideally, a manufacturing system should be computer-controlled and should communicate with controllers and materials-handling systems at higher levels of the hierarchy, as shown in Fig. 1.3.

www.TechnicalBooksPDF.com

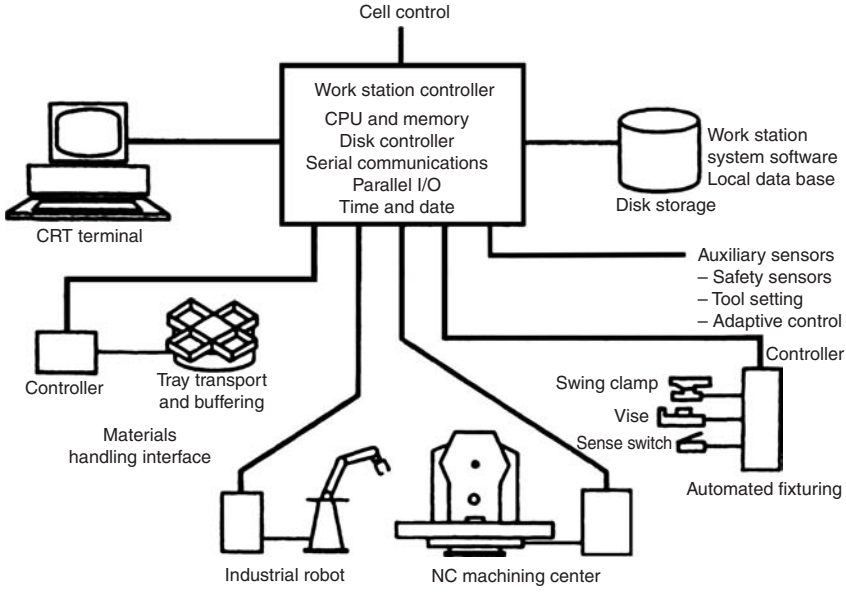


Figure 1.3 A computer-controlled manufacturing system.

1.3 Photo Sensing Fluorescence in Genome Sequencing

The sequencing of DNA molecules began in the 1970s with development of the Maxam-Gilbert method, and later the Sanger method. Originally developed by Frederick Sanger in 1975, most DNA sequencing that occurs in medical and research laboratories today is performed using sequencers employing variations of the Sanger method. Termed the chain-termination method, it involves a reaction where chain-terminator nucleotides are labeled with fluorescent dyes, combined with fragmented DNA, DNA sequencing primers, and DNA polymerase. Each nucleotide in the DNA sequence is labeled with a different dye color, and a chromatogram is produced, with each color representing a different letter in the DNA code—A, T, C, or G. Advances in sequencing technology and computer programming enabled relatively fast and cost-efficient DNA sequencing. However, sequencing of entire genomes of organisms was difficult and time consuming. At the time the Human Genome Project was officially started in 1990, it was thought that sequencing the human genome would take 15 years. The sequence was released in 2003, although some gaps still exist. The estimated project cost for the entire Human Genome Project was *around $3 billion*, although this figure represents a wide range of scientific activities that went into the project, not exclusively genome sequencing. Optimally, current sequencers are able to sequence approximately 2.8 million base pairs

www.TechnicalBooksPDF.com

per 24-hour period. However, even the smallest organisms such as bacteria are hundreds of millions of base-pairs long and the human genome is about 3 billion base pairs. At this rate, using the most modern Sanger sequencers, it will take almost *three years* to sequence the human genome.

1.4 A New Sensing Tool for Decoding the Genome

Whole genome sequencing (WGS) has been the standard approach for producing high-quality genome sequence data for more than 20 years. While other technologies exist, none has challenged the WGS method significantly. There is, however, a growing need for a more efficient and cost-effective approach for genome sequencing that will deliver the high-quality data of conventional sequencing at a low cost.

To that end, U.S. and Australian scientists have pioneered a new hybrid method for genomic sequencing that is faster and more cost-effective than the WGS method alone. The new approach, combines the best of new and old code-cracking methods for "fingerprinting" the genetic basis of life. Using the genomes of six ocean bacteria, the researchers tested the utility and cost-effectiveness of the new and old methods and found that the new hybrid method was better than either method on its own (see Fig. 1.4).

1.5 Mapping RNA Protein Folding Energy Through Bio-Sensors

All the crucial proteins in our bodies must fold into complex shapes to accomplish their predestined tasks. The snarled molecules grip other molecules to move them around, to speed up important chemical reactions, or to take hold of our genes, turning them "on" and "off" to affect which proteins our cells make.

In 2008, scientists discovered that RNA—the stringy molecule that translates human genetic code into protein—can act as a protein itself. RNA can form winding bundles that shut genes down or start them up without the help of proteins. Since the discovery of these

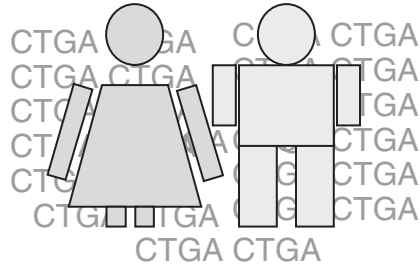


Figure 1.4 CTGA—The genome sequencing technique.

www.TechnicalBooksPDF.com

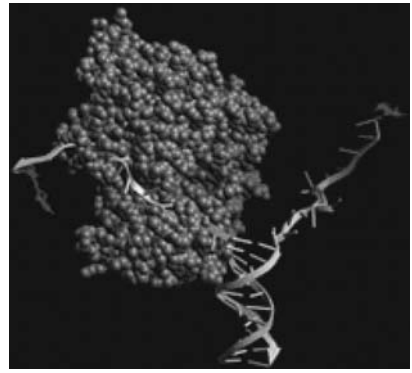


Figure 1.5
RNA forms winding bundles.

RNA clusters, called "riboswitches," in 2002, scientists have been eager to understand how they work and how they form. Now, researchers at Stanford University in California are deeply involved in understanding how the three-dimensional twists and turns in a riboswitch come together by taking hold of it and tugging it straight. By physically pulling on this winding RNA, scientists have determined how a three-dimensional molecular structure folds, step by step (Fig. 1.5.)

1.6 Nano-Sensor Manufacturing Challenges

Nanotechnology offers the scientific communities and end users valuable opportunities, but not without substantial obstacles and difficulties, in particularly in methodology and business management. The current nanotechnology industry is worth $40 billion, with applications in microelectronics, specialty coatings, tooling, and cosmetics. These and other emerging nanotechnology markets are forecast to grow over the next half decade to more than $1 trillion by 2015, according to analysts at the National Science Foundation (NSF) and independent research groups. In order to effectively commercialize nanotechnology and harness its unique benefits, it's widely believed that all sectors of industry will have to change their management approaches, business models, and corporate strategies. A recent study by the National Center for Manufacturing Sciences (NCMS) in Ann Arbor, Michigan was funded by the NSF. The NCMS outlined how well prepared U.S. industries must be to face these changes to take advantage of nanotechnology's values.

1.7 Nano-Crystals Enable Scalable Memory Technologies

The 24-Mb memory array of "silicon nano-crystals" is reducing the cost of embedded flash memory as researchers have demonstrated

www.TechnicalBooksPDF.com

the advancement in 2008. However, in 1984, Fujio Masuoka, Prof. of Solid State Electronics at Tohuku University in Sendai, Japan named this breakthrough "flash" since its erasures of data in the device is reminiscent of a camera flash. Today, flash has become one of the most widely used memory systems in devices such as USB drives, digital camera memory cards, and stand-alone chips in circuit boards. However, with a set of recent advancements, flash memory is poised to make an even greater impact on electronic devices.

The building block on which sensors and control technology is erected is the clear understanding of light energy, and the methodology of light generated and employed in making sensor devices. Therefore, a clear understanding of light-emitting diodes (LEDs), light amplification by stimulated emission of radiation (laser), the visible light of through beams, infrared light energy, and ultraviolet light energy will enhance the application of sensors and control systems in manufacturing applications.

1.8 Chromaticity—Color Rendering Index (CRI)

Chromaticity describes the lamp itself, or a neutral surface illuminated by a lamp. Chromaticity sets the "tone" or atmosphere of an environment: warm or cool. Chromaticity is often called color temperature and is usually measured in Kelvins. It can also be defined by using x and y coordinates against a standard chromaticity scale developed by the Commission Internationale de l'Éclairage (CIE).

An RGB color space is any additive color space based on the RGB color model. A particular RGB color space is defined by the three chromaticity of the red, green, and blue additive primaries, and can produce any chromaticity that is the triangle defined by those primary colors (Fig. 1.6). The complete specification of an RGB color space also requires a white point chromaticity and a gamma correction curve.

RGB is an acronym for Red, Green, Blue.

An RGB color space can be easily understood by thinking of it as "all possible colors" that can be made from three colorants for red, green and blue. Imagine, for example, shining three lights together onto a white wall: one red light, one green light, and one blue light, each with dimmer switches. If only the red light is on, the wall will look red. If only the green light is on, the wall will look green. If the red and green lights are on together, the wall will look yellow. Dim the red light some and the wall will become more of a yellow-green. Dim the green light instead, and the wall will become more orange. Bringing up the blue light a bit will cause the orange to become less saturated and more whitish. In all, each setting of the three dimmer switches will produce a different result, either in color or in brightness or both. The set of all possible results is the gamut defined by those particular color light bulbs. Swap out the red light bulb for one

www.TechnicalBooksPDF.com

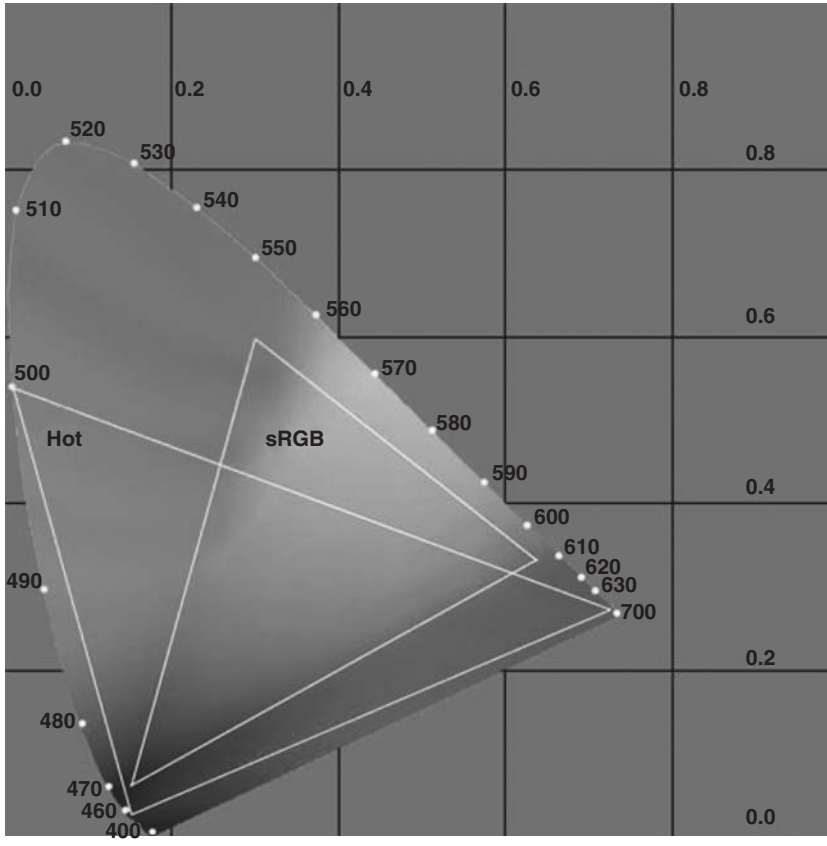


Figure 1.6 Chromaticity—color rendering index (CRI). (*See also color insert.*)

of a different brand that is slightly more orange, and there will be slightly different gamut, since the set of all colors that can be produced with the three lights will be changed (Fig. 1.7).

An LCD display can be thought of as a grid of thousands of little red, green, and blue light bulbs, each with their own dimmer switch. The gamut of the display will depend on the three colors used for the red, green and blue lights. A wide-gamut display will have very saturated, "pure" light colors, and thus be able to display very saturated, deep colors.

RGB is a convenient color model for computer graphics because the human visual system works in a way that is similar—though not quite identical—to an RGB color space. The most commonly used RGB color spaces are sRGB and Adobe RGB (which has a significantly larger gamut). Adobe has recently developed another color space called Adobe Wide Gamut RGB, which is even larger, in detriment to gamut density.

www.TechnicalBooksPDF.com

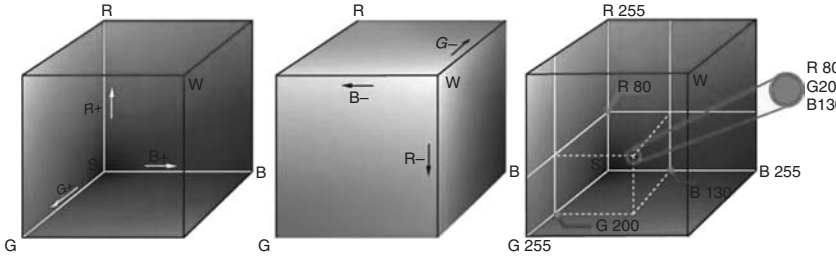


Figure 1.7 Chromaticity—RGB point in space.

As of 2007, sRGB is by far the most commonly used RGB color space, particularly in consumer grade digital cameras, HD video cameras, computer monitors and HDTVs, because it is considered adequate for most consumer applications. Having all devices use the same color space is convenient in that an image does not need to be converted from one color space to another before being displayed. However, sRGB's limited gamut leaves out many highly saturated colors that can be produced by printers or in film, and thus is not ideal for some high quality applications. The wider gamut Adobe RGB is being built into more medium-grade digital cameras, and is favored by many professional graphic artists for its larger gamut.

RGB spaces are generally specified by defining three primary colors and a white point. In the Table 1.1, the three primary colors and white points for various RGB spaces are given. The primary colors are specified in terms of their CIE 1931 color space chromaticity coordinates (x,y).

The CIE 1931 color space standard defines both the CIE RGB space, which is an RGB color space with monochromatic primaries, and the CIE XYZ color space, which works like an RGB color space except that it has nonphysical primaries that can not be said to be red, green, and blue.

A chromaticity chart is an attempt to represent colors from the 3D color space of human visual perception on a 2D graph. Obviously, one of the dimensions must be imagined. We give up on color "intensity" and use the chart as an indication of hue and saturation. The axes of the chart can be thought of as two of the dimensions of the human visual system, denoted as x and y. The horseshoe-like curve is the locus of color perceptions that occur when we look at pure spectral wavelengths of light, such as from a laser.

The point at the far right corresponds to light having a wavelength of 700 nm (red). As the wavelength shortens, the perceptual response in (x, y) moves up the curve, reaches a peak at 520 nm (green), then continues to the point at the bottom for 400-nm light (blue).

The straight line connecting the 400-nm locus to the 700-nm point is called the "line of purples." No spectral wavelength can stimulate

www.TechnicalBooksPDF.com

			xR	yR	xG	yG	xB	yB
ISO RGB	Limited floating		floating					
Extended ISO RGB	Unlimited (signed)	floating	floating					
sRGB, HDTV (ITU-R BT.709-5)	CRT	D65	0.64	0.33	0.30	0.60	0.15	0.06
scRGB	Unlimited (signed)	D65	0.64	0.33	0.30	0.60	0.15	0.06
ROMM RGB	Wide	D50	0.7347	0.2653	0.1596	0.8404	0.0366	0.0001
Adobe RGB 98	CRT	D65	0.64	0.34	0.21	0.71	0.15	0.06
Apple RGB	CRT	D65	0.625	0.34	0.28	0.595	0.155	0.07
NTSC (FCC 1953, ITU-R BT.470-2 System M)	CRT	C	0.67	0.33	0.21	0.71	0.14	0.08
NTSC (1979) (SMPTE 170M/240M, SMPTE RP 145 “SMPTE C”)	CRT	D65	0.63	0.34	0.31	0.595	0.155	0.07
PAL/SECAM (EBU 3213, ITU-R BT.470-2 System B, G)	CRT	D65	0.64	0.33	0.29	0.60	0.15	0.06
Adobe Wide Gamut RGB	Wide	D50	0.735	0.265	0.115	0.826	0.157	0.018
CIE (1931)	Wide	E	0.7347	0.2653	0.2738	0.7174	0.1666	0.0089

From “Susstrunk, Buckley and Swen—2005.”

Table 1.1 Some RGB Volor Space Parameters

www.TechnicalBooksPDF.com

these colors; rather, it is light that lacks wavelengths between red and blue that reside here.

All of the colors humans can perceive fall *inside* the spectral locus. It is an interesting feature of this chart that given two source colors, all of the colors that can be made by blending those colors in different amounts will fall on the line that connects them. An important extension of this is that the colors that can be made by blending *three* sources will fall inside the triangle defined by those sources. The vertices of the triangle are regarded as the *primaries* of that particular color system. This is very useful in predicting the colors that can be made by three different phosphors, as used in video displays. The colors inside the triangle represent the *color gamut* of the display, the colors that can be generated by the display.

The exact primary colors for a given display technology are carefully selected to balance a set of tradeoffs between saturation, hue, and brightness. Some very successful phosphor combinations have been found over the years and are used in various broadcast television and video standards. One successful set comprises the Trinitron phosphors used in a vast number of television and computer displays. It embodies a design choice where the extent of the color gamut is diminished slightly in favor of a significantly brighter image. Because of its ubiquity, it forms the foundation for the sRGB color space, a standard used in PC and world-wide-web graphic design.

The spectral locus represents the highest degree of purity possible for a color. As one moves away from this boundary toward the interior, colors become less saturated. In the center area, the colors become nearly neutral tints of gray. A display system, defined by a triangle of primaries, will select a point in the center to be the *white-point* for the display. It need not be the geometric center of the triangle, and for many systems it can be quite arbitrary, since the visual system will adapt, attempting to make an image look "natural." The exact mechanism of adaptation is complex and the subject of current color research, but enough is known that it is now possible to translate images from one display system to another, or to a hardcopy print, and retain the natural appearance. The circle in the RGB space cube representation shown in Fig. 1.7 represent the white-points for the two different display systems charted.

A set of primaries and a white-point are enough to define a (linear) color space. These are usually called RGB color spaces because the primaries for most useful systems are distinctly red, green, and blue. A characteristic of these color spaces is that they can be implemented using linear algebra. Converting from one space to another is a matter of applying the correct 3×3 matrix operation.

There is one more characteristic of most displays that modifies the color space and destroys its linearity: the gamma.

www.TechnicalBooksPDF.com

1.9 The LED Color Chart

The LED color chart in Table 1.2 does NOT represent what the LED light carries. This chart is only to be used as a reference for the various types of LEDs being manufactured today, and to show what their basic properties are.

1.10 The Color Rendering Index (CRI)

The color rendering index (CRI) is a measure of the quality of light. A measurement of the amount of color shift that objects undergo when lighted by a light source as compared with the color of those same objects when seen under a reference light source of comparable color temperature. LED light CRI values generally range from 60 (average) to 90 (best). High CRI equates to sharper, crisper, more natural-colored pictures, while at the same time reducing glare (Fig. 1.8).

A measure of the color of a light source relative to a blackbody at a particular temperature expressed in degrees Kelvin (K). Incandescent lights have a low color temperature (approximately 2800K) and have a red-yellowish tone; daylight has a high color temperature (approximately 6000K) and appears bluish (the most popular fluorescent light, Cool White, is rated at 4100K). Lamps with color temperatures below 5000K tend to be more yellow/red, lamps rated between 5000 and 6000K are viewed as white, while lamps above 6000K tend to have a blue cast.

1.10.1 The Foot-Candle

The foot-candle unit is defined as the amount of illumination the inside surface of an imaginary 1-foot radius sphere would be receiving if there was a uniform point source of one candela in the exact center of the sphere. Basically, the amount of light that a single candle would provide to a 1-foot radius sphere.

1.10.2 Full Spectrum

A light bulb or lamp that produces a light spectrum that covers the entire range of visible light (400–700 nm) without gaps in its spectral output. White LEDs are inherently a full spectrum light source.

1.10.3 Intensity

Is a measure of the time-averaged energy flux or amount of light striking a given area. For bulbs alone, this is measured in terms of lumens, while for lighting fixtures it is measured in lux (lumens/sq. meter).

www.TechnicalBooksPDF.com

Wavelength (nm)	Color Name	Fwd Voltage (V_f @ 20 mA)	Intensity 5 mm LEDs	Viewing Angle	LED Dye Material
635	High Eff. Red	2.0	200 mcd @20 mA	15°	GaAsP/GaP—Gallium Arsenic Phosphide/ Gallium Phosphide
633	Super Red	2.2	3500 mcd @20 mA	15°	InGaAlP—Indium Gallium Aluminum Phosphide
623	Red-Orange	2.2	4500 mcd @20 mA	15°	InGaAlP—Indium Gallium Aluminum Phosphide
612	Orange	2.1	160 mcd @20 mA	15°	GaAsP/GaP—Gallium Arsenic Phosphide/ Gallium Phosphide
592	Amber Yellow	2.1	7000 mcd @20 mA	15°	InGaAlP—Indium Gallium Aluminum Phosphide
585	Yellow	2.1	100 mcd @20 mA	15°	GaAsP/GaP—Gallium Arsenic Phosphide/ Gallium Phosphide
3500K	“Incan-descent” White	3.6	2000 mcd @20 mA	20°	SiC/GaN—Silicon Carbide/Gallium Nitride
5000K	Pale White	3.6	4000 mcd @20 mA	20°	SiC/GaN—Silicon Carbide/Gallium Nitride
6500+K	Cool White	3.6	6000 mcd @20 mA	20°	SiC/GaN—Silicon Carbide/Gallium Nitride
574	Super Lime Yellow	2.4	1000 mcd @20 mA	15°	InGaAlP—Indium Gallium Aluminum Phosphide

www.TechnicalBooksPDF.com

570	Super Lime Green	2.0	1000 mcd @20 mA	15°	InGaAlP—Indium Gallium Aluminum Phosphide
565	High Efficiency Green	2.1	200 mcd @20 mA	15°	GaP/GaP—Gallium Phosphide/Gallium Phosphide
560	Super Pure Green	2.1	350 mcd @20 mA	15°	InGaAlP—Indium Gallium Aluminum Phosphide
555	Pure Green	2.1	80 mcd @20 mA	15°	GaP/GaP—Gallium Phosphide/Gallium Phosphide
525	Aqua Green	3.5	10,000 mcd @20 mA	15°	SiC/GaN—Silicon Carbide/Gallium Nitride
505	Blue Green	3.5	2000 mcd @20 mA	45°	SiC/GaN—Silicon Carbide/Gallium Nitride
470	Super Blue	3.6	3000 mcd @20 mA	15°	SiC/GaN—Silicon Carbide/Gallium Nitride
430	Ultra Blue	3.8	100 mcd @20 mA	15°	SiC/GaN—Silicon Carbide/Gallium Nitride

TABLE 1.2 The LED Color Chart (*See also color insert.*)

www.TechnicalBooksPDF.com

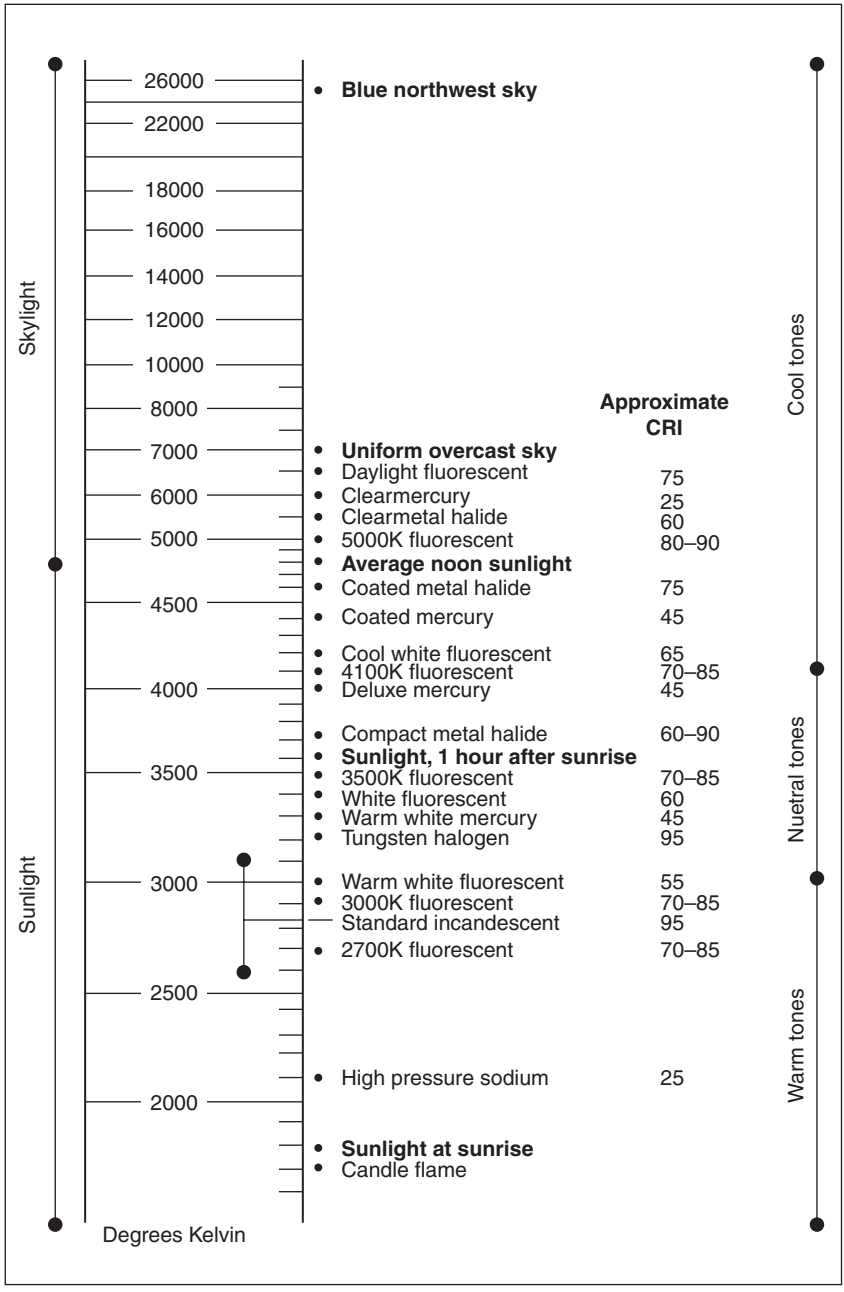


Figure 1.8 The color rendering index (CRI).

www.TechnicalBooksPDF.com

1.10.4 Kelvin Color Temperature

The Kelvin color temperature is a measure of the color of a light source relative to a blackbody at a particular temperature expressed in degrees Kelvin (K). Incandescent lights have a low color temperature (approximately 2800K) and have a red-yellowish tone; daylight has a high color temperature (approximately 6000K) and appears bluish (the most popular fluorescent light, Cool White, is rated at 4100K). Today, the phosphors used in fluorescent lamps can be blended to provide any desired color temperature in the range from 2800K to 6000K. Lamps with color temperatures below 5000K tend to be more yellow/red, lamps rated between 5000 and 6000K are viewed as white, while lamps above 6000K tend to have a blue cast.

1.11 LEDs—Light-Emitting Diodes

An LED is a solid-state device and does not require the heating of a filament to create light. Rather, electricity is passed through a chemical compound that is excited and thus generates light. LEDs are not bulbs or lamps. LEDs require precise manufacturing operation to make them ready to be used. They need to be placed on a circuit board or other material that will allow electricity to pass through it at a specific voltage and current, and with components required to operate them at specific voltages such as 12 V DC, 24 V DC, or 120 V AC. They are not readily available to be plugged into a 12-volt or 120-volt power source.

1.11.1 LED Bars

An LED bar is a solid strip of material to which LEDs have been soldered, along with resistors and other components that a specific product requires to make them operate at the stated operating voltage. The bars are usually an enclosed strip of LEDs. Enclosures are plastics, or aluminum, or metal composites with various types of lens/cover plates.

1.11.2 LED Clusters or Arrays

A group of LEDs set in a square, rectangular, or linear pattern, and formatted to be operated at a specific voltage. They will always include two wires called leads. One is positive, the other negative.

1.11.3 LED Drivers

LED drivers are current control devices that replace the need for resistors. LED drivers respond to the changing input voltage while maintaining a constant amount of current (output power) to the LED as its electrical properties change with temperature.

www.TechnicalBooksPDF.com

1.11.4 LED Lighting

LED lighting is a general term used by those who do not know the specific type or category of LED lighting they are after. LED lighting includes LED bulbs and fixtures, flashlights, strips, clusters, and other LED light sources.

1.11.5 LED Strips

LED strips are usually printed circuit boards with LEDs soldered to the board. The strip can be rigid, or flexible and without any enclosure to protect the LED and circuit.

1.11.6 Low Voltage

LEDs require 12 V DC, 24 V DC, or 48 V DC to be powered, as opposed to 110/120 V AC, which is high voltage. Usually, LEDs utilize a low voltage of 12 V DC; however, sometimes 24 V DC is used. To run these low voltage lights, power must be sent to the light through a power supply, transformer, or adapter that is connected to 110/120/240 V AC power lines. The actual voltage reaching the light will be at 12 V DC.

1.11.7 Lumen Maintenance

Lumen maintenance is defined as how well an LED light bulb is able to retain its intensity when compared to a new one. Typically, a high-power SMD LED bulb will retain 70 percent of its intensity for 40,000–50,000 hours. This means a good-quality LED bulb will run eight hours a day for 13 years at 70 percent of its new condition. It is interesting to state that no other light source can achieve such results.

1.11.8 Lumens

The lumens is the unit of luminous flux use by the International System and is equal to the amount of light given out through a solid angle by a source of one candela intensity radiating equally in all directions. It is used to measure light bulbs as stand-alone light sources. Lighting fixtures are measured by lux output, which is lumens per square meter.

1.11.9 Lux

The lux is typically used to measure the light intensity produced by a lighting fixture. The higher the lux reading, the more light the lighting fixture is producing for a given area, known as lumens per square meter.

1.11.10 LED mA

All LEDs run on current, and current is measured in milliamps (mA). All LED products have a mA rating at which they are powered.

www.TechnicalBooksPDF.com

1.11.11 Max Rated Temperature

Max rated temperature is often called the "operating temperature." It is the ambient temperature at which the LED light source is installed and at which it should be maintained. In most cases, this is around 40–50°C, which is comparable to 104 to 122°F. Operating an LED light source beyond the operating temperature will lower the LED's life span, leading to its destruction.

1.11.12 mcd

mcd stands for millicandela. It is used to rank the brightness of an LED. 1000 mcd is equal to one candela. The higher the mcd number, the brighter the light the LED emits. The human eye may sense visible light, as described in Fig. 1.9. The visible light spectra varies between ultraviolet 350 to red 700 nm (Fig. 1.10). Light between 315 and 380 nm is often called long wave or "black-light."

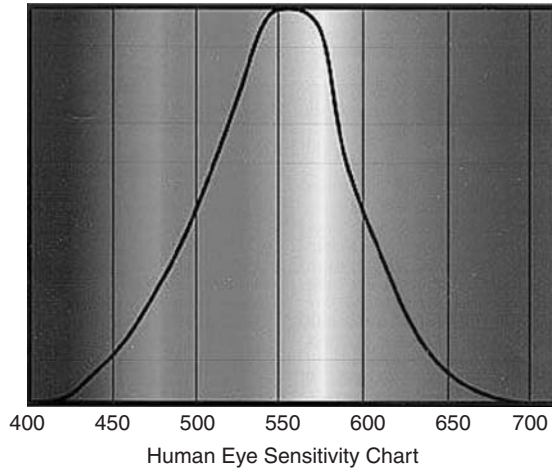


Figure 1.9 Visible light spectra. (*See also color insert.*)

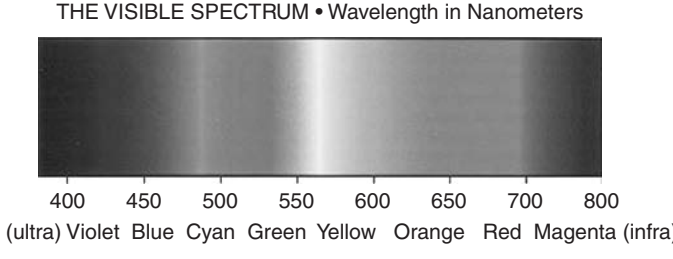


Figure 1.10 Visible ultraviolet light to red light spectra. (*See also color insert.*)

www.TechnicalBooksPDF.com

1.11.13 Nanometers (nm)

Nanometers (nm) are used to measure the wavelengths of light. The lower the wavelength (e.g., 400 nm), the bluer and stronger the light source. Longer wavelengths above 600 nm are red. Above 680 nm, they fall into the infrared range, which is colorless to our eyes. White LEDs have no specific wavelength. They are measured by the color of white against the chromaticity scale.

1.11.14 Operating Life

Operating life is usually referred to as the number of hours a specific type of LED is expected to be operational. With high-powered LEDs, that usually means life after it loses 10–15 percent or more rated output after 1000 or more hours of runtime. The Luxeon LEDs are rated for 50,000 at 75 percent maintenance for 50,000 hours.

1.11.15 Printed Circuit Boards

Printed circuit boards (PCBs) are made from various materials, including fiberglass and aluminum. The PCB has an electrical circuit imprinted in silver etching. The PCB is the platform by which LEDs are employed in various applications. It can be a rigid board, flexible, or twistable.

1.11.16 Power Supply

The power supply transforms voltage and current, causing electrical conversion of 110/120/240 V AC line power into 12 V DC or any other values that will then be applied directly to the LED light product. Power supplies are rated according to the current/amperage load capacity each can handle. It is an electrical or electromechanical device sometimes referred to as a LED driver.

1.11.17 PWM

The Pulse Width Modulation (PWM) with respect to LEDs is the LED pulsed or strobed at a fast rate that the eye will see the light as being constant. This pulsing or turning on and off of the LED lowers the potential heat-stress on the chemical that makes the light, thus allowing the LED to often have a longer life than anticipated. This is one of the reasons why a dimmer/PWM is recommended with the use of a 12 V DC LED product.

1.11.18 RGB

RGB stands for red, green, and blue, the three primary colors that make up white light and all other colors. It can be preprogrammed to

www.TechnicalBooksPDF.com

seven colors, automatically changing the LED bar or strip that is non-adjustable. Also, it means an RGB color changing system, which allows adjustment of color change frequency, strobing, chasing, and other action modes.

1.11.19 SMD/SMT

SMD/SMT is a type of low-profile LED that is surface-mounted to a PCB. These types of LEDs are very powerful and range in lumen output from 35 up to 170 lumens. With the latest LED technology being applied today, these have shown to have the most robust system, delivering light levels and color consistency. These SMD LEDs are usually in the 0.5-watt, 1-watt, 3-watt, and 5-watt power range. In this instance, a 7-watt or 9-watt LED light will contain 1-watt LEDs × 7, or 1-watt LEDs × 9, or 3-watt LEDs × 3.

1.11.20 SSL

SSL means "solid-state lighting." It does not use heating of a thin fragile filament to create light. Instead, it uses electrical current passing through a chemical, which will then become excited and emit light.

1.12 Task Lighting-Lamps

A task lighting-lamp is an LED light used to specifically light a particular area used for work or reading. It is typically found in the form of a desk, floor, or clamp-on lamp, and can be a high-powered LED light in any form.

1.12.1 UV-A

UV-A is in the range of 315 to 380 nm and is also called "long wave" or "black-light" since it is invisible to the human eye. It can cause skin irritation and fading of fabrics.

1.12.2 UV-B

UV-B is in the range of 280 to 315 nm and is also called "medium wave radiation." It can cause severe damage to the skin and human eye through exposure.

1.12.3 UV-C

UV-C is below 280 and is also called "short wave" or "germicidal" for its ability to destroy even bacterial life forms. It is extremely hazardous to all life forms due to its ability to cause immediate damage to cellular DNA.

www.TechnicalBooksPDF.com

1.12.4 View Angle Degree

The view angle degree is referred to as directivity, or the directional pattern of an LED light beam. The expressed degree dictates the width of the light beam. Also, it controls to some extent the light intensity of an LED. View angles range from 8 to 160 degrees and are provided through the use of optics, which are special lenses made to collimate light into a desired view angle.

1.12.5 Voltage

Voltage is the rate at which energy is drawn from a source, thus producing a flow of electricity, or amperage, in a circuit. The difference in electrical charge between two points in a circuit is expressed as volts.

1.12.6 Voltage Regulator

A voltage regulator is a device that limits or controls and stabilizes the voltage being applied to usage units such as LED lights and motors. A regulator also takes higher voltages than required and reduces them to the working voltage that makes a specific product run correctly. In many instances, the lack of a voltage regulator will allow higher voltage in than a product can manage and as a result will cause irreparable damage to the product.

1.12.7 Volts

A volt is the International System unit of electric potential and electromotive force, and is equal to the difference of electric potential between two points on a conducting wire carrying a constant current of one ampere when the power dissipated between the points is one watt.

1.12.8 Waterproof

Waterproof means that the LED product can be submerged into calm water. However, there is a limited depth as stated for each specific product, and this must be specifically outlined regarding "submersion." It also means that the product is made to withstand water being splashed onto it like rain, or having snow fall on it. Waterproof products are made of high-grade stainless steel, aluminum, or high-strength plastics.

1.12.9 Watts

A watt is the unit for measuring electrical power. It defines the rate of energy consumption by an electrical device when it is in operation. The energy cost of operating an electrical device is calculated as its wattage times the hours of use. In single-phase circuits,

www.TechnicalBooksPDF.com

it is related to volts and amps by the formula: volts × amps × power factor = watts.

1.12.10 Watts per LED

It can be confusing when two watt numbers are used in product specifications. For the application to SMD high-powered LEDs, the 1 watt, 3 watt, and 5 watt classifications (and so on), refers to the power consumption of that specific LED installed in that product. The watt numbers expressed as light output are a comparison to an incandescent light bulb light output (e.g., a 60-watt light output is equal to a 60-watt incandescent light bulb). The watt output is equipment measured.

1.12.11 Weatherproof

Weatherproof means the product can endure splashing water and high humidity without deterioration to the LED or circuit. Weatherproof LED products cannot be submerged into water.

1.12.12 White

White is defined by Kelvin temperature or degrees Kelvin. A Kelvin temperature of 6000 K is white with a blue tint. 5000–5500 K is a daylight/sunlight white. At 4200–4500 K, it is called cool white. At 2800–3300 K, it is warm white, which is the color temperature that the majority of incandescent light bulbs emit. The color becomes "warmer" from 5500 K down the scale, due to the dominance of red and yellow hues. In the opposite direction, whites will have cooler colors, with blues and greens becoming more apparent, thus they are called cool whites.

1.13 The Basics on LEDs

How does an LED work?

The following outlines some basic principles of an LED.

- White LEDs need 3.6 V DC and use approximately 30 milliamps of current; a power dissipation of 100 milliwatts (mW).
- The positive power is applied to one side of an LED semiconductor through a lead often called an anode, and another lead called a whisker.
- The other side of the semiconductor is attached to the top of the anvil. This is the negative power lead, which is called the cathode.
- It is the chemical makeup of the LED semiconductor that determines the color of the light the LED produces.

www.TechnicalBooksPDF.com

- The epoxy resin enclosure has three functions. It is designed to allow the most light to escape from the semiconductor, it focuses the light and protects the LED semiconductor from the elements.
- The entire LED unit is totally embedded in an epoxy. This is what make LEDs virtually indestructible. There are no loose or moving parts within the solid epoxy enclosure.

Thus, a light-emitting diode (LED) is essentially a PN junction semiconductor diode that emits light when current is applied. It is a solid-state device that controls current without heated filaments and is therefore highly reliable. LED performance is based on a few primary characteristics:

1.13.1 Monochromatic LEDs

LEDs are highly monochromatic, emitting a pure color in a narrow frequency range. The color emitted from an LED is identified by a peak wavelength (L_{pk}) and measured in nanometers (nm). See Fig. 1.11.

Peak wavelength is a function of the LED chip material. Although process variations are ±10 nm, the 565 to 600 nm wavelength spectral region is where the sensitivity level of the human eye is highest. Therefore, it is easier to perceive color variations in yellow and amber LEDs than other colors.

LEDs are made from gallium-based crystals that contain one or more additional materials (such as phosphorous) to produce a distinct color. Different LED chip technologies emit light in specific regions of the visible light spectrum and produce different intensity levels. See Table 1.3.

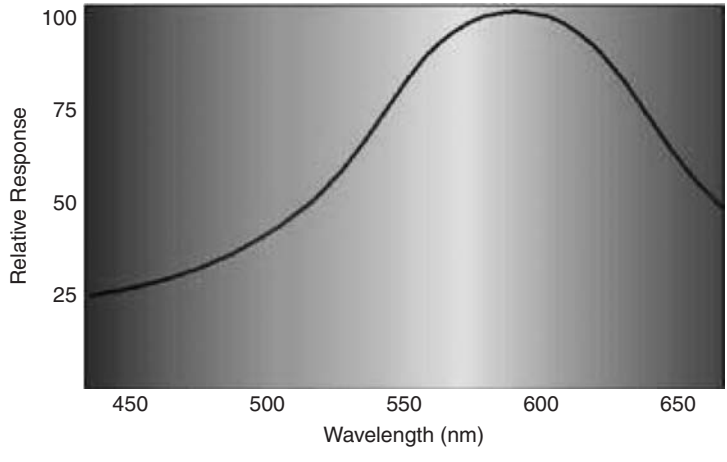


Figure 1.11 An LED peak wavelength (L_{pk}). (*See also color insert.*)

www.TechnicalBooksPDF.com

LED Color	Standard Brightness				High Brightness			
	Chip Material	Ipk (nm)	Iv (mcd)	Viewing Angle	Chip Material	Ipk (nm)	Iv[3] (mcd)	Viewing Angle
Red	GaAsP/ GaP	635	120	35	AS AlInGaP	635	900	30
Orange	GaAsP/ GaP	605	90	30	AS AlInGaP	609	1,300	30
Amber	GaAsP/ GaP	583	100	35	AS AlInGaP	592	1,300	30
Yellow	GaP	570	160	30	—	—	—	—
Green	GaP	565	140	24	GaN	520	1,200	45
Turquoise	—	—	—	—	GaN	495	2,000	30
Blue	—	—	—	—	GaN	465	325	45

Table 1.3 Comparison of Chip Technologies for Wide-Angle Non-Diffused LEDs (*See also color insert.*)

1.13.2 LED White Light

When light from all parts of the visible spectrum overlap one another (Fig. 1.12), the additive mixture of colors appears white. However, the eye does not require a mixture of all the colors of the spectrum to perceive white light. Primary colors from the upper, middle, and lower parts of the spectrum (red, green, and blue), when combined, appear white. To achieve this combination with LEDs requires a sophisticated electrooptical design to control the blend and diffusion of colors. Variations in LED color and intensity further complicate this process.

Presently, it is possible to produce white light with a single LED using a phosphor layer (Yttrium Aluminum Garnet) on the surface of a blue (Gallium Nitride) chip. Although this technology produces various hues, white LEDs may be appropriate to illuminate opaque lenses or backlight legends. However, using colored LEDs to illuminate similarly colored lenses produces better visibility and overall appearance.

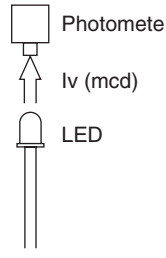


Figure 1.12 The structure of an LED to produce a mixture of light: white light.

www.TechnicalBooksPDF.com

1.13.3 LED Intensity

LED light output varies with the type of chip, encapsulation, efficiency of individual wafer lots, and other variables. Several LED manufacturers use terms such as "superbright" and "ultrabright" to describe LED intensity. Such terminology is entirely subjective since there is no industry standard for LED brightness.

The amount of light emitted from an LED is quantified by a single-point on-axis luminous intensity value (Iv). LED intensity is specified in terms of millicandela (mcd). This on-axis measurement is not comparable to mean spherical candlepower (MSCP) values used to quantify the light produced by incandescent lamps.

Luminous intensity is roughly proportional to the amount of current (If) supplied to the LED. The greater the current, the higher the intensity. Of course, design limits exist. Generally, LEDs are designed to operate at 20 milliamps (mA). However, the operating current must be reduced relative to the amount of heat in the application. For example, six-chip LEDs produce more heat than single-chip LEDs because they incorporate multiple wire bonds and junction points that are affected more by thermal stress than single-chip LEDs. Similarly, LEDs designed to operate at higher design voltages are subject to greater heat. LEDs are designed to provide lifelong operation because of optimal design currents, considering heat dissipation and other degradation factors.

1.13.4 Visibility

Luminous intensity (Iv) does not represent the total light output from an LED. Both the luminous intensity and the spatial radiation pattern (viewing angle) must be taken into account, Figure 1.13. If two LEDs have the same luminous intensity value, the lamp with the larger viewing angle will have the higher total light output.

Theta one-half (q½) is the off-axis angle where the LED's luminous intensity is half the intensity at direct on-axis view. Two times q½ is the LED's full viewing angle; however, light emission is visible beyond the q½ point. Viewing angles listed in this catalog are identified by their full viewing angle (2q½°).

The LED viewing angle is a function of the LED chip type and the epoxy lens that distributes the light. The highest luminous intensity (mcd rating) does not equate to the highest visibility. The light output from an LED chip is very directional. A higher light output is achieved by concentrating the light in a tight beam. Generally, the higher the mcd rating, the narrower the viewing angle.

The shape of the encapsulation acts as a lens magnifying the light from the LED chip. Additionally, the tint of the encapsulation affects the LED's visibility. If the encapsulation is diffused, the light emitted by the chip is more dispersed throughout the encapsulation. If the encapsulation is nondiffused or water clear, the light is more intense

www.TechnicalBooksPDF.com

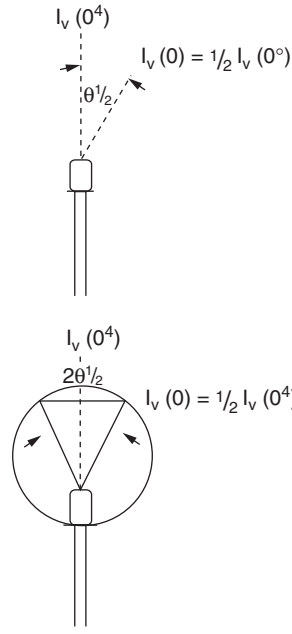


Figure 1.13 LED actual radiation pattern.

but has a narrower viewing angle. Nondiffused and water clear LEDs have identical viewing angles; the only difference is, water clear encapsulations do not have a tint to indicate color when the LED is not illuminated.

Overall visibility can be enhanced by increasing the number of LED chips in the encapsulation, increasing the number of individual LEDs, and utilizing secondary optics to distribute light. To illustrate, consider similar red GaAlAs LED chip technologies in four different configurations.

In each case, the amount of visible light depends on how the LED is being viewed. The single chip may be appropriate for direct viewing in competition with high ambient light. The six-chip LED may be better suited to backlight a switch or small legend, while the cluster or lensed LED may be best to illuminate a pilot light or larger lens.

1.13.5 Operating Life

Because LEDs are solid-state devices, they are not subject to catastrophic failure when operated within design parameters. DDP LEDs are designed to operate upwards of 50,000 hours at 25°C ambient temperature. Operating life is characterized by the degradation of LED intensity over time. When the LED degrades to half of its

www.TechnicalBooksPDF.com

original intensity after 50,000 hours, it is at the end of its useful life, although the LED will continue to operate as output diminishes. Unlike standard incandescent bulbs, DDP® LEDs resist shock and vibration and can be cycled on and off without excessive degradation.

1.13.6 Voltage/Design Currents

LEDs are voltage-driven—in other words, they are current-driven devices. Although drive current and light output are directly related, exceeding the maximum current rating will produce excessive heat within the LED chip due to excessive power dissipation. The result will be reduced light output and reduced operating life.

LEDs that are designed to operate at a specific voltage contain a built-in current-limiting resistor. Additional circuitry may include a protection diode for AC operation or a full-bridge rectifier for bipolar operation. The operating current for a particular voltage is designed to maintain LED reliability over its operating life.

1.13.7 Heat Dissipation

When many LEDs are mounted into a small area, heat generation must be taken into consideration. If there is a possibility that the ambient temperature may exceed 60 degrees centigrade, some kind of forced cooling system will be needed.

The ambient operating temperature must be taken into consideration when a product/system is being designed. Certain limits exist to maximum current at certain temperatures, which must be kept in mind.

1.13.8 EMI Countermeasures

When instantaneous power failure or a current surge by lightning stops the controller at abnormal conditions, the abnormally high electric current may continue running through the LEDs for an extended period of time. This can damage the LEDs in the system. Circuit protection against abnormally high current must be built into the system to protect against this.

1.14 Non-Phosphor White LEDs at a Viewing Angle of 30°

A non-phosphor white LED is shown in Fig. 1.14.

Description:

- Superbright LED Lamp
- Round type

www.TechnicalBooksPDF.com

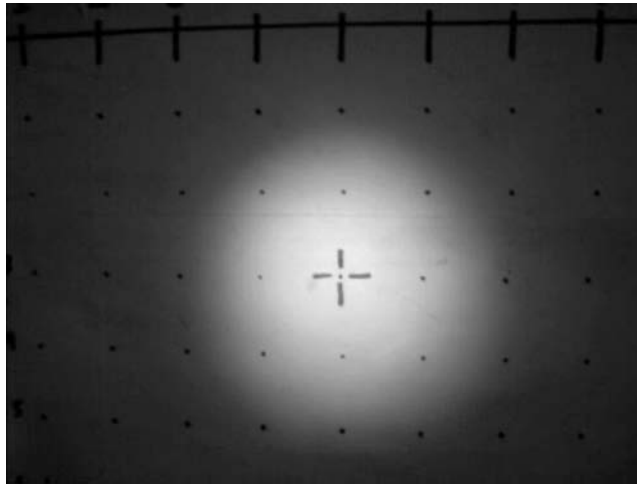


Figure 1.14 A non-phosphor white LED at 19.28 mA.

- T1-3/4 (5mm) diameter
- Lens color: water clear
- With flange
- Solder leads without stand-off

Features:

- Emitted color: white
- High luminous intensity
- Technology: InGaN/Sapphire

1.14.1 Non-Phosphor White LEDs

A non-phosphor white LED is a 5-mm round LED made of a clear epoxy case. This type is unlike most other white LEDs because it does not use a phosphor but emits white light directly. The operating chemistry is like ZnSe (zinc selenide)—there is one peak at ~460 nm and another stronger peak at ~560 nm. This is a normal trend for a non-phosphor white LED.

The beam consists of a small hotspot surrounded by a larger, mainly circular corona (Fig. 1.14). Unlike the other ZnSe, a white LED *does not* have a yellow color—dark amber to orange—or a brown perimeter in its beam. The beam of this LED mostly lacks that particular feature. It provides only a very narrow ring at the perimeter that may have a yellow tint but that is rather faint.

Two peaks are spaced relatively close to one another, Fig. 1.15. These LED emissions have no phosphor, Figs. 1.16 to 1.19.

www.TechnicalBooksPDF.com

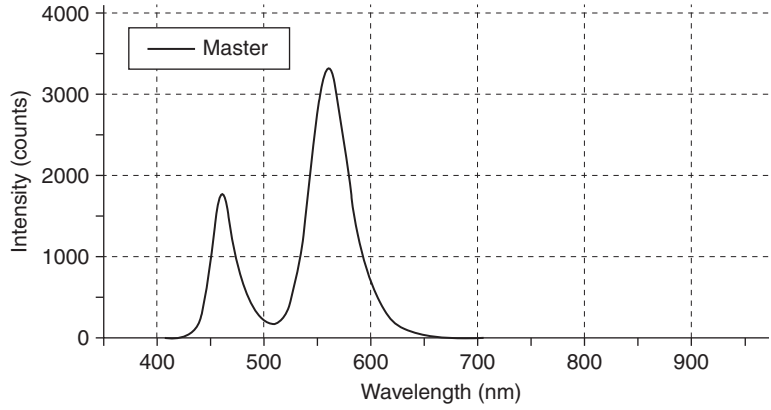


Figure 1.15 Spectrographic analysis of a non-phosphor white LED at 19.28 mA.

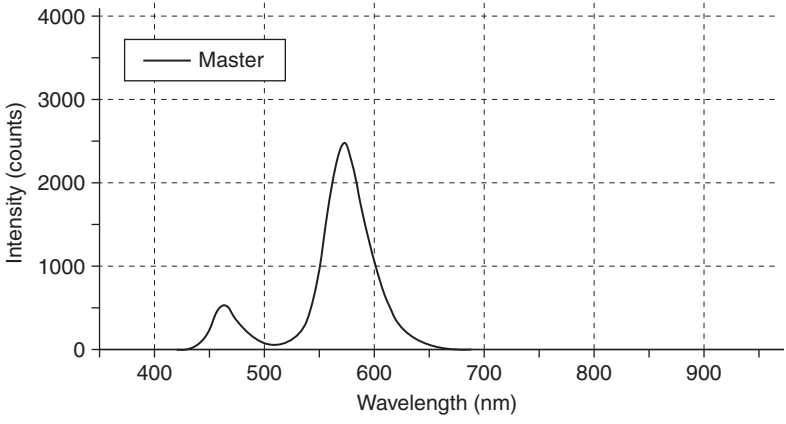


Figure 1.16 Spectrographic analysis of a non-phosphor white LED at ~1.8 mA.

1.15 Luminous Intensity (Candlepower)

The luminous intensity, or candlepower, is the light density within a very small solid angle, in a specified direction. In other words, this is the total number of lumens from a surface emitted in a given direction. The unit of measure is candela. In modern standards, the *candela* is the basic of all measurements of light, and all other units are derived from it. Candlepower measurements are often taken at various angles around the source, and the results are then plotted to give a candlepower distribution curve. Such a curve shows *luminous intensity* (how "bright" the source seems) in any direction.

www.TechnicalBooksPDF.com

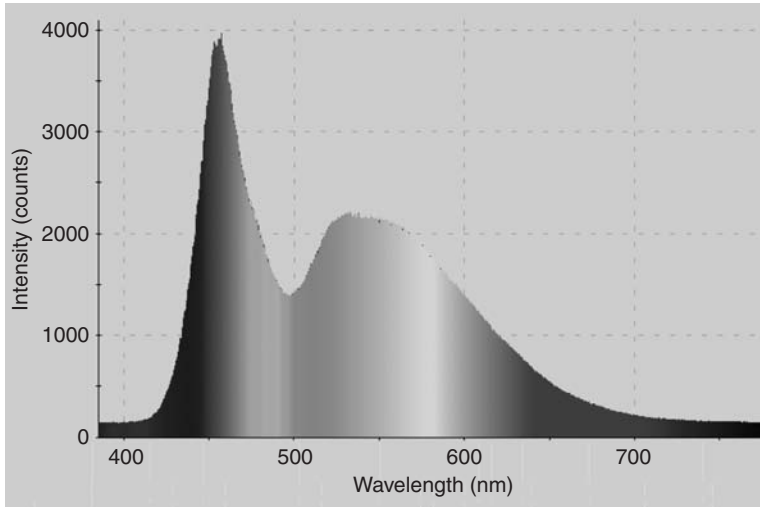


Figure 1.17 Spectrographic analysis of a 5-mm non-phosphor white LED. (*See also color insert.*)

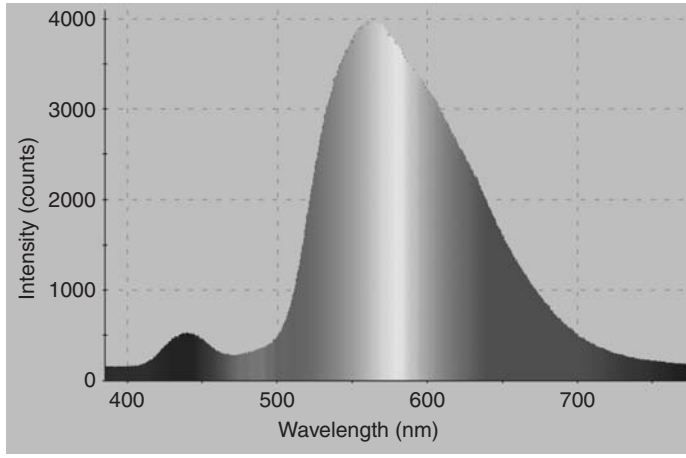


Figure 1.18 Spectrographic analysis of a 10-mm non-phosphor warm white LED. (*See also color insert.*)

1.15.1 Luminous Flux

The luminous flux is the time rate of the flow of light. The unit of measure is the lumen. One lumen may be defined as the light flux emitted in a one-unit solid angle by a one-candela uniform-point source. The *lumen* differs from the candela in that it is a measure of light flux irrespective of direction. The lumen is used to express a quantity of light flux: total output of a source; output within a specific

www.TechnicalBooksPDF.com

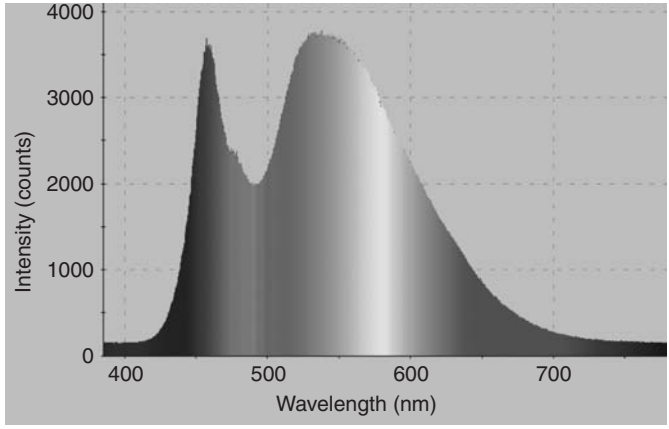


Figure 1.19 Spectrographic analysis of a 10-mm non-phosphor LED, when driven at 350.7 mA. (*See also color insert.*)

angular zone; amount of absorbed light, etc. However, if you need to calculate a characteristic that is not related to the human eye—for example, the temperature increase due to absorbed light, you cannot use luminous flux. Instead, you must use the correct unit of power—the watt.

1.15.2 Illumination

Illumination is the density of luminous flux on a surface This parameter shows how "bright" the surface point appears to the human eye. The appropriate units of measure are *foot-candle* and *lux*. One foot-candle is the illumination produced by one lumen uniformly distributed over one square foot of a surface, or conversely this is the illumination at the point of a surface which is one foot from, and perpendicular to, a uniform point source of one candela. So, foot-candles incident on a surface = lumens/area (sq. feet). Lux is used in the International System. Both have a similar objective, but meters are used for lux and feet are used for candelas.

Therefore, one lux = 0.0929 foot-candles.

Or,

$$\approx 1 \text{ fc} = 10 \text{ lux}$$

1.15.3 Luminance (Brightness)

Luminance (or brightness) is the luminous intensity of a surface in a given direction per unit of projected area of the surface. It can be expressed in two ways: in candelas per unit area, or in lumens per

www.TechnicalBooksPDF.com

unit area. We won't delve any further into this subject because it is so seldom used. Many different standard units of measurement exist for luminance, some of which are listed next.

- Candela per square inch (cd/in²)
- Foot-lambert (luminance of a surface emitting one lumen per square foot)
- Lambert (similar, but per square cm).

$$1 \text{ cd/in}^2 = 452 \text{ foot-lambert}$$

$$1 \text{ lambert} = 929 \text{ foot-lambert} = 2.054 \text{ cd/in}^2$$

Our eye sees brightness, not illumination. Every visible object has brightness. Usually, brightness is proportional to the object's illumination, so a well-illuminated object seems brighter. For a properly diffusing reflecting surface:

$$\text{foot-lamberts} = \text{foot-candles} \times \text{surface reflectance}$$

1.15.4 Foot-Candle to Lux Conversion

One foot-candle is the illumination produced by one lumen uniformly distributed over one square foot of surface, and lux is the illumination over one square meter of surface (Fig. 1.20). Therefore, one lux = 0.0929 foot-candles. Or, approximately:

$$\approx 1 \text{ Fc} = 10 \text{ lux}$$

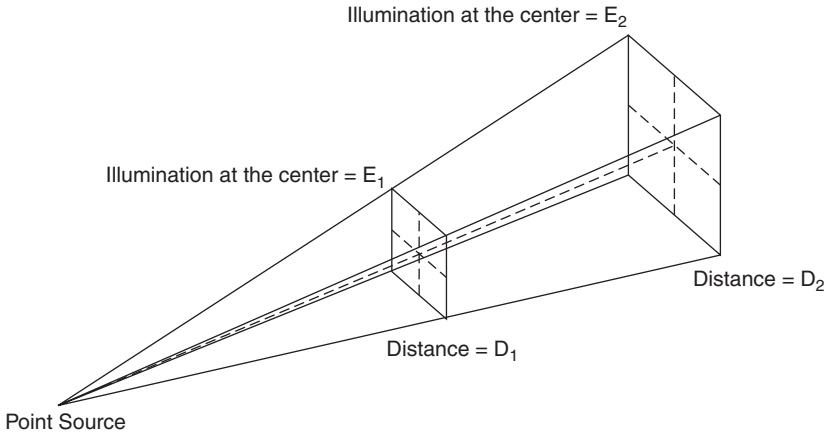


Figure 1.20 Illumination inverse square law representation.

www.TechnicalBooksPDF.com

1.15.5 The Inverse Square Law

The inverse square law tells us that the illumination is inversely proportional to the square of the distance between the point source and the surface, in other words:

$$\frac{E_1}{E_2} = \frac{D_2^2}{D_1^2}$$

If you have a fixture that can be treated as a point source, where the distance from the surface is large, and you measure the illumination at 20 feet as 2000 Fc at the beam center, then at 40 feet the illumination is 500 Fc at the beam center.

1.15.6 The Cosine Law

Effective illumination is proportional to the cosine of the angle of incidence of the light on the surface. It is the angle between the direction of the light and the perpendicular to the surface (Fig. 1.21).

$$E_2 = E_1 \cos\alpha = E_1 \sin\beta$$

Here are a few cases:

When the surface is tilted by an angle of 30°, the illumination is reduced by a factor of 0.87

45° – 0.71

60° – 0.5

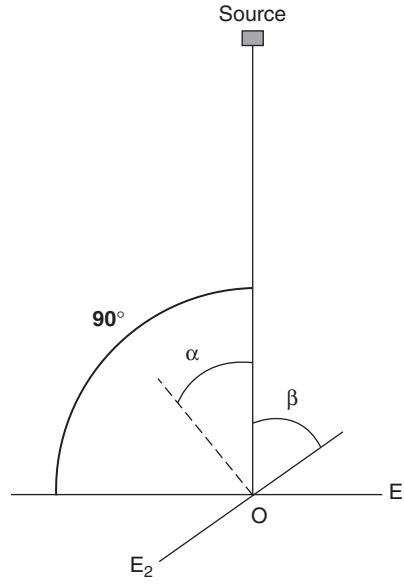


Figure 1.21 Illumination at the "O" point on surfaces "1" and "2."

www.TechnicalBooksPDF.com

1.15.7 The Difference Between Lumen and Watt

Lumen is a unit of the photometric system and Watt belongs to the radiometry system.

Both characterize a power of light flow. However, lumen is power "related" to human eye sensitivity. Therefore, lights with the same power in watts, but that are different colors have different luminous fluxes, because the human eye has different sensitivity at different wavelengths. At a wavelength of 555 nm (maximum eye sensitivity), 1 watt equals 683 lumens (lm).

High-power sources of infrared radiation produce no lumen output, so the human eye can't see it. However, in order to calculate the total power absorbed by a surface to estimate temperature increase, for example, lumen flux must be transferred to watts. This can be done by using a spectral luminous efficiency curve, which can be found in many photometry handbooks.

1.15.8 Calculating Beam Angles

If the distance from a fixture to the screen is known (which is a much larger value than fixture length), as well as the image diameter (Fig. 1.22), then:

$$\alpha = 2\arctan\frac{D}{2L}$$

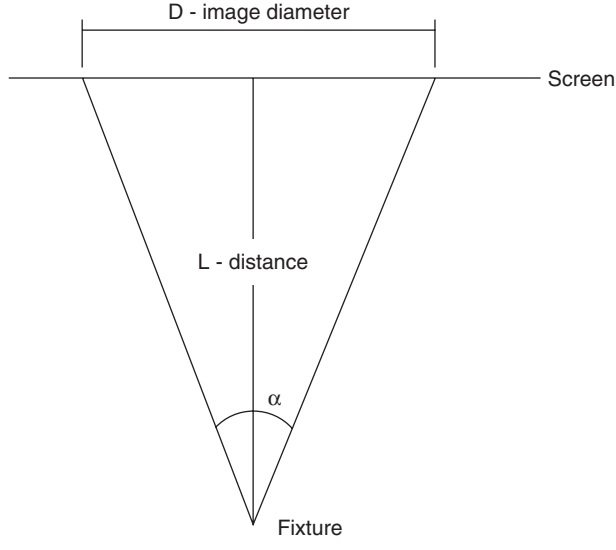


Figure 1.22 Calculation of beam angle.

www.TechnicalBooksPDF.com

In most practical cases, the following approximation is true:

$$\alpha = 57.3 * \frac{D}{L}$$

Example:

distance = 20 feet

image diameter = 5 feet

Exact formula gives the angle = 14.25°

In the case of a "soft edge" light image diameter, it is usually measured at the point where illumination is 50 percent (beam angle) or 10 percent (field angle) of the center illumination.

1.15.9 Calculating Lumens Output—Luminous Flux

The beam radius is divided into "n" equal parts (radii and illumination reading values are indexed from 0, at the beam center, to n–1, at the beam edge). See Fig. 1.23.

1.15.10 Calculating Center and Edge Points—Two Points:

$$P = 2.1R^2(E_1 + 0.5E_0)$$

1.15.11 Calculating Center, Middle, and Edge Points—Three Points:

$$P = 1.3R^2(E_2 + 1.2E_1 + 0.2E_0)$$

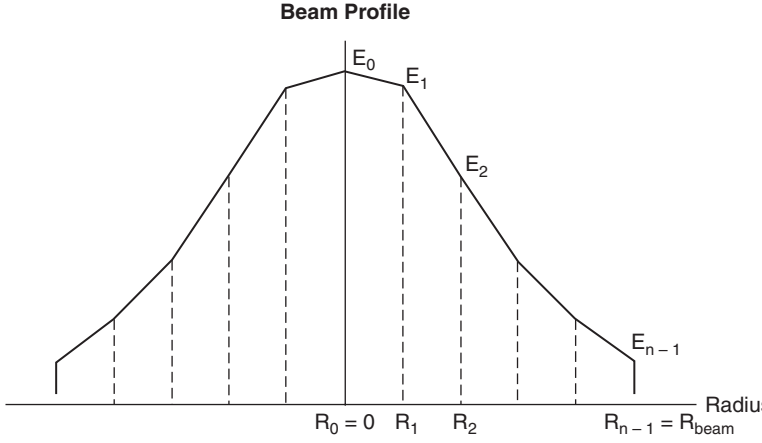


Figure 1.23 Calculation of lumens output.

www.TechnicalBooksPDF.com

1.15.12 Calculating Four Calculation Points:

$$P = 0.93R^2(E_3 + 1.5E_2 + 0.75E_1 + 0.125E_0)$$

1.15.13 Calculating Five Calculation Points:

$$P = 0.72R^2(E_4 + 1.64E_3 + 1.09E_2 + 0.55E_1 + 0.09E_0)$$

1.15.14 Calculating Eight Points:

$$P = 0.43R^2(E_7 + 1.8E_6 + 1.5E_5 + 1.2E_4 + 0.9E_3 + 0.6E_2 + 0.3E_1 + 0.05E_0)$$

where P = total lumens
R = beam radius
E = illumination

In the case of a "soft edge" fixture, where the image size is taken at 10 percent of the center illumination (field angle), the formula becomes:

$$P = 1.26R^2E_{\text{center}}$$

1.16 LED and Spectralon

1.16.1 Spectralon® Material—Light Reflectance Calibration

Spectralon reflectance standards (Fig. 1.24) are available individually and in sets consisting of a diffuse white standard and a selection of diffuse gray standards. Each is supplied with reflectance data from 250 nm to 2500 nm, in 50-nm increments. These durable, chemically

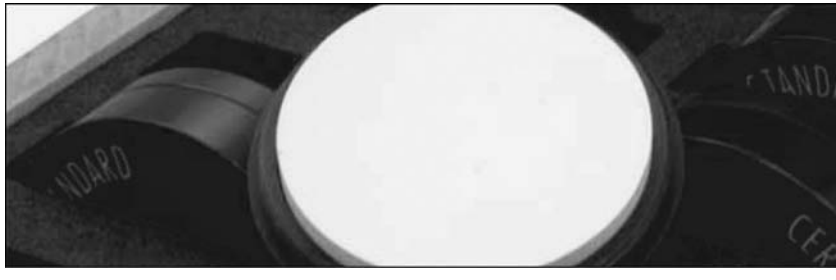


Figure 1.24 Spectralon® diffuse white standards provide the highest diffuse reflectance values of any known substance.

www.TechnicalBooksPDF.com



Figure 1.25 Spectralon® calibration material.

inert, and washable standards have typical reflectance values of 95 to 99 percent and are spectrally flat over the UV-VIS-NIR spectrum. Spectralon SRM-99 reflectance material is the most Lambertian reflector available for use over the wavelength range from 250–2500 nm. All reflectance calibration tests are directly traceable to NIST (Fig. 1.25).

It is used for calibrating:

- Densitometers
- Integrating sphere systems
- Optical equipment
- Photographic equipment
- Reflectometers
- Remote sensing
- Spectroscopy instruments

1.16.2 Spectralon® Features

- Highest diffuse reflectance values of any known substance
- Durable, chemically inert, and washable standards
- Reflectance values of 95–99 percent
- Spectrally flat over the UV-VIS-NIR spectrum
- Data from 250–2500 nm, in 50-nm increments
- NIST traceable calibration

Further Reading

Chappel, A. (ed.), *Optoelectronics: Theory and Practice*, McGraw-Hill, New York, 1978.

Doebelin, E. O., *Measurement Systems: Application and Design*, 4th ed., McGraw-Hill, New York, 1990.

www.TechnicalBooksPDF.com

Holliday, D., and R. Resnick, *Physics*, Wiley, New York, 1975.
International Organization for Standardization, "Statistical Interpretation of Data: Comparison of Two Means in the Case of Paird Observations," ISO 3301–1975.
Lion, K. L., *Elements of Electrical and Electronic Instrumentation*, McGraw-Hill, New York, 1975.
Neubert, H.K.P., *Instrument Transducers*, 2d ed., Clarendon Press, Oxford, 1975.
Ogata, K., *Modern Control Engineering*, 2d ed., Prentice-Hall, Englewood Cliffs, N.J, 1990.
Rock, I., *Lightness Constancy, Perception*, W. H. Freeman, New York, 1984.
Seippel, R. G., *Optoelectronics*, Reston Publishing Co., Reston, Va., 1981.
Shortley, G., and D. Williams, *Quantum Property of Radiation*, Prentice-Hall, Englewood Cliffs, N.J., 1971.
Todd, C. D. (Bourns Inc.), *The Potentiometer Handbook*, McGraw-Hill, New York, 1975.
White, R. M., "A Sensor Classification Scheme," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, March, 1987.

www.TechnicalBooksPDF.com

This page intentionally left blank

www.TechnicalBooksPDF.com

CHAPTER 2
Classification and Types of Sensors

2.0 Introduction

Workstations, work cells, and work centers represent a coordinated cluster of a production system. A production machine with several processes is considered a workstation. A machine tool is also considered a workstation. Integrated workstations form a work cell. Several complementary workstations may be grouped together to construct a work cell. Similarly, integrated work cells may form a work center. This structure is the basic concept in modeling a flexible manufacturing system. The flexible manufacturing system is also the cornerstone of the computer-integrated manufacturing strategy (Fig. 2.1).

The goal is to provide the management and project development team with an overview of major tasks to be solved during the planning, design, implementation, and operation phases of computer-integrated machining, inspection, and assembly systems. Financial and technical disasters can be avoided if a clear understanding of the role of sensors and control systems in the computer-integrated manufacturing strategy is asserted.

Sensors are largely applied within the workstations and are the only practical means of operating a manufacturing system and tracking its performance continuously.

Sensors and control systems in manufacturing provide the means of integrating different, properly defined processes as input to create the expected output. Input may be raw material and/or data that have to be processed by various auxiliary components such as tools, fixtures, and clamping devices. Sensors provide the feedback data to describe the status of each process. The output may also be data and/or materials that can be processed by further cells of the manufacturing system. A flexible manufacturing system, which contains workstations, work cells, and work centers and is equipped with appropriate sensors and control systems, is a distributed management

www.TechnicalBooksPDF.com

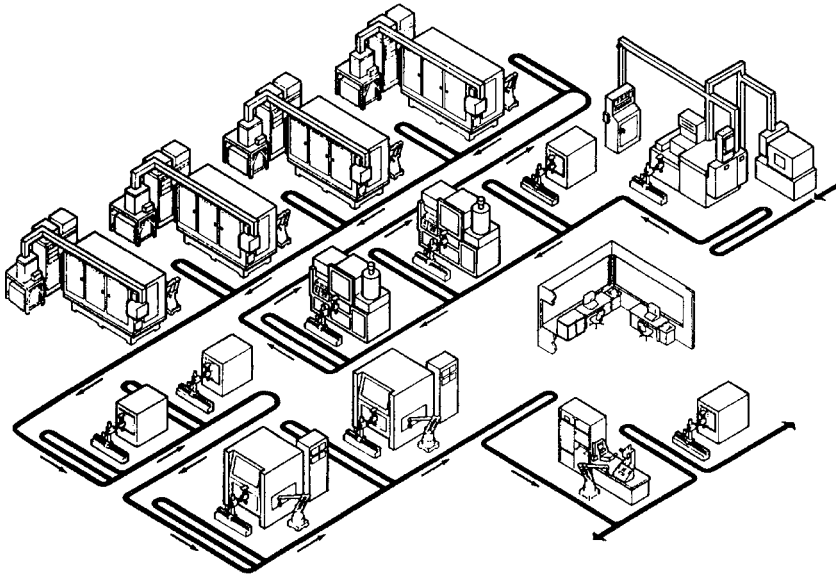


Figure 2.1 Workstation, work cell, and work center.

information system, linking together subsystems of machining, packaging, welding, painting, flame cutting, sheet-metal manufacturing, inspection, and assembling with material-handling and storage processes.

In designing various workstations, work cells, and work centers in a flexible manufacturing system within the computer-integrated manufacturing strategy, the basic task is to create a variety of sensors interconnecting different material-handling systems, such as robots, automated guided-vehicle systems, conveyers, and pallet loading and unloading carts, to allow them to communicate with data processing networks for successful integration with the system.

Figure 2.2 illustrates a cell consisting of several workstations with its input and output, and indicates its basic functions in performing the conversion process, storing workpieces, linking material-handling systems to other cells, and providing data communication to the control system.

The data processing links enable communication with the databases containing part programs, inspection programs, robot programs, packaging programs, machining data, and real-time control data through suitable sensors. The data processing links also enable communication of the feedback data to the upper level of the control hierarchy. Accordingly, the entire work-cell facility is equipped with current data for real-time analysis and fault recovery.

A cluster of manufacturing cells grouped together for particular production operations is called a *work center*. Various work centers

www.TechnicalBooksPDF.com

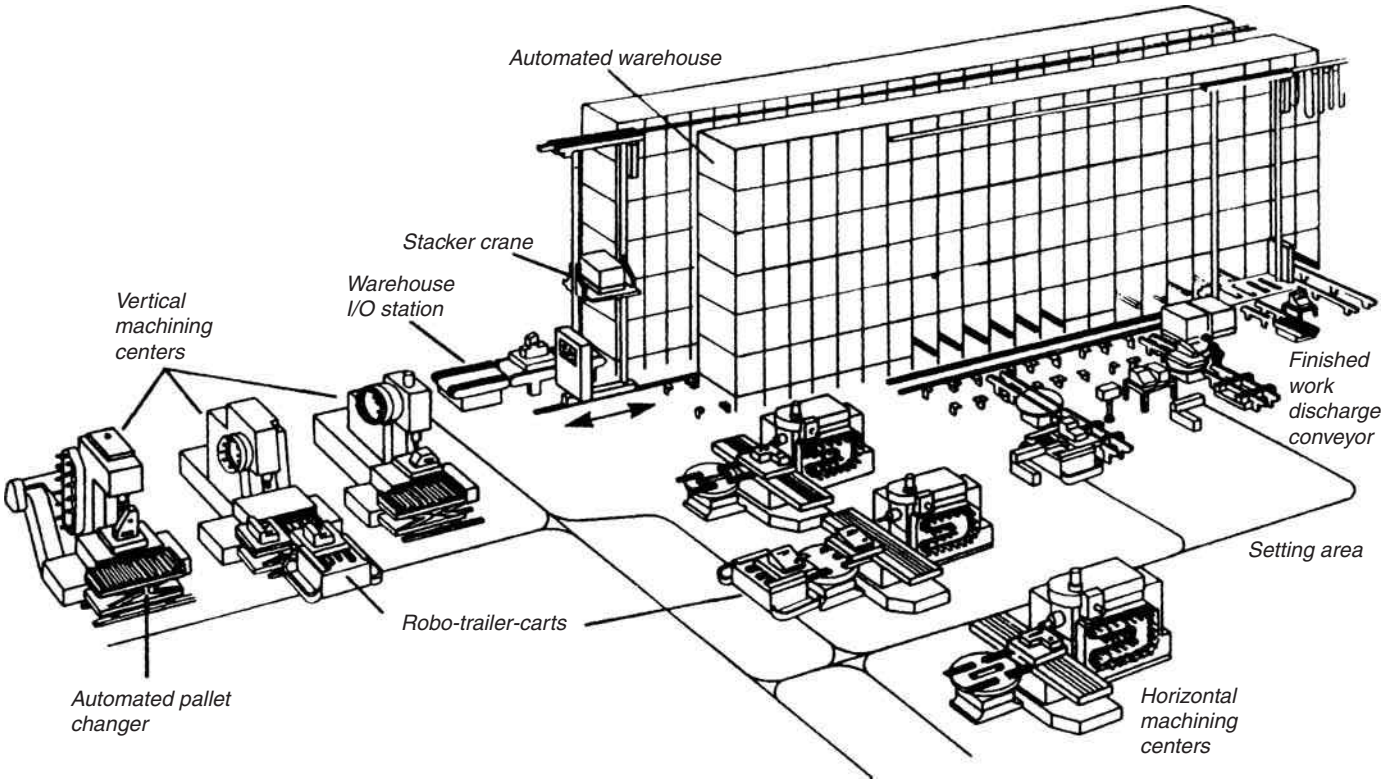


Figure 2.2 Conversion process in a manufacturing cell.

www.TechnicalBooksPDF.com

can be linked together via satellite communication links irrespective of the location of each center. Manufacturing centers can be located several hundred feet apart or several thousand miles apart. Adequate sensors and control systems together with effective communication links will provide practical real-time data analysis for further determination.

The output of the cell is the product of the module of the flexible manufacturing system. It consists of a finished or semi-finished part as well as data in a computer-readable format that will instruct the next cell how to achieve its output requirement. The data are conveyed through the distributed communication networks. If, for example, a part is required to be surfaced to a specific datum in a particular cell, sensors will be adjusted to read the required acceptable datum during the surfacing process. Once the operation is successfully completed, the part must once again be transferred to another cell for further machining or inspection processes. The next cell is not necessarily physically adjacent; it may be the previous cell, as programmed for the required conversion process.

The primary reason for the emphasis on integrating sensors and control systems into every manufacturing operation is the worldwide exponentially increasing demand for error-free production operations. Sensors and control technology can achieve impressive results only if effectively integrated with corporate manufacturing strategy.

The following benefits can be achieved:

- *Productivity*. Greater output and a lower unit cost.
- *Quality*. The product is more uniform and consistent.
- *Production reliability*. The intelligent self-correcting sensory and feedback system increases the overall reliability of production.
- *Lead time*. Parts can be randomly produced in batches of one or in reasonably high numbers, and the lead time can be reduced by 50 to 75 percent.
- *Expenses*. Overall capital expenses are 5 to 10 percent lower. The cost of integrating sensors and feedback control systems into the manufacturing source is less than that of stand-alone sensors and feedback systems.
- *Greater utilization*. Integration is the only available technology with which a machine tool can be utilized up to 85 percent of the time—and the time spent cutting can also be over 90 percent.

In contrast, a part (from stock to finished item) spends only 5 percent of its time on the machine tool, and actual productive work takes only 30 percent of this 5 percent. The time for useful work on

www.TechnicalBooksPDF.com

	Active, %	Idle, %
Tool positioning and tool changing	25	
Machining process	5	
Loading and inspection	15	
Maintenance	20	
Setup	15	
Idle time		15
Total	85	15

Table 2.1 Time Utilization of Integrated Manufacturing Center Carrying Sensory and Control Systems

	Active, %	Idle, %
Machine tool in wait mode		35
Labor control		35
Support services		15
Machining process	15	
Total	15	85

Table 2.2 Productivity Losses of Stand-alone Manufacturing Center Excluding Sensory and Control Systems

stand-alone machines without integrated sensory and control systems is as little as 1 to 1.5 percent of the time available (see Tables 2.1 and 2.2).

To achieve the impressive results indicated in Table 2.1, the integrated manufacturing system carrying the sensory and control feedback systems must maintain a high degree of flexibility. If any cell breaks down for any reason, the production planning and control system can reroute and reschedule the production or, in other words, reassign the system environment. This can be achieved only if both the processes and the routing of parts are programmable. The sensory and control systems will provide instantaneous descriptions of the status of parts to the production and planning system.

If different processes are rigidly integrated into a special-purpose, highly productive system such as a transfer line for large batch production, then neither modular development nor flexible operation is possible.

However, if the cells and their communication links to the outside world are programmable, much useful feedback data may be gained. Data on tool life, measured dimensions of machined surfaces by

www.TechnicalBooksPDF.com

in-process gauging and production control, and fault recovery derived from sensors and control systems can enable the manufacturing system to increase its own productivity, learn its own limits, and inform the part programmers of them. The data may also be very useful to the flexible manufacturing system designers for further analysis. In non-real-time control systems, the data cannot usually be collected, except by manual methods, which are time-consuming and unreliable.

2.1 Classification of Control Processes

An *engineering integrated system* can be defined as a machine responsible for certain production output, a controller to execute certain commands, and sensors to determine the status of the production processes. The machine is expected to provide a certain product as an output, such as computer numerical control (CNC) machines, packaging machines, and high-speed press machines. The controller provides certain commands arranged in a specific sequence designed for a particular operation. The controller sends its commands in the form of signals, usually electric pulses. The machine is equipped with various devices, such as solenoid valves and step motors, that receive the signals and respond according to their functions. The sensors provide a clear description of the status of the machine performance. They give detailed accounts of every process in the production operation (Fig. 2.3).

Once a process is executed successfully, according to a specific sequence of operations, the controller can send additional commands for further processes until all processes are executed. This completes one cycle. At the end of each cycle, a command is sent to begin a new loop until the production demand is met.

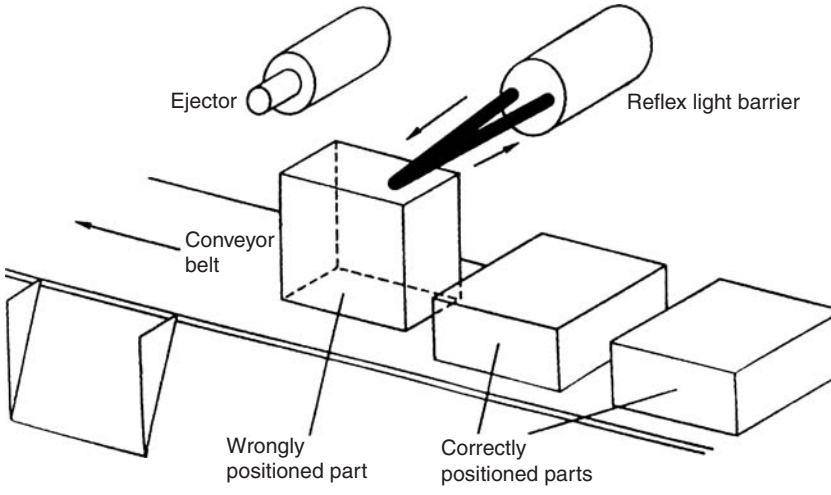


Figure 2.3 Sensors providing machine status.

www.TechnicalBooksPDF.com

In an automatic process, the machine, the controller, and the sensors interact with one another to exchange information. Mainly, there are two types of interaction between the controller and the rest of the system: through either an open-loop control system or a closed-loop control system.

An *open-loop control system* (Fig. 2.4) can be defined as a system in which there is no feedback. Motor motion is expected to faithfully follow the input command. Stepping motors are an example of open-loop control.

A *closed-loop control system* (Fig. 2.5) can be defined as a system in which the output is compared to the command, with the result being used to force the output to follow the command. Servo systems are an example of closed-loop control.

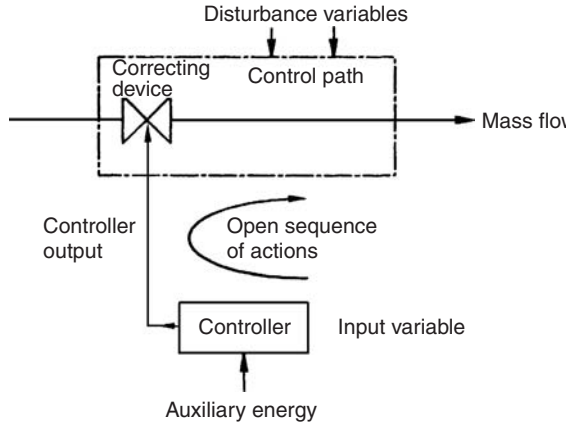


Figure 2.4 An open-loop control system.

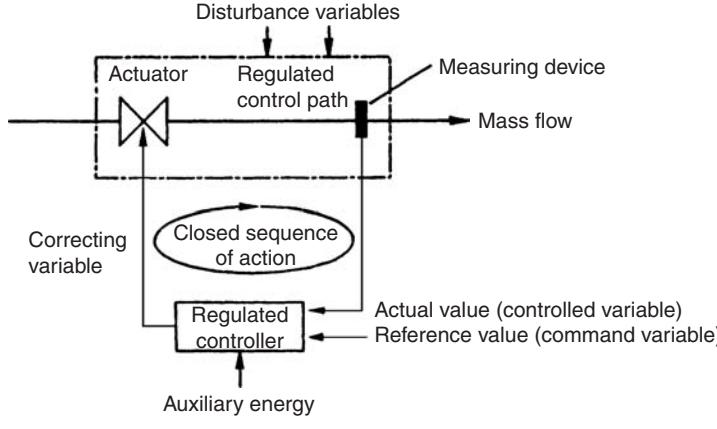


Figure 2.5 A closed-loop control system.

www.TechnicalBooksPDF.com

2.2 Open- and Closed-Loop Control Systems

In an open-loop control system, the actual value in Fig. 2.4 may differ from the reference value in the system. In a closed-loop system, the actual value is constantly monitored against the reference value described in Fig. 2.5.

The mass flow illustrated in Fig. 2.6 describes the amount of matter per unit time flowing through a pipeline that must be regulated. The current flow rate can be recorded by a measuring device, and a correcting device such as a valve may be set to a specific flow rate. The system, if left on its own, may suffer fluctuations and disturbances that will change the flow rate. In such an open-loop system, the reading of the current flow rate is the *actual value*, and the *reference value* is the desired value of the flow rate. The reference value may differ from the actual value, which then remains unaltered.

If the flow rate falls below the reference value because of a drop in pressure, as illustrated in Fig. 2.7, the valve must be opened further

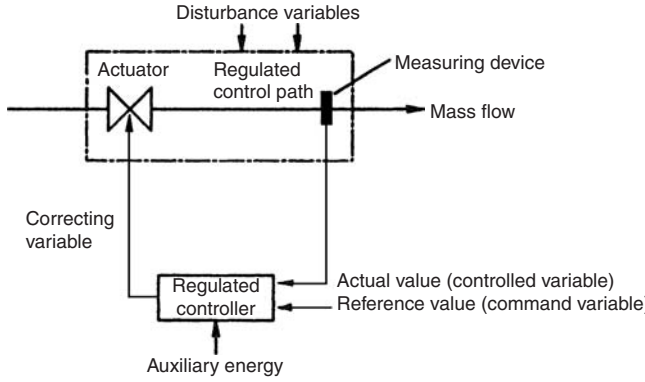


Figure 2.6 Regulation of mass flow.

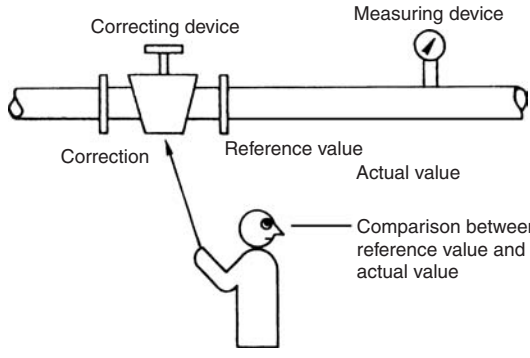


Figure 2.7 Reference value.

www.TechnicalBooksPDF.com

to maintain the desired actual value. Where disturbances occur, the course of the actual value must be continuously observed. When adjustment is made to continuously regulate the actual value, the loop of action governing measurement, comparison, adjustment, and reaction within the process is called a *closed loop*.

2.3 Understanding Photoelectric Sensors

In order to successfully automate a process, it is necessary to obtain information about its status. The sensors are the part of the control system responsible for collecting and preparing process status data and passing it on to a processor (Fig. 2.8).

2.3.1 Principles of Operation

Photoelectric controls use light to detect the presence or absence of an object. All photoelectric controls consist of a sensor, a control unit, and an output device. A logic module or other accessories can be added to the basic control to add versatility. The sensor consists of a source and a detector. The source is a light-emitting diode (LED) that emits a powerful beam of light either in the infrared or visible light spectrum. The detector is typically a photodiode that senses the presence or absence of light. The detection amplifier in all photoelectric controls is designed so it responds to the light emitted by the source. Ambient light, including sunlight up to 3100 metercandles, does not affect operation.

The source and detector may be separated or may be mounted in the same sensor head, depending on the particular series and application (Fig. 2.9).

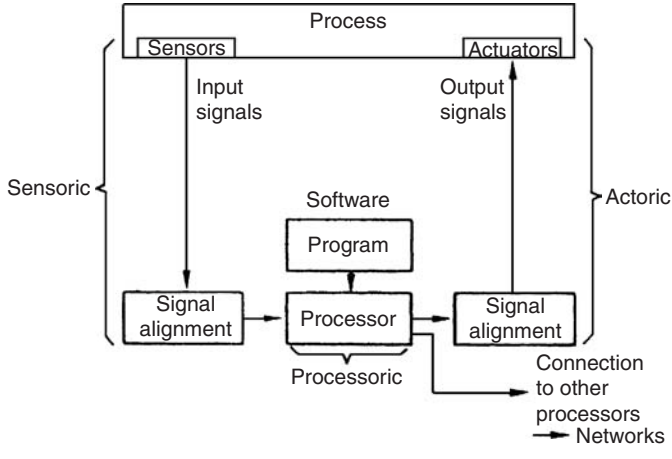


Figure 2.8 Components of controlled process.

www.TechnicalBooksPDF.com

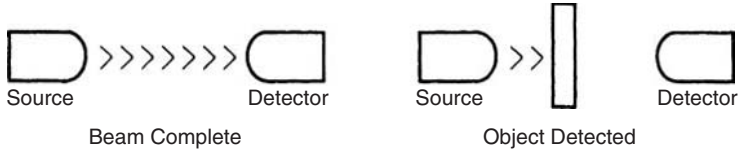


Figure 2.9 Components of photoelectric control.

The control unit modulates and demodulates the light sent and received by the source and detector. This assures that the photoelectric control responds only to its light source. The control unit also controls the output device in *self-contained* photoelectric controls. The control unit and sensor are built into an integral unit.

Controls can be configured to operate as light-actuated devices. The output is triggered when the detector sees light. They can also be dark-actuated devices, where the output is triggered when the detector does not see light.

Output devices may include relays such as *double pole, double throw* (*DPDT*) and *single pole, double throw* (*SPDT*). Output devices may also include a triac or other high-current device and may be programmable-controller-compatible.

Logic modules are optional devices that allow addition of logic functions to a photoelectric control. For example, instead of providing a simple ON/OFF signal, a photoelectric control can (with a logic module) provide time-delay, one-shot, retriggerable one-shot, motion-detection, and counting functions.

2.3.2 Manufacturing Applications of Photodetectors

The following applications of photoelectric sensors are based on normal practices at home, at the workplace, and in various industries. The effective employment of photoelectric sensors can lead to successful integration of data in manufacturing operations to maintain an error-free environment and assist in obtaining instantaneous information for dynamic interaction.

A photoelectric sensor is a semiconductor component that reacts to light or emits light. The light may be either in the visible range or the invisible infrared range. These characteristics of photoelectric components have led to the development of a wide range of photoelectric sensors.

A photoelectric reflex sensor equipped with a time-delay module set for *delay dark* ignores momentary beam breaks. If the beam is blocked longer than the predetermined delay period, the output energizes to sound an alarm or stop the conveyer (Fig. 2.10).

A set of photoelectric through-beam sensors can determine the height of a scissor lift, as illustrated in Fig. 2.11. For example, when

www.TechnicalBooksPDF.com

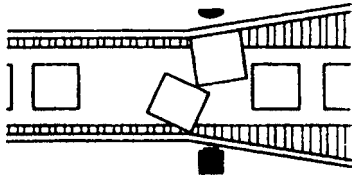


Figure 2.10 Jam detection with photoelectric sensor.

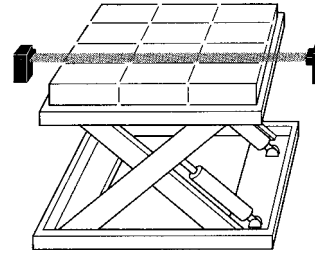


Figure 2.11 Stack height measurement with photoelectric sensor.

the control is set for *dark-to-light* energizing, the lift rises after a layer has been removed and stops when the next layer breaks the beam again.

Cans on a conveyer are diverted to two other conveyers controlled by a polarized photoelectric reflex sensor with a divider module (Fig. 2.12). Items can be counted and diverted in groups of 2, 6, 12, or 24. A polarized sensor is used so that shiny surfaces may not falsely trigger the sensor control.

Two photoelectric control sensors can work together to inspect a fill level in cartons on a conveyer (Fig. 2.13). A reflex photoelectric sensor detects the position of the carton and energizes another synchronized photoelectric sensor located above the contents. If the photoelectric sensor positioned above the carton does not "see" the fill level, the carton does not pass inspection.

A single reflex photoelectric sensor detects boxes anywhere across the width of a conveyer. Interfacing the sensor with a programmable controller provides totals at specific time intervals (Fig. 2.14).

High-temperature environments are accommodated by the use of fiber optics. The conveyer motion in a 450°F cookie oven can be detected as shown in Fig. 2.15. If the motion stops, the one-shot logic

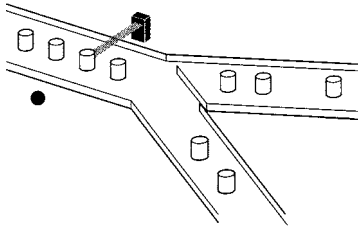


Figure 2.12 Batch counting and diverting with photoelectric sensor.

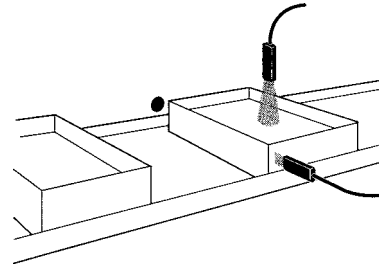


Figure 2.13 Measuring carton fill with photoelectric sensor.

www.TechnicalBooksPDF.com

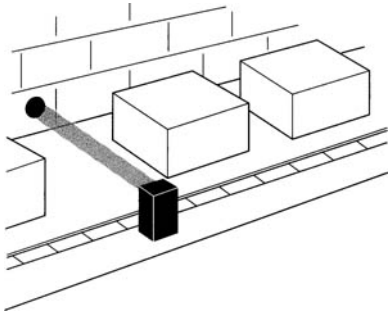


Figure 2.14 Box counting with photoelectric sensor.

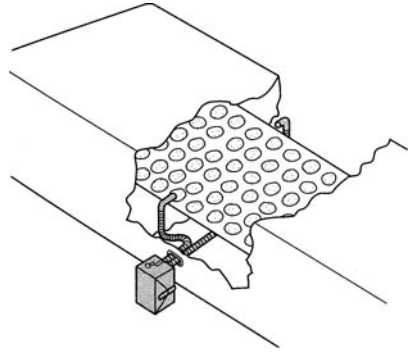


Figure 2.15 Detecting proper cookie arrangement.

module detects light or dark that lasts too long, and the output device shuts the oven down.

Placing the photoelectric sensor to detect a saw tooth (Fig. 2.16) enables the programmable controller to receive an input signal that rotates the blade into position for sharpening of the next tooth.

A through-beam photoelectric sensor is used to time the toll gate in Fig. 2.17. To eliminate toll cheating, the gate lowers the instant the rear of the paid car passes the control. The rugged sensor can handle harsh weather, abuse, and 24-hour operation.

A safe and secure garage is achieved through the use of a through-beam photoelectric sensor interfaced to the door controller. The door shuts automatically after a car leaves, and if the beam is broken while the door is lowering, the motor reverses direction and raises the door again (Fig. 2.18).

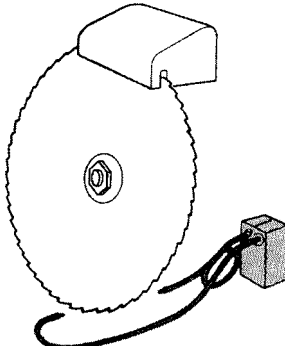


Figure 2.16 Sawtooth inspection.

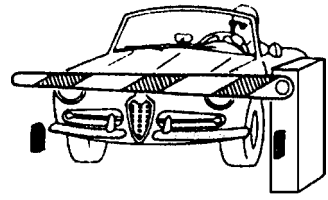


Figure 2.17 Toll-booth control with photoelectrical sensor.

www.TechnicalBooksPDF.com

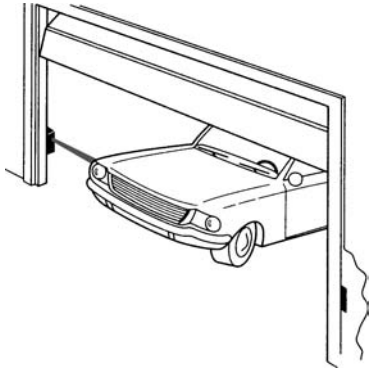


Figure 2.18 Garage door control with photoelectric control.

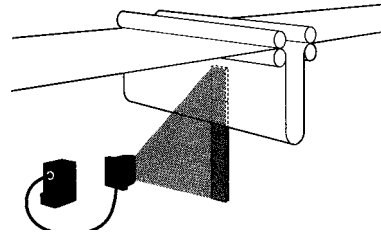


Figure 2.19 Web loop control.

A photoelectric sensor that generates a "curtain of light" detects the length of a loop on a web drive system by measuring the amount of light returned from an array of retro-reflectors. With this information, the analog control unit instructs a motor controller to speed up or slow down the web drive (Fig. 2.19).

Small objects moving through a curtain of light are counted by a change in reflected light. A low-contrast logic module inside the photoelectric sensor unit responds to slight but abrupt signal variations while ignoring slow changes such as those caused by dust buildup (Fig. 2.20).

A pair of through-beam photoelectric sensors scan over and under multiple strands of thread. If a thread breaks and passes through one of the beams, the low-contrast logic module detects the sudden changes in signal strength and energizes the output. As this logic module does not react to slow changes in signal strength, it can operate in a dusty environment with little maintenance (Fig. 2.21).

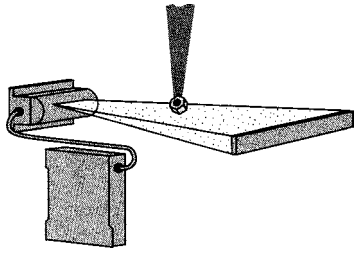


Figure 2.20 Small parts detection.

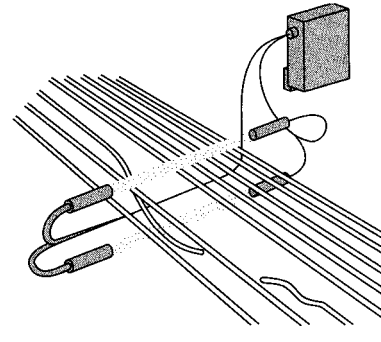


Figure 2.21 Broken-thread detection.

www.TechnicalBooksPDF.com

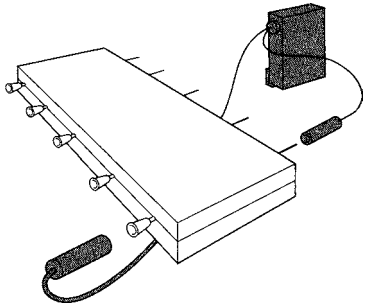


Figure 2.22 Hypodermic needle quality assurance.

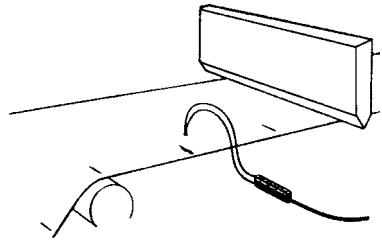


Figure 2.23 Indexing mark detection.

A remote photoelectric source and detector pair inspects for passage of light through a hypodermic needle (Fig. 2.22). The small, waterproof stainless-steel housing is appropriate for crowded machinery spaces and frequent wash-downs. High signal strength allows quality inspection of hole sizes down to 0.015 mm.

Index marks on the edge of a paper are detected by a fiber-optic photoelectric source/detector sensor to control a cutting shear down line (Fig. 2.23).

Liquids are monitored in a clear tank through beam sensors and an analog control. Because the control produces a voltage signal proportional to the amount of detected light, liquid mixtures and densities can be controlled (Fig. 2.24).

Remote photoelectric sensors look for the presence of holes in a metal casting (Fig. 2.25). Because each hole is inspected, accurate information is recorded. A rugged sensor housing an extremely high signal strength handle dirt and grease with minimum maintenance. The modular control unit allows for dense packaging in small enclosures.

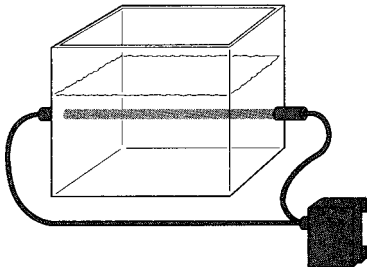


Figure 2.24 Liquid clarity control.

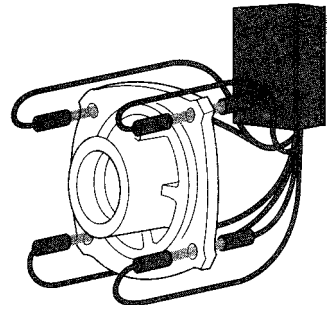


Figure 2.25 Multihole casting inspection.

www.TechnicalBooksPDF.com

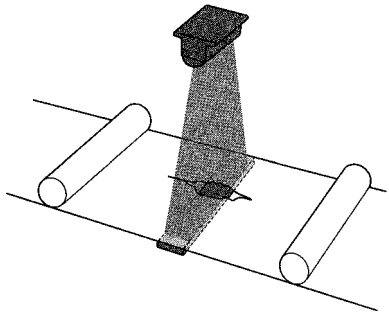


Figure 2.26 Web flaw detection.

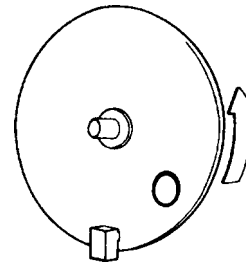


Figure 2.27 Over/under-speed of rotating disk.

In a web flaw detection application, a web passes over an array of retroreflectors (Fig. 2.26). When light is returned to the sensor head, the output is energized and the web shuts down. High web speeds can be maintained because of the superior response time of the control unit.

A reflex photoelectric sensor with a motion control module counts the revolutions of a wheel to monitor over/underspeed of a rotating object. Speed is controlled by a programmable controller. The rate ranges from 2.4 to 12,000 counts per minute (Fig. 2.27).

When the two through-beam photoelectric sensors in Fig. 2.28 observe the same signal strength, the output is zero. When the capacity of the web changes, as in a splice, the signal strengths are thrown out of balance and the output is energized. This system can be used on webs of different colors and opacities with no system reconfiguration.

Understanding the environment is important to effective implementation of an error-free environment. An awareness of the characteristics of photoelectric controls and the different ways in which they

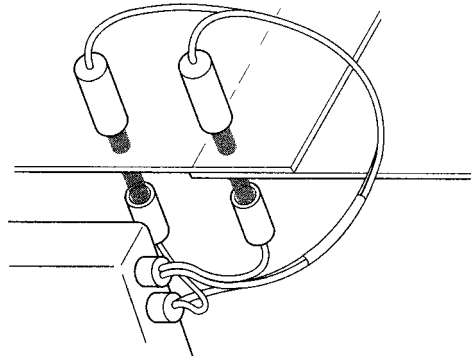


Figure 2.28 Web splice detection.

www.TechnicalBooksPDF.com

Key Point	Consideration
Range	How far is the object to be detected?
Environment	How dirty or dark is the environment?
Accessibility	What accessibility is there to both sides of the object to be detected?
Wiring	Is wiring possible to one or both sides of the object?
Size	What size is the object?
Consistency	Is object consistent in size, shape, and reflectivity?
Requirements	What are the mechanical and electrical requirements?
Output signal	What kind of output is needed?
Logic functions	Are logic functions needed at the sensing point?
Integration	Is the system required to be integrated?

Table 2.3 Key Characteristics of Sensors

can be used will establish a strong foundation. This understanding also will allow the user to obtain a descriptive picture of the condition of each manufacturing process in the production environment.

Table 2.3 highlights key questions the user must consider.

2.4 Detection Methods

The three modes of detection used by photoelectric sensors are:

- Through-beam detection
- Reflex detection
- Proximity detection

2.4.1 The Through-Beam Detection Method

The through-beam method requires that the source and detector are positioned opposite each other and that the light beam is sent directly from source to detector (Fig. 2.29). When an object passes between the source and detector, the beam is broken, signaling detection of the object.

Through-beam detection generally provides the longest range of the three operating modes and provides high power at shorter range to penetrate steam, dirt, or other contaminants between the source and detector. Alignment of the source and detector must be accurate.

2.4.2 The Reflex Detection Method

The reflex method requires that the source and detector be installed on the same side of the object to be detected (Fig. 2.30). The light beam is transmitted from the source to a retroreflector that returns

www.TechnicalBooksPDF.com

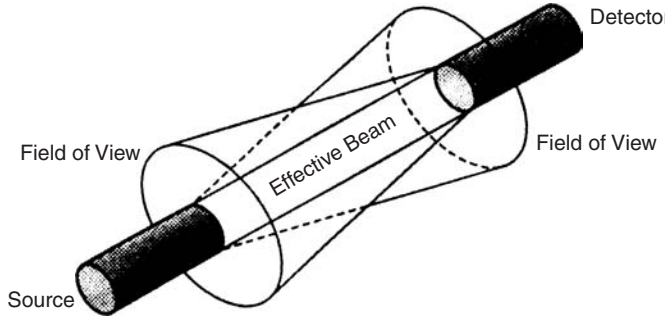


Figure 2.29 Through-beam detection.

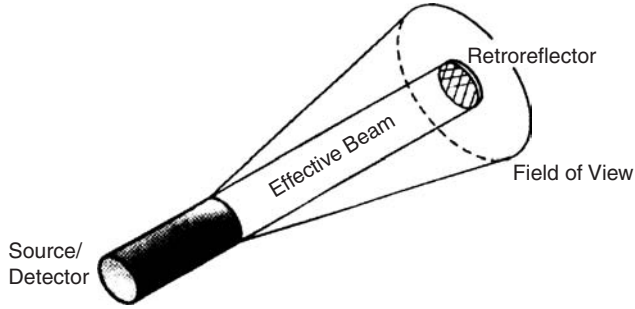


Figure 2.30 Reflex detection.

the light to the detector. When an object breaks a reflected beam, the object is detected.

The reflex method is widely used because it is flexible and easy to install and provides the best cost-performance ratio of the three methods. The object to be detected must be less reflective than the retroreflector.

2.4.3 The Proximity Detection Method

The proximity method requires that the source and detector are installed on the same side of the object to be detected and aimed at a point in front of the sensor (Fig. 2.31). When an object passes in front of the source and detector, light from the source is reflected from the object's surface back to the detector, and the object is detected.

Each sensor type has a specific operating range. In general, through-beam sensors offer the greatest range, followed by reflex sensors, then proximity sensors.

The maximum range for through-beam sensors is of primary importance. At any distance less than the maximum range, the sensor has more than enough power to detect an object.

www.TechnicalBooksPDF.com

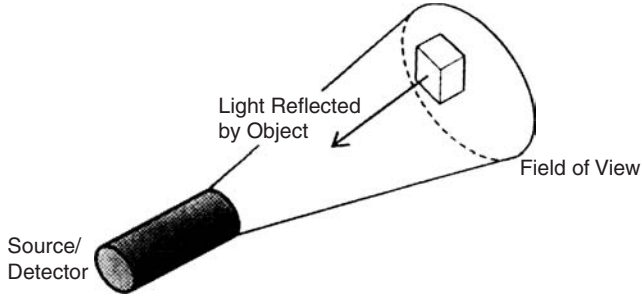


Figure 2.31 Proximity detection.

The optimum range for the proximity and reflex sensors is more significant than the maximum range. The optimum range is the range at which the sensor has the most power available to detect objects. The optimum range is best shown by an excess gain chart (Fig. 2.32).

Excess gain is a measure of sensing power available in excess of that required to detect an object. An excess gain of 1 means there is just enough power to detect an object under the best conditions without obstacles placed in the light beam. The distance at which the excess gain equals 1 is the maximum range. An excess gain of 100 means there is 100 times the power required to detect an object. Generally, the more excess gain available at the required range, the more consistently the control will operate.

For each distance within the range of the sensor, there is a specific excess gain. Through-beam controls generally provide the most excess gain, followed by reflex and then proximity sensors.

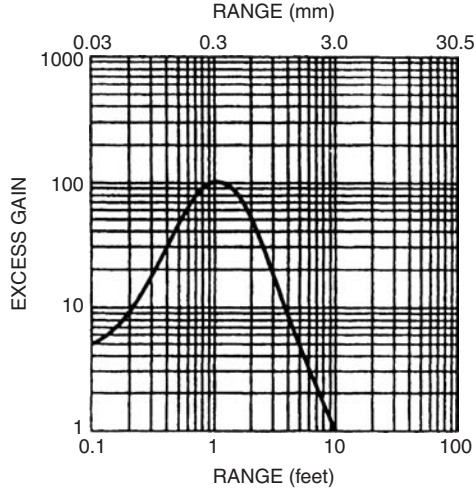


Figure 2.32 Photoelectric excess gain and range.

www.TechnicalBooksPDF.com

Environment	Through-beam	Reflex	Proximity
Relatively clean	1.25 per side	1.6 per side	
Office clean	1.6 total	2.6 total	2.6 total
Lightly dirty	1.8 per side	3.2 per side	
Warehouse, post office	3.2 total	10.5 total	3.2 total
Dirty	8 per side	64 per side	
Steel mill, saw mill	64 total		64 total
Very dirty	25 per side		
Steam tunnel, painting rubber or grinding, cutting with coolant, paper plant	626 total		
Extremely dirty	100 per side		
Coal bins or areas where thick layers build quickly	10,000 total		

TABLE 2.4 Excess Gain Chart

General guidelines can be provided for the quantity of excess gain required for the amount of contamination in an environment. Environments are classified as one of the following: relatively clean, lightly dirty, dirty, very dirty, and extremely dirty. Table 1.4 illustrates the excess gain recommended for these types of environments for each sensing mode.

Example. If in a through-beam setup, the source is in a lightly dirty environment where excess gain is 1.8, and the detector is in a very dirty environment where excess gain is 25, the recommended excess gain is $1.8 \times 25 = 45$, from Table 1.4.

2.5 Proximity Sensors

Proximity sensing is the technique of detecting the presence or absence of an object with an electronic noncontact sensor.

Mechanical limit switches were the first devices to detect objects in industrial applications. A mechanical arm touching the target object moves a plunger or rotates a shaft, which causes an electrical contact to close or open. Subsequent signals will produce other control functions through the connecting system. The switch may be activating a simple control relay, or a sophisticated programmable logic

www.TechnicalBooksPDF.com

control device, or a direct interface to a computer network. This simple activity, once done successfully, will enable varieties of manufacturing operations to direct a combination of production plans according to the computer-integrated manufacturing strategy.

Inductive proximity sensors are used in place of limit switches for noncontact sensing of metallic objects. Capacitive proximity switches are used on the same basis as inductive proximity sensors; however, capacitive sensors can also detect nonmetallic objects. Both inductive and capacitive sensors are limit switches with ranges up to 100 mm.

The distinct advantage of photoelectric sensors over inductive or capacitive sensors is their increased range. However, dirt, oil mist, and other environmental factors will hinder operation of photoelectric sensors during the vital operation of reporting the status of a manufacturing process. This may lead to significant waste and buildup of false data.

2.5.1 Typical Applications of Inductive Proximity Sensors

Motion position detection (Fig. 2.33):

- Detection of rotating motion
- Zero-speed indication
- Speed regulation

Motion control (Fig. 2.34):

- Shaft travel limiting
- Movement indication
- Valve open/closed

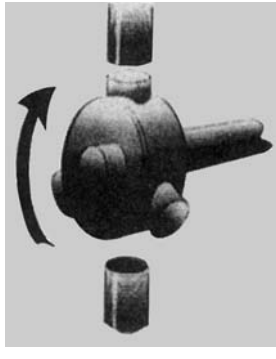


Figure 2.33 Motion/position detection with inductive proximity sensor.

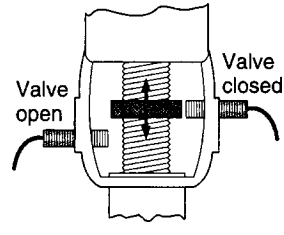


Figure 2.34 Motion control, inductive proximity sensor.

www.TechnicalBooksPDF.com

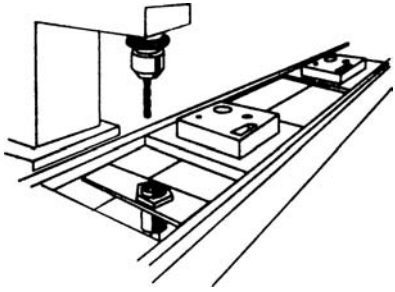


Figure 2.35 Conveyer system control, inductive proximity sensor.

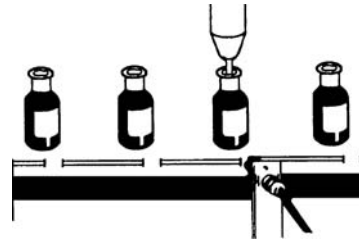


Figure 2.36 Process control, inductive proximity sensor.

Conveyer system control (Fig. 2.35):

- Transfer lines
- Assembly line control
- Packaging machine control

Process control (Fig 2.36):

- Product complete
- Automatic filling
- Product selection

Machine control (Fig. 2.37):

- Fault condition indication
- Broken tool indication
- Sequence control

Verification and counting (Fig. 2.38):

- Product selection
- Return loop control
- Product count

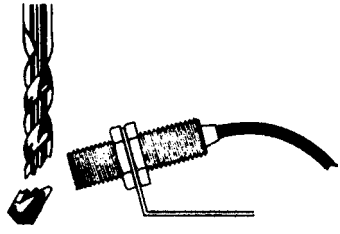


Figure 2.37 Machine control, inductive proximity sensor.

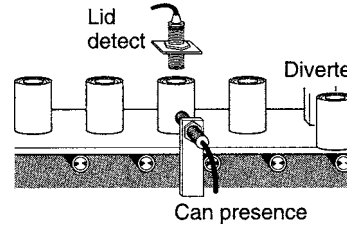


Figure 2.38 Verification and counting, inductive proximity sensor.

www.TechnicalBooksPDF.com

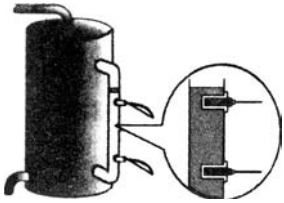


Figure 2.39 Liquid level detection, capacitive proximity sensor.

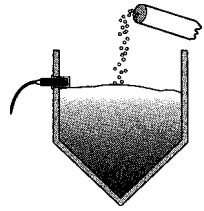


Figure 2.40 Bulk material level control, capacitive proximity sensor.

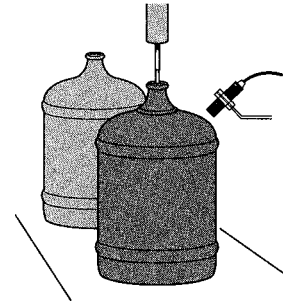


Figure 2.41 Process control, capacitive proximity sensor.

2.5.2 Typical Applications of Capacitive Proximity Sensors

Liquid level detection (Fig. 2.39):

- Tube high/low liquid level
- Overflow limit
- Dry tank

Bulk material level control (Fig. 2.40):

- Low level limit
- Overflow limit
- Material present

Process control (Fig. 2.41):

- Product present
- Bottle fill level
- Product count

2.6 Understanding Inductive Proximity Sensors

2.6.1 Principles of Operation

An inductive proximity sensor consists of four basic elements (Fig. 2.42):

- Sensor coil and ferrite core
- Oscillator circuit
- Detector circuit
- Solid-state output circuit

www.TechnicalBooksPDF.com

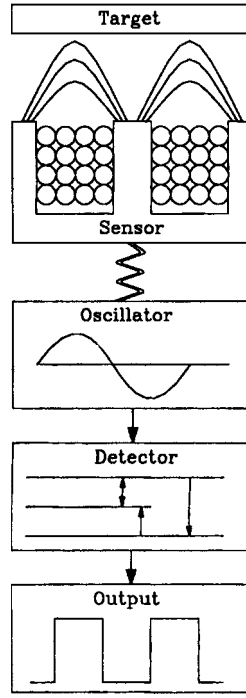


Figure 2.42 Operating principle of inductive proximity sensor.

The oscillator circuit generates a radio-frequency electromagnetic field that radiates from the ferrite core and coil assembly. The field is centered around the axis of the ferrite core, which shapes the field and directs it at the sensor face. When a metal target approaches and enters the field, eddy currents are induced into the surfaces of the target. This results in a loading effect, or "damping," that causes a reduction in amplitude of the oscillator signal (Fig. 2.43).

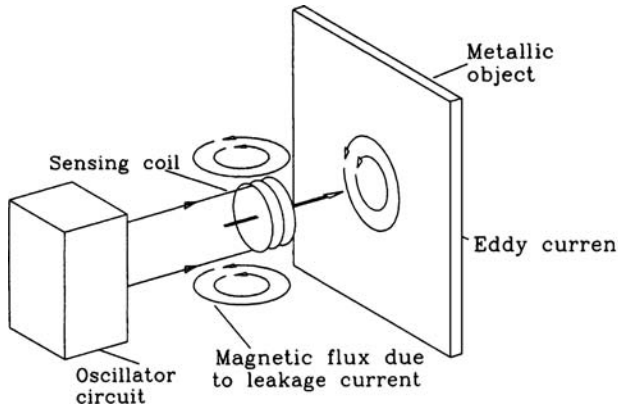


Figure 2.43 Induced eddy current.

www.TechnicalBooksPDF.com

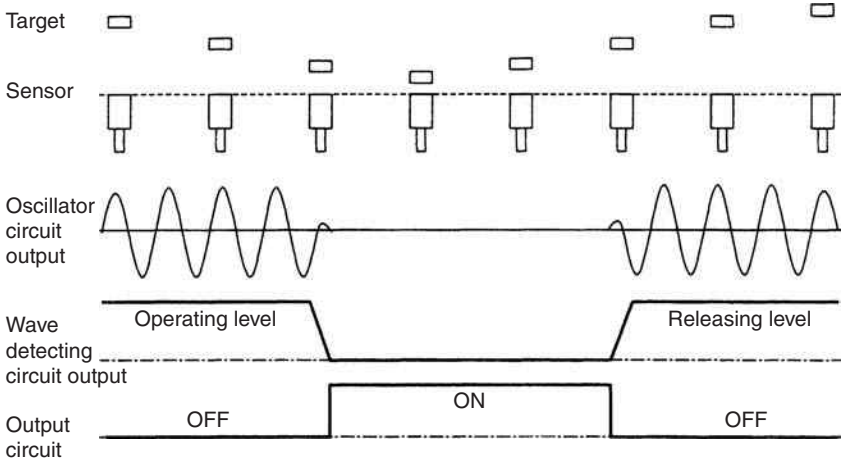
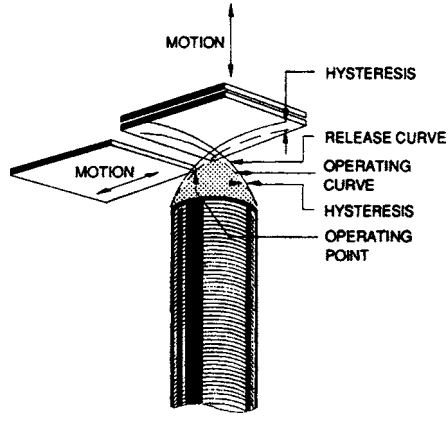


Figure 2.44 Detection cycle.

The detector circuit detects the change in oscillator amplitude (Fig. 2.44). The detector circuit will switch ON at a specific operate amplitude. This ON signal generates a signal to turn ON the solid-state output. This is often referred to as the *damped* condition. As the target leaves the sensing field, the oscillator responds with an increase in amplitude. As the amplitude increases above a specific value, it is detected by the detector circuit, which switches OFF, causing the output signal to return the normal or OFF (*undamped*) state.

The difference between the operate and the release amplitude in the oscillator and corresponding detector circuit is referred to as the *hysteresis* (H) of the sensor. It corresponds to a difference in the point of target detection and the release distance between the sensor face and the target surface (Fig. 2.45).

Figure 2.45
Core assembly.



2.6.2 Inductive Proximity Sensing Range

The *sensing range* of an inductive proximity sensor refers to the distance between the sensor face and the target. It also includes the shape of the sensing field generated through the coil and core. Several mechanical and environmental factors affect the sensing range, such as those shown in the following table.

Mechanical Factors	Environmental Factors
Core size	Ambient temperature
Core shield	Surrounding electrical conditions
Target material	Surrounding mechanical conditions
Target size	Variation between devices
	Target shape

The geometry of the sensing field can be determined by the construction factor of the core and coil. An open coil with no core produces an omnidirectional field. The geometry of an air-core is a toroid. Such sensors could be actuated by a target approaching from any direction, making them undesirable for practical industrial applications (Fig. 2.46).

Ferrite material in the shape of a cup core is used to shape the sensing field. The ferrite material absorbs the magnetic field, but enhances the field intensity and directs the field out of the open end of the core (Fig. 2.47).

A standard field range sensor is illustrated in Fig. 2.48. It is often referred to as *shielded* sensing coil. The ferrite contains the field so that

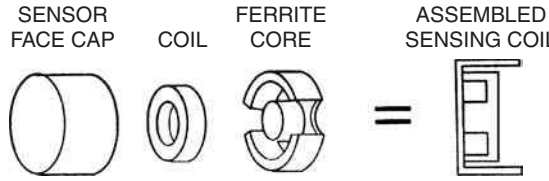


Figure 2.46 Open coil without core.

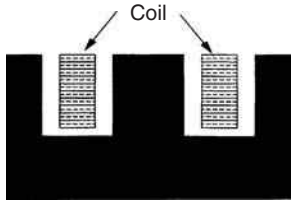


Figure 2.47 Cup-shaped coil/core assembly.

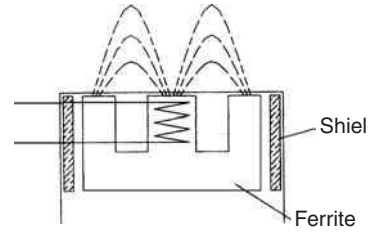


Figure 2.48 Standard range core coil.

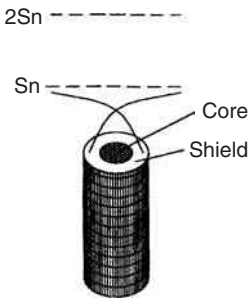


Figure 2.49 Standard range field plot.

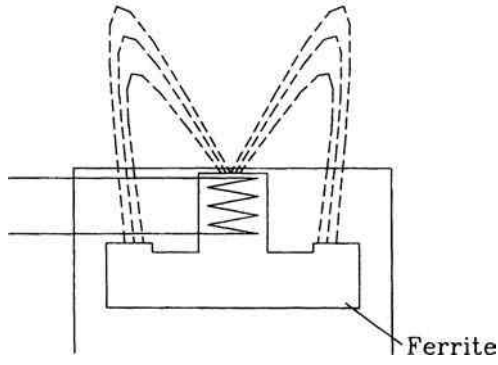


Figure 2.50 Extended range core and coil.

it emanates straight from the sensing face. Figure 2.49 shows the typical standard-range sensing-field plot.

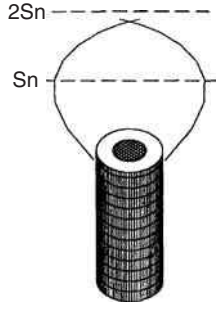
An extended range coil and core assembly does not have the ferrite around the perimeter of the coil (Fig. 2.50). This unshielded device accordingly has an extended range. Figure 2.51 illustrates a typical extended sensing-field plot.

2.6.3 Sensing Distance

The electromagnetic field emanates from the coil and core at the face of the sensor and is centered around the axis of the core. The nominal sensing range is a function of the coil diameter and the power that is available to operate the electromagnetic field.

The sensing range is subject to manufacturing tolerances and circuit variations. Typically, it varies by 10 percent. Similarly, temperature drift can affect the sensing range by 10 percent. Applied to the nominal sensing switch, these variations mean that the sensing

Figure 2.51 Extended range field plot.



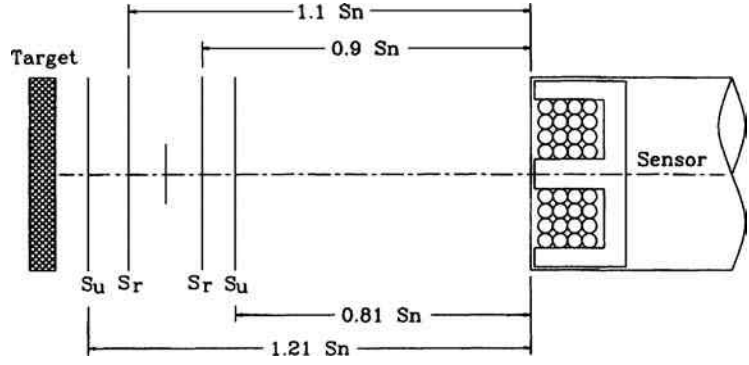


FIGURE 2.52 Sensing distance tolerances.

range can be as much as 120 percent or as little as 81 percent of the nominal stated range (Fig. 2.52).

$$S_r = 0.9 < S_n < 1.1$$

$$S_n = 0.9 < S_r < 1.1$$

where S_n = nominal sensing range
S_r = effective sensing range
S_u = usable sensing range

2.6.4 Target Material and Size

As a target approaches the sensing field, eddy currents are induced in the target. In order to ensure that the target has the desired damping affect on the sensor, the target must be of the appropriate size and material. Metallic targets can be defined as:

- *Ferrous.* Containing iron, nickel, or cobalt
- *Nonferrous.* All other metallic materials, such as aluminum, copper, and brass.

Eddy currents induced in ferrous targets are stronger than in nonferrous targets, as illustrated in Table 1.5.

An increase in target size will not produce an increase in sensing range. However, a decrease in target size will produce a decrease in sensing range, and may also increase response time. Figure 2.53 illustrates the relationship of target size and target material to the sensing range of a limit-switch-type sensor with a nominal 13-mm sensing range.

Table 1.6 shows correction factors by which the rated nominal sensing range of most inductive proximity sensors can be multiplied.

Device	Standard Target Dimensions	Standard Target Material
Modular limit switch type	45 mm square × 1 mm thick	Mild steel
8 mm tubular	8 mm square × 1 mm thick	Mild steel
12 mm tubular	12 mm square × 1 mm thick	Mild steel
18 mm tubular	18 mm square × 1 mm thick	Mild steel
30 mm tubular	30 mm square × 1 mm thick	Mild steel

Table 2.5 Standard Target

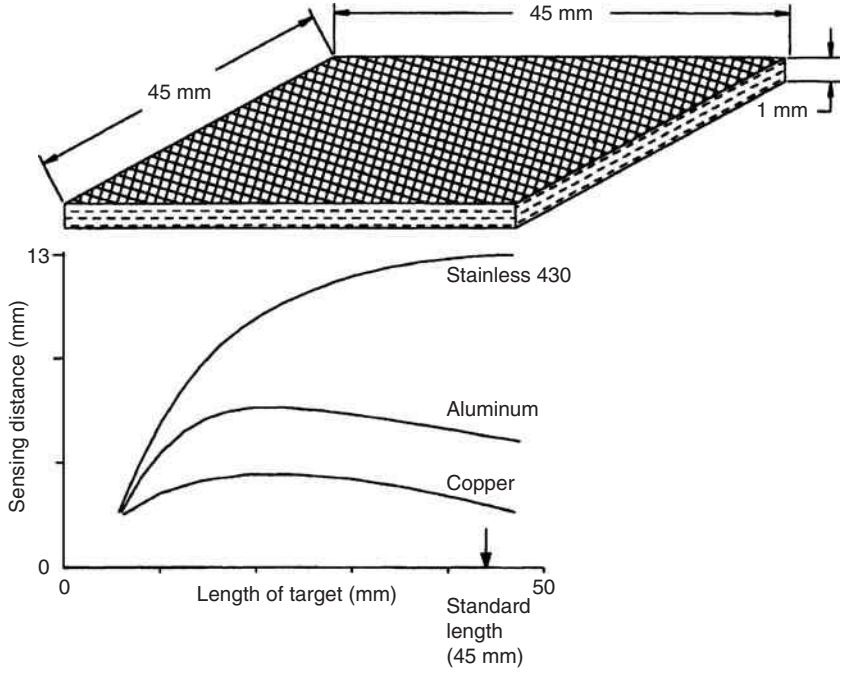


Figure 2.53 Sensing range correction factor.

This will determine the effective sensing range for a device sensing a stated target material of standard target size.

Example.

18-mm tubular extended range	8 mm
Copper target correction factor	× 0.35
Sensing range detecting copper standard target	2.80 mm

Target Material	Limit-Switch Type	Pancake Type	Tubular, mm			
			8	12	18	30
Steel (1020)	1.0	1.0	1.0	1.0	1.0	1.0
Stainless steel (400)	1.03	0.90	0.90	0.90	1.0	1.0
Stainless steel (300)	0.85	0.70	0.60	0.70	0.70	0.65
Brass	0.50	0.54	0.35	0.45	0.45	0.45
Aluminum	0.47	0.50	0.35	0.40	0.45	0.40
Copper	0.40	0.46	0.30	0.25	0.35	0.30

TABLE 2.6 Target Correction

2.6.5 Target Shape

Standard targets are assumed to be a flat square shape with the stated dimensions. Targets of round shape or with a pocketed surface must be of adequate dimensions to cause the necessary dampening effect on the sensor. Allowing the sensor-to-target distance less than the nominal range will help assure proper sensing. Also, using the next larger size or an extended-range sensor will help minimize problems with other-than-standard target dimensions or shapes. Figure 2.54 illustrates the axial (head-on) approach, indicating that the target approaches the face of the sensor on the axis of the coil core. When the target approaches axially, the sensor should not be located such that it becomes an end stop. If axial operation is considered, a good application practice is to allow for 25 percent overtravel.

Lateral (side by) approach means the target approaches the face of the sensor perpendicular to the axis of the coil core (Fig. 2.55). Good application practice (GAP), a term often used in "world-class" manufacturing strategies, dictates that the tip of the sensing field envelope

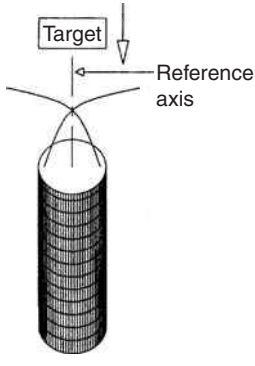


FIGURE 2.54 Axial approach.

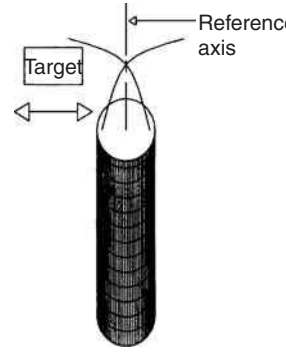


FIGURE 2.55 Lateral approach.

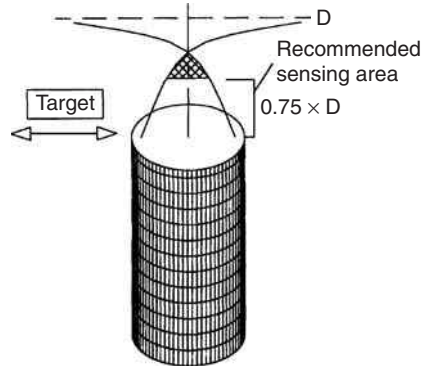


Figure 2.56 Lateral approach—recommended sensing distance.

should not be used. That is the point where sensing range variations start to occur. Therefore, it is recommended that the target pass not more than 75 percent of the sensing distance *D* from the sensor face. Also, the target should not pass any closer than the basic tolerance incorporated in the machine design in order to prevent damage to the sensor. Hysteresis can be greater for an axial approach (Fig. 2.56).

2.6.6 Variation Between Devices

Variations of sensing range between sensors of the same type often occur. With modern manufacturing technologies and techniques, variations are held to a minimum. The variations can be attributed to collective tolerance variations of the electrical components in the sensor circuit and to subtle differences in the manufacturing process from one device to the next; 5 percent variation is typical (Fig. 2.57).

Sensing distance also will vary from one temperature extreme to the other because of the effect of temperature change on the components of the sensor. Typical temperature ranges are –25°C (–3°F) to +70°C (+180°F). Figure 2.58 illustrates sensing range variations with temperature.

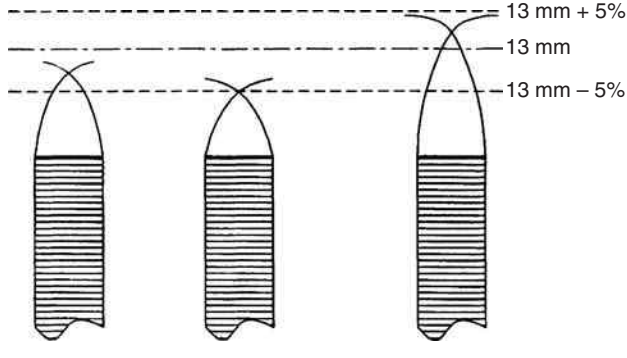


Figure 2.57 Sensing range variation.

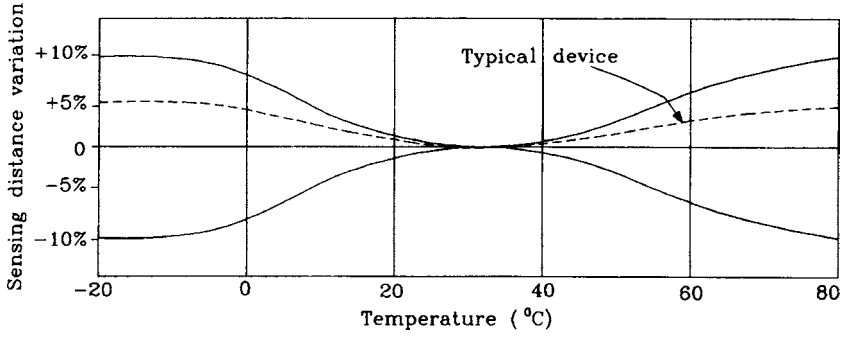


Figure 2.58 Sensing range variation—with temperature.

2.6.7 Surrounding Conditions

Several environmental factors must also be considered in order to obtain reliable information from inductive proximity sensors. These surrounding factors are:

- *Embeddable mounting.* The shielded sensor in Fig. 2.59 is often referred to as a *flush-mounted* sensor. Shielded sensors are not affected by the surrounding metal.
- *Flying metal chips.* A chip removed from metal during milling and drilling operations may affect the sensor performance depending on the size of the chip, its location on the sensing face, and type of material. In these applications, the sensor face should be oriented so that gravity will prevent chips from accumulating on the sensor face. If this is not possible, then coolant fluid should wash the sensor face to remove the chips. Generally, a chip does not have sufficient surface area to cause a sensor to turn on. If a chip lands on the center of the sensor face, it will have a negligible effect, but elsewhere on the sensor face, it will extend the range of the sensor.

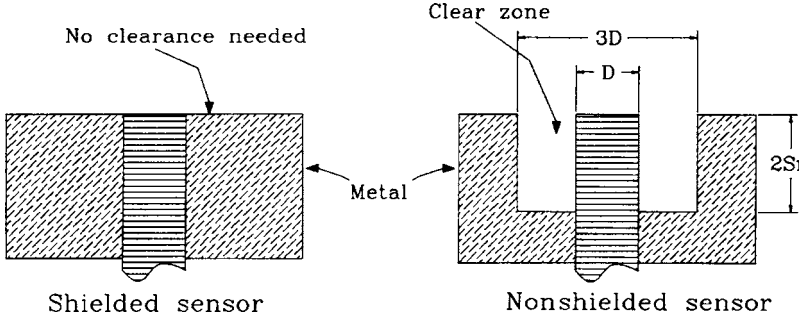


Figure 2.59 Embeddable and nonembeddable sensors.

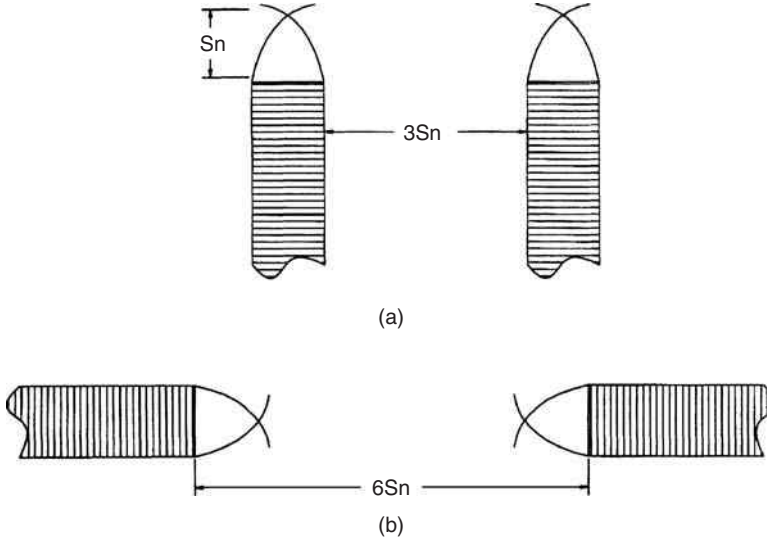


Figure 2.60 Adjacent sensors.

- *Adjacent sensors.* When two similar sensors are located adjacent to or opposite each other, the interaction of their fields can affect operation. Figure 2.60 provides the guidelines for placing two similar sensors adjacent to each other. *Alternate-frequency* heads will allow adjacent mounting of sensors without interaction of their sensing fields.
- *Magnetic fields.* Electrical wiring in the vicinity of the sensor face may affect sensor operation. If the magnetic field around the electrical wiring reaches an intensity that would saturate the sensor ferrite or coil, the sensor will not function properly. Use of inductive sensors in the presence of high-frequency radiation can also unexpectedly affect their operation. Sensors specially designed for welding applications can be used with programmable logic control (PLC). The PLC can be programmed to ignore the signal from the sensor for the period that the high-frequency welder is operated. A slight OFF-time delay assures proper operation of the sensor.
- *Radio-frequency interference (RFI).* Radio transceivers, often called *walkie-talkie* devices, can produce a signal that can cause an inductive proximity sensor to operate falsely. The radio transceiver produces a radio-frequency signal similar to the signal produced by the oscillator circuit of the sensor. The effect that RFI has on an inductive proximity switch can vary. The factors that determine this variation are as follows:

- *Distance between RFI source and the sensor.* Typically, inductive proximity switches are not affected by RFI when a transceiver is 1 ft away from the inductive switch. However, if closer than 1 ft, the switch may operate without a target present.
- *Signal frequency.* The signal frequency may be the determining factor that will cause a particular device to false-operate
- *Signal intensity.* Radio-frequency transceivers usually are portable devices with a power rating of 5 W maximum.
- *Inductive proximity package.* The sensor package construction may determine how well the device resists RFI.
- *Approach to the sensor.* A transceiver approaching the connecting cable of a switch may affect it at a greater distance than if it was brought closer to the sensing face. As RFI protection varies from device to device and manufacturer to manufacturer, most manufacturers have taken steps to provide the maximum protection against false operation due to RFI.

- *Showering arc.* Showering arc is the term applied to induced line current/voltage spikes. The spike is produced by the electrical arc on an electromechanical switch or contactor closure. The current spike is induced from lines connected to the electromechanical switch to the lines connected to the inductive proximity switch, if the lines are adjacent and parallel to one another. The result can be false operation of the inductive proximity switch. The spike intensity is determined by the level of induced voltage and the duration of the spike. Avoiding running cables for control devices in the same wiring channel as those for the contactor or similar leads may eliminate spikes. Most electrical code specifications require separation of control device leads from electromechanical switch and contractor leads.

2.7 Understanding Capacitive Proximity Sensors

2.7.1 Principles of Operation

A capacitive proximity sensor operates much like an inductive proximity sensor. However, the means of sensing is considerably different. Capacitive sensing is based on dielectric capacitance. *Capacitance* is the property of insulators to store an electric charge. A capacitor consists of two plates separated by an insulator, usually called a *dielectric*. When the switch is closed (Fig. 2.61) a charge is stored on the two plates.

The distance between the plates determines the ability of a capacitor to store a charge and can be calibrated as a function of stored charge to determine discrete ON and OFF switching status.

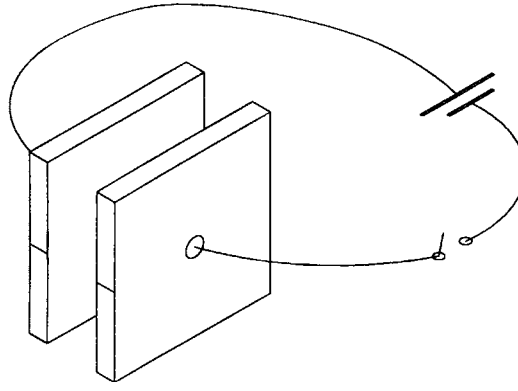


Figure 2.61 Capacitive principle.

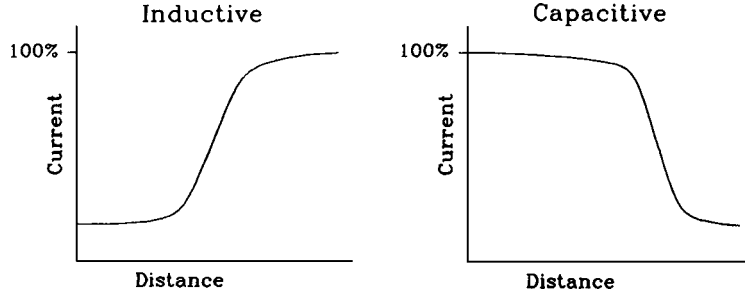


Figure 2.62 Capacitive sensor.

Figure 2.62 illustrates the principle as it applies to the capacitive sensor. One capacitive plate is part of the switch, the sensor face (the enclosure) is the insulator, and the target is the other plate. Ground is the common path.

The capacitive proximity sensor has the same four basic elements as an inductive proximity sensor:

- Sensor (the dielectric plate)
- Oscillator circuit
- Detector circuit
- Solid-state output circuit

The oscillator circuit in a capacitive switch operates like one in an inductive proximity switch. The oscillator circuit includes feedback capacitance from the external target plate and the internal plate. In a capacitive switch, the oscillator starts oscillating when sufficient

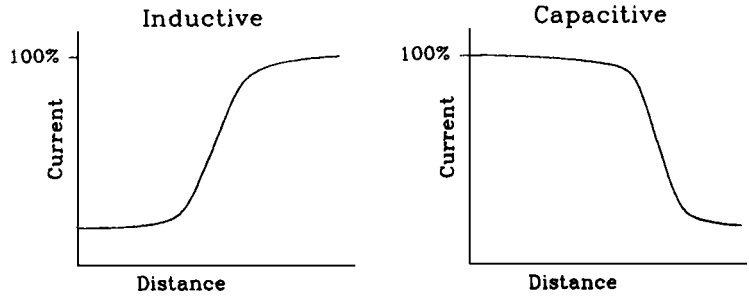


Figure 2.63 Oscillator damping of inductive and capacitive sensors.

feedback capacitance is detected. In an inductive proximity switch, the oscillation is damped when the target is present (Fig. 2.63).

In both capacitive and inductive switch types, the difference between the operate and the release amplitude in the oscillator and corresponding detector circuit is referred to as *hysteresis* of the sensor. It corresponds to the difference between target detection and release distances from the sensor face.

2.7.2 Features of Capacitive Sensors

The major characteristics of capacitive proximity sensors are:

- They can detect nonmetallic targets.
- They can detect lightweight or small objects that cannot be detected by mechanical limit switches.
- They provide a high switching rate for rapid response in object counting applications.
- They can detect liquid targets through nonmetallic barriers (glass, plastic, etc.).
- They have a long operational life with a virtually unlimited number of operating cycles.
- The solid-state output provides a bounce-free contact signal.

Capacitive proximity sensors have two major limitations.

- They are affected by moisture and humidity.
- They must have an extended range for effective sensing.

2.7.3 Sensing Range

Capacitive proximity sensors have a greater sensing range than inductive proximity sensors as shown in the following table.

Tubular Diameter, mm	Inductive Extended Range, mm	Capacitive Extended Range, mm
18	8	10
30	15	20
34	—	40

Sensing distance for capacitive proximity sensors is a matter of plate diameter, as coil size is for inductive proximity sensors. Capacitive sensors basically measure a dielectric gap. Accordingly, it is desirable to be able to compensate for target and application conditions with a sensitivity adjustment for the sensing range. Most capacitive proximity switches are equipped with a sensitivity adjustment potentiometer (Fig. 2.64).

2.7.4 Target Material and Size

The sensing range of capacitive sensors, like that of inductive proximity sensors, is determined by the type of material. Table 1.7 lists the

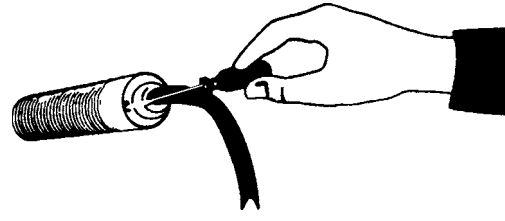


Figure 2.64 Sensitivity adjustment.

Material	Factor
Mild steel	1.0
Cast iron	1.0
Aluminum and copper	1.0
Stainless steel	1.0
Brass	1.0
Water	1.0
Polyvinylchloride (PVC)	0.5
Glass	0.5
Ceramic	0.4
Wood	≥0.2
Lubrication oil	0.1

Table 2.7 Target Material Correction

sensing-range derating factors that apply to capacitive proximity sensors. Capacitive sensors can be used to detect a target material through a nonmetallic interposing material like glass or plastic. This is beneficial in detecting a liquid through the wall of a plastic tank or through a glass sight tube. The transparent interposing material has no effect on sensing. For all practical purposes, the target size can be determined in the same manner as for inductive proximity sensors.

2.7.5 Surrounding Conditions

Capacitive proximity devices are affected by component tolerances and temperature variations. As with inductive devices, capacitive proximity devices are affected by the following surrounding conditions:

- *Embeddable mounting.* Capacitive sensors are generally treated as nonshielded nonembeddable devices.
- *Flying chips.* Capacitive devices are more sensitive to metallic and nonmetallic chips.
- *Adjacent sensors.* Allow more space than with inductive proximity devices because of the greater sensing range of capacitive devices.
- *Target background.* Relative humidity may cause a capacitive device to operate even when a target is not present. Also, the greater sensing range and ability to sense nonmetallic target materials dictate greater care in applying capacitive devices with target background conditions.
- *Magnetic fields.* Capacitive devices are not usually applied in welding environment.
- *Radio-frequency interference.* Capacitive sensor circuitry can be affected by RFI in the same way that an inductive device can.
- *Showering arc.* An induced electrical noise will affect the circuitry of a capacitive device in the same way it does an inductive device.

2.8 Understanding Limit Switches

A limit switch is constructed much like the ordinary light switch used in home and office. It has the same ON/OFF characteristics. The limit switch usually has a pressure-sensitive mechanical arm. When an object applies pressure on the mechanical arm, the switch circuit is energized. An object might have a magnet attached that causes a contact to rise and close when the object passes over the arm.

Limit switches can be either *normally open* (*NO*) or *normally closed* (*NC*) and may have multiple poles (Fig. 2.65). A normally open switch has continuity when pressure is applied and a contact is made, while

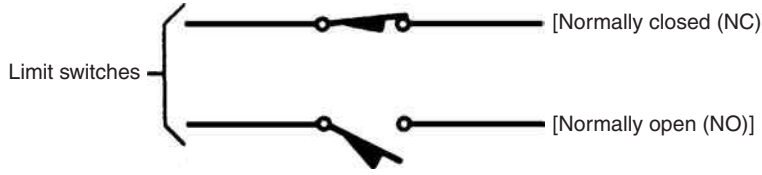


Figure 2.65 Normally open and normally closed microswitches.

a normally closed switch opens when pressure is applied and a contact is separated. A single-pole switch allows one circuit to be opened or closed upon switch contact, whereas a multiple-pole switch allows multiple circuits to be opened or closed.

Limit switches are mechanical devices. They have three potential problems:

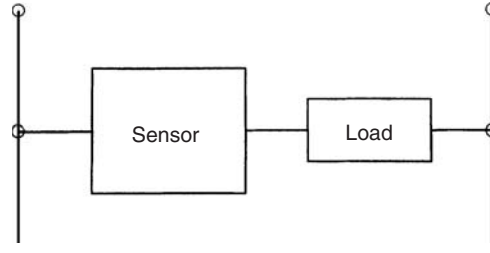
- They are subject to mechanical failure.
- Their mean time between failures (MTBFs) is low compared to non-contact sensors.
- Their speed of operation is relatively slow; the switching speed of photoelectric microsensors is up to 3000 times faster.

2.9 Inductive and Capacitive Sensors in Manufacturing

Inductive and capacitive proximity sensors interface to control circuits through an output circuit, for manufacturing applications. Also, the control circuit type is a determining factor in choosing an output circuit. Control circuits, whether powered by AC, DC, or AC/DC, can be categorized as either *load powered* or *line powered*.

The load-powered devices are similar to limit switches. They are connected in series with the controlled load. These devices have two connection points and are often referred to as *two-wire switches*. Operating current is drawn through the load. When the switch is not operated, the switch must draw a minimum operating current referred to as *residual current*. When the switch is operated or damped (i.e., a target is present), the current required to keep the sensor operating is the minimum holding current (Fig. 2.66). The residual current is not a

Figure 2.66 Load-powered residual current.



consideration for low-impedance loads such as relays and motor starters. However, high-impedance loads, most commonly programmable logic controllers, require residual current of less than 2 mA. Most sensors offer 1.7 mA or less.

In some manufacturing applications, a particular type of PLC will require less than 1.7 mA residual current. In such applications, a loading resistor is added in parallel to the input to the PLC load. Then, the minimum holding current may range up to 20 mA, depending on the sensor specification. If the load impedance is too high, there will not be enough load current level to sustain the switch state.

Inductive proximity sensors with a holding current of 4 mA or less can be considered low-holding-current sensors. These devices can be used with PLCs without concern for minimum holding current.

Line-powered devices derive current, usually called *burden current*, from the line and not through the controller load. These devices are called three-wire switches because they have three connections (Fig. 2.67).

The operating current for a three-wire sensor is burden current, and is typically 20 mA. Since the operating current does not pass through the load, it is not a major concern for the circuit design.

2.9.1 Relays

An output circuit relay is a mechanical switch available in a variety of contact configurations. Relays can handle load currents at high voltages, allowing the sensor to directly interface with motors, large solenoids, and other inductive loads. They can switch either AC or DC loads. Contact life depends on the load current and frequency of operation. Relays are subject to contact wear and resistance buildup. Because of contact bounce, they can produce erratic results with counters and programmable controllers unless the input is filtered. They can add 10 to 25 ms to an inductive or capacitive switch response time because of their mechanical nature (Fig. 2.68).

Relays are familiar to manufacturing personnel. They are often used with inductive or capacitive proximity sensors since they

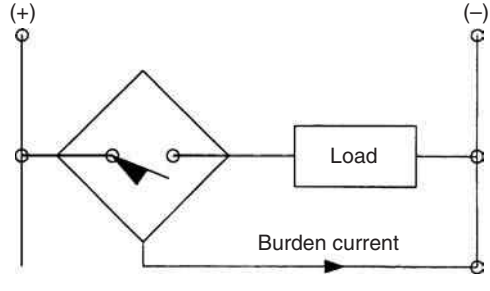


Figure 2.67 Line-powered burden current.

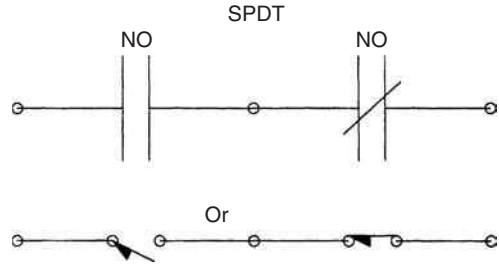


Figure 2.68 Relay output.

provide multiple contacts. The good and bad features of a relay are summarized in the following table.

Relay Advantages	Relay Disadvantages
Switches high currents/loads	Slow response time
Multiple contacts	Mechanical wear
Switches AC or DC voltages	Contact bounce
Tolerant of inrush current	Affected by shocks and vibration

2.9.2 Triac Devices

A triac is a solid-state device designed to control AC current (Fig. 2.69). Triac switches turn ON in less than a microsecond when the gate (control leg) is energized, and shut OFF at the zero crossing of the AC power cycle.

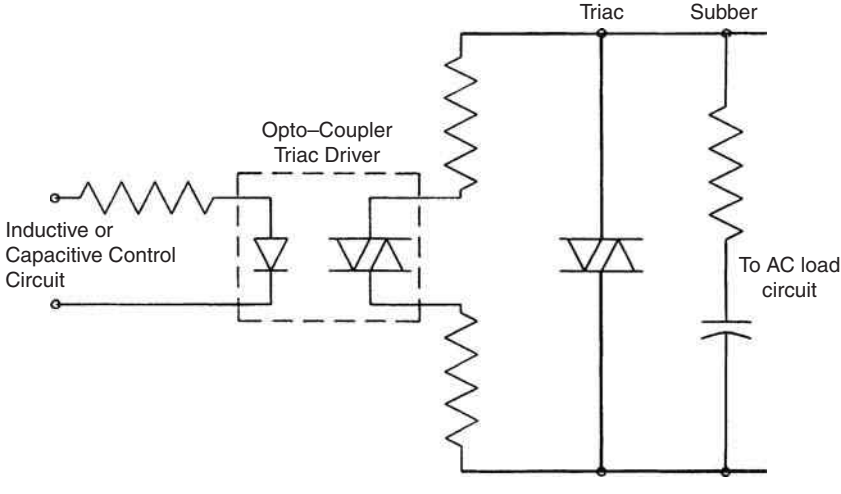


Figure 2.69 Triac circuit.

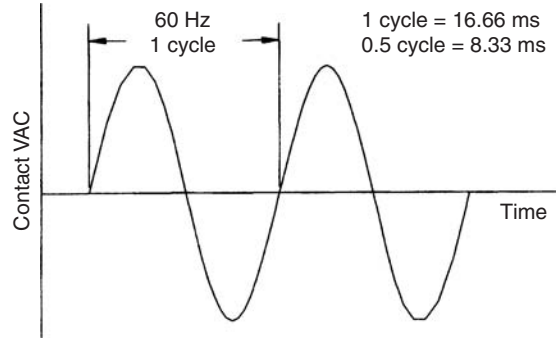


Figure 2.70 AC power cycle.

Because a triac is a solid-state device, it is not subject to the mechanical limitations of a relay such as mechanical bounce, pitting, corrosion of contacts, and shock and vibration. Switching response time is limited only by the time it takes the 60-Hz AC power to go through one half of a cycle (8.33 ms) (Fig. 2.70).

As long as a triac is used within its rated maximum current and voltage specifications, life expectancy is virtually infinite. Triac devices used with inductive or capacitive sensors generally are rated at 2-A loads or less. Triac limitations can be summarized as follows: (1) shorting the load will destroy a triac; and (2) directly connected inductive loads or large voltage spikes from other sources can false-trigger a triac.

To reduce the effect of these spikes, a snubber circuit composed of a resistor and capacitor in series is connected across the device. Depending on the maximum switching load, an appropriate snubber network for switch protection is used. The snubber network contributes to the OFF state leakage to the load. The leakage must be considered when loads requiring little current, such as PLCs, are switched. In the ON state, a drop of about 1.7 V rms is common (Fig. 2.71). Good and bad features of triacs are listed in the following table.

Triac Advantages	Triac Disadvantages
Fast response time (8.33 ms)	Can be falsely triggered by large inductive current
Tolerant of large inrush currents	Snubber contributes to OFF state leakage current
Can be directly interfaced with programmable controllers	Can be destroyed by short circuits
Infinite life when operated within rated voltage/current limits	

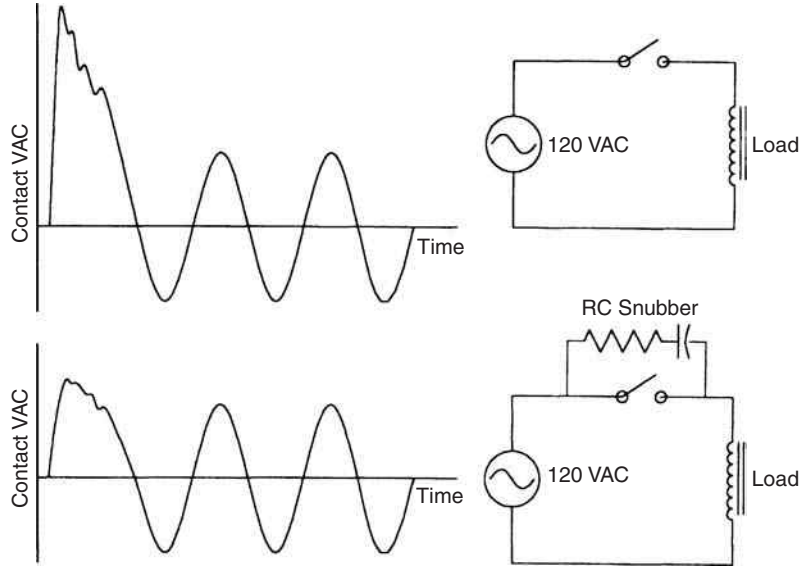


Figure 2.71 Snubber circuit.

2.9.3 Transistor DC Switches

Transistors are solid-state DC switching devices. They are most commonly used with low-voltage DC-powered inductive and capacitive sensors as the output switch. Two types are employed, depending on the function (Fig. 2.72).

In an NPN transistor, the current source provides a contact closure to the DC positive rail. The NPN current sink provides a contact to the DC common. The transistor can be thought of as a single-pole switch that must be operated within its voltage and maximum current ratings (Fig. 2.73).

Any short circuit on the load will immediately destroy a transistor that is not short-circuit protected. Switching inductive loads creates voltage spikes that exceed many times the maximum rating of

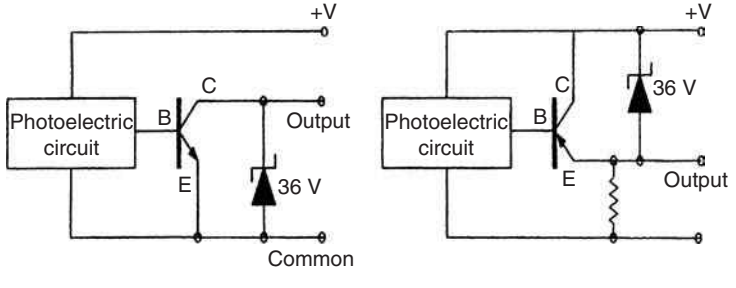


Figure 2.72 DC circuit logic.

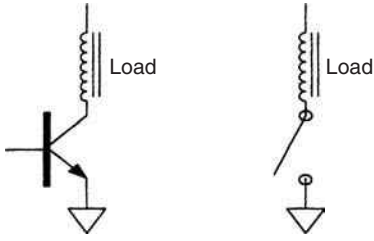


Figure 2.73 Transistor switch.

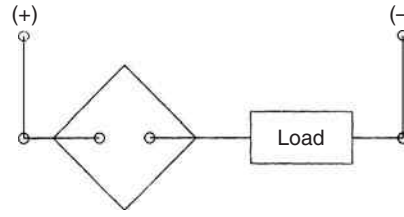


Figure 2.74 Voltage clamp.

the transistor. Peak voltage clamps such as zener diodes or transorbs are utilized to protect the output device. Transistor outputs are typically rated to switch loads of 250 mA at 30 V DC maximum (Fig. 2.74).

The following outlines the advantages and disadvantages of transistors.

Transistor Advantages	Transistor Disadvantages
Virtually instantaneous response	Low current handling capacity
Low OFF state leakage and voltage drop	Cannot handle inrush current unless clamped
Infinite life when operated within rated current/voltage	Can be destroyed by short circuit unless protected
Not affected by shock and vibration	

2.9.3.1 Output Configuration

Output configurations are categorized as follows:

- Single output—normally open (NO)
- Single output—normally closed (NC)
- Programmable output—NO or NC
- Complementary output—NO and NC

The functions of normally open and normally closed output types are defined in Table 1.8.

2.9.4 Inductive and Capacitive Control/Output Circuits

A single output sensor has either an NO or an NC configuration and cannot be changed to the other configuration (Fig. 2.75).

A programmable output sensor has one output, NO or NC, depending on how the output is wired when installed. These sensors are exclusively two-wire AC or DC (Fig. 2.76).

Output Configuration	Target State	Oscillator State	Output
NO	Absent	Undamped	Nonconducting (OFF)
	Present	Damped	Conducting (ON)
NC	Absent	Undamped	Conducting (ON)
	Present	Damped	Nonconducting (OFF)

Table 2.8 Output Logic

Figure 2.75 Single output.

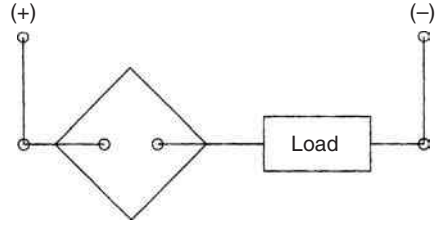
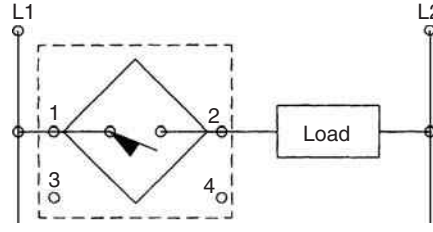


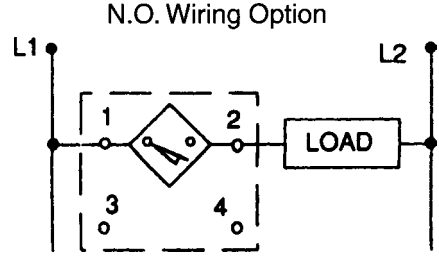
Figure 2.76 Programmable output.



A complementary output sensor has two outputs, one NO and one NC. Both outputs change state simultaneously when the target enters or leaves the sensing field. These sensors are exclusively three-wire AC or DC (Fig. 2.77).

The choice of control circuit and output logic plays an important part in determining the reliability of data collection. The choice of control circuit and output logic depends on the following parameters:

Figure 2.77 Complementary output.



- *AC or DC control voltage.* Use of AC control may seem to require the use of an AC-configured sensor. However, interface circuitry can allow for DC sensors even if the main control voltage source is AC.
- *Control circuit current requirements.* Usually control circuits operating in the 200- to 300-mA range can use either AC or DC sensors. Circuits with 0.5-A and higher current will dictate the type of sensor to be used.
- *Application output requirements.* NO output is the most commonly used output type. Controlled circuit configurations may dictate use of NC or complementary-type configured sensors.
- *Switching speed requirements.* AC circuits are limited in their operations per second. DC circuits may be required for applications involving counting or high speed.
- *Connecting logic device.* The device to which the sensor is connected—such as programmable controller, relay, solenoid, or timer/counter—is usually the most important factor in sensor circuit and output configuration.

2.9.5 Accessories for Sensor Circuits

Sensor circuits and their output configurations must have various types of indicators and protection devices, such as:

- Light-emitting diode (LED) indicators
- Short-circuit protectors
- Reverse-polarity protectors—DC three-wire
- Wire terminators—color-coded wire
- Pin connector type and pin-out designator

2.9.5.1 LED Indicators

LED indicators provide diagnostic information on the status of sensors (e.g., operated or not operated) that is vital in computer-integrated manufacturing. Two LEDs also indicate the status of complementary-type sensor switches and power ON/OFF status, as well as short-circuit condition.

2.9.5.2 Short-Circuit Protection

Short-circuit protection is intended to protect the switch circuit from excessive current caused by wiring short circuits, line power spikes from high inrush sources, or lightning strikes. This option involves special circuitry that either limits the current through the output device or turns the switch OFF. The turn off–type switch remains inoperative until the short circuit has been cleared—with power disconnected. Afterwards, power is reapplied to the sensor. A second

LED is usually furnished with this type of device to indicate the shorted condition.

2.9.5.3 Reverse-Polarity Protection

Reverse-polarity protection is special circuitry that prevents damage in a three-wire DC sourcing (PNP) or sinking (NPN) device when it is connected to control circuitry incorrectly. Although reverse polarity is relatively common, not all switches are equipped with this option.

2.9.5.4 Wire Termination

Wire terminals are common on limit-switch enclosure-type sensors. The terminal designations are numbered and correspond to the device wiring diagram (Fig. 2.78). Cable/wire stub terminations are most common on tubular sensors. Color-coded conductors are essential for correct wiring. Most sensor wires are color-coded to comply with industry wire color-code standards.

2.9.5.5 Pin Connectors

Pin-connector-terminal sensors feature a male pin connector receptacle on the switch or at the end of the wire/cable stub. The female receptacle is at the end of the matching cable cord. Most industry-standard pin connectors are either the mini type—approximately 18 mm in diameter—or the micro type—approximately 12 mm in diameter (Fig. 2.79).

2.9.6 Inductive and Capacitive Switching Logic

The outputs of two or more inductive or capacitive proximity sensors can be wired together in series or parallel to perform logic functions. The ON, or activated, switch function can be either a normally open or a normally closed output function, depending on the desired control logic.

Although sensors are the most effective means for the data acquisition role in manufacturing, care must be exercised when sensors are

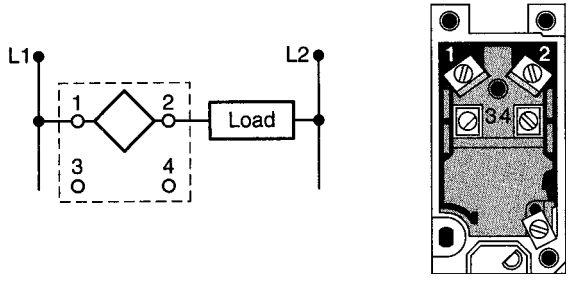


Figure 2.78 Wire terminal.

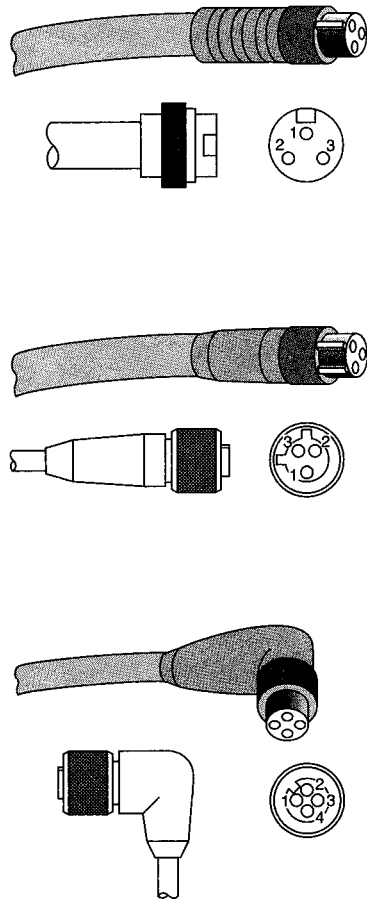


FIGURE 2.79
Pin connector.

integrated with various production operations. The following factors will affect the performance of the switch logic circuit:

- Excessive leakage current in parallel-connected load-powered devices
- Excessive voltage drop in series-connected devices
- Inductive feedback with line-powered sensors with parallel connections

2.9.6.1 Parallel-Connection Logic-OR Function

The binary OR logic in Table 1.9 indicates that the circuit output is ON (1) if one or more of the sensors in parallel connection is ON.

0 = OFF

1 = ON

A	B	C	Out
0	0	0	0
0	0	1	1
0	1	0	1
1	0	0	1

TABLE 2.9 Binary Logic Chart–Parallel Connection

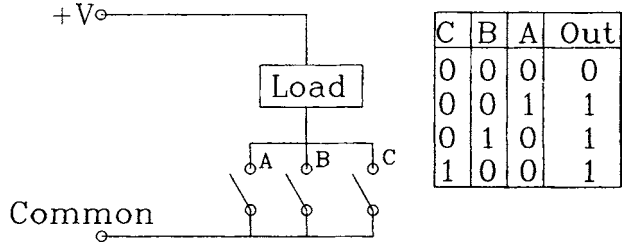


FIGURE 2.80 Parallel sensor arrangement.

It is important to note that, in two-wire devices, the OFF state residual current is additive (Fig. 2.80). If the circuit is affected by the total leakage applied, a shunt (loading) resistor may have to be applied. This is a problem in switching to a programmable controller or other high-impedance device.

Example.

$$I_a + I_b + I_c = I_t$$

$$1.7 + 1.7 + 1.7 = 5.1 \text{ mA}$$

Three-wire 10 to 30 V can also be connected in parallel for a logic OR circuit configuration. Figure 2.81 shows a current *sourcing* (PNP) parallel connection.

Figure 2.82 displays a current *sinking* (NPN) parallel connection. It may be necessary to utilize blocking diodes to prevent inductive

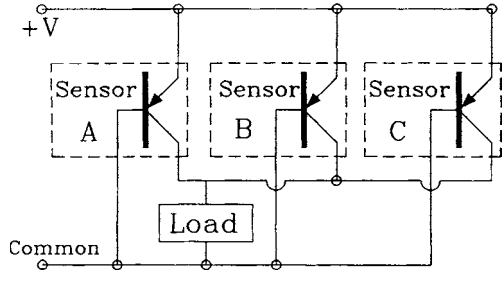


FIGURE 2.81 Sourcing (PNP) parallel sensor arrangement.

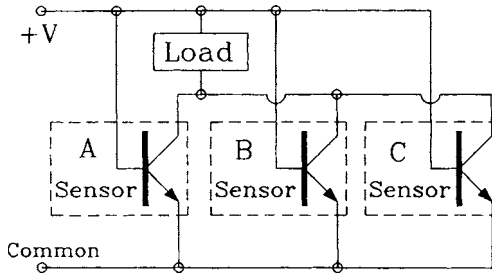


Figure 2.82 Sinking (NPN) parallel sensor arrangement.

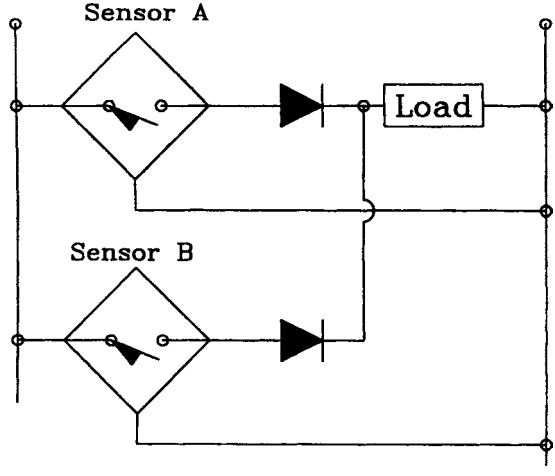


Figure 2.83 Blocking diodes.

feedback (or reverse polarity) when one of the sensors in parallel is damped while the other is undamped. Figure 2.83 demonstrates the use of blocking diodes in this type of parallel connection.

2.9.6.2 Series-Connection Logic—AND Function

Figure 2.84 shows AND function logic indicating that the series-connected devices must be ON (1) in order for the series-connected circuit to be ON.

B	A	Out
0	0	0
0	1	0
1	0	0
1	1	1

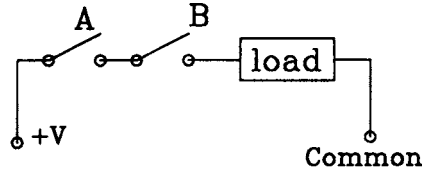


Figure 2.84 Series AND logic.

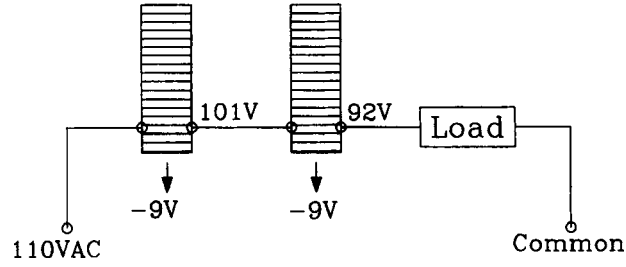


Figure 2.85 Series connected, load powered.

The voltage drop across each device in series will reduce the available voltage the load will receive. Sensors, as a general rule, have a 7-to-9-V drop per device. The minimum operating voltage of the circuit and the sum of the voltage drop per sensor will determine the number of devices in a series-connected circuit. Figure 2.85 shows a typical two-wire AC series-connected circuit.

Series connection is generally applied to two-wire devices, most commonly two-wire AC. A 10 to 30-V DC two-wire connection is not usually practical for series connection because of the voltage drop per device and minimum operating voltage. Three-wire devices are generally not used for series connection. However, the following characteristics should be considered for three-wire series-connected circuits (Fig. 2.86):

- Each sensor must carry the load current and the burden current for all the downstream sensors (Fig. 2.86).
- When conducting, each sensor will have a voltage drop in series with the load, reducing the available voltage to the load. As with two-wire devices, this and the minimum operating voltage will limit the number of devices wired in series.

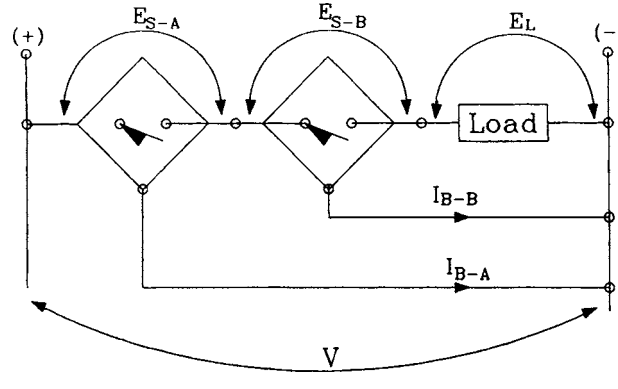


Figure 2.86 Series connected, line powered.

- When upstream sensors are not conducting, the downstream sensors are disconnected from their power source and are incapable of responding to a target until the upstream sensors are activated (damped). Time before availability will be increased due to the response in series.

Series and parallel connections that perform logic functions with connection to a PLC are not common practice. Utilizing sensors this way involves the preceding considerations. It is usually easier to connect directly to the PLC inputs and perform the desired logic function through the PLC program.

2.9.7 Inductive and Capacitive Sensor Response Time—Speed of Operation

When a sensor receives initial power on system power-up, the sensor cannot operate. The sensor operates only after a delay called *time delay before availability* (Fig. 2.87).

In AC sensors, this delay is typically 35 ms. It can be as high as 100 ms in AC circuits with very low residual current and high noise immunity. In DC sensors, the time delay is typically 30 ms.

2.9.7.1 Response and Release Time

A target entering the sensing field of either an inductive or a capacitive sensor will cause the detector circuit to change state and initiate an output. This process takes a certain amount of time, called *response time* (Fig. 2.88).

Response time for an AC sensor is typically less than 10 ms. DC devices respond in microseconds. Similarly, when a target leaves

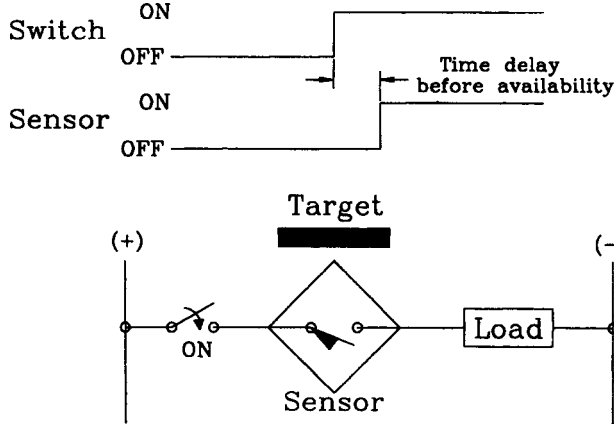


Figure 2.87 Time delay prior to availability.

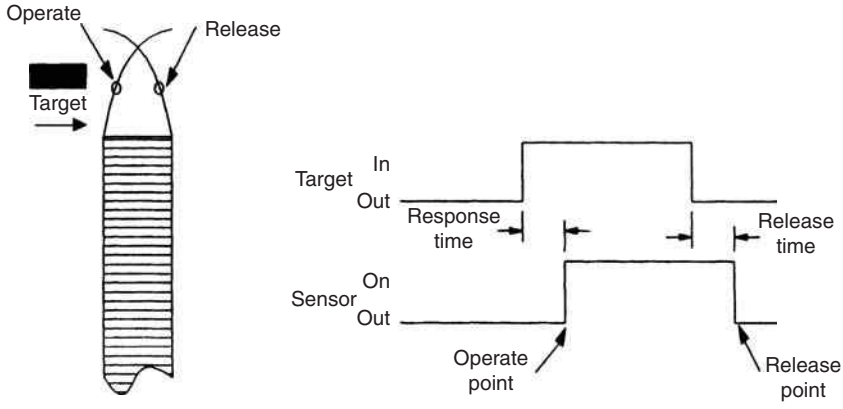


Figure 2.88 Response/release times.

the sensing field, there is a slight delay before the switch restores to the OFF state. This is the *release time*. Release time for an AC device is typically one cycle (16.66 ms). The DC device release time is typically 3 ms.

2.9.7.2 High-Speed Operation

Mechanical devices such as limit switches and relays do not operate at speeds suitable for high-speed counting or other fast-operating-circuit needs. Solid-state devices, however, can operate at speeds of 10, 15, or more operations per second. DC devices can operate at speeds of 500, 1000, or more operations per second.

In order to properly achieve high-speed operation, some basic principles must be applied.

2.9.7.3 Maximum Target Length

A response delay occurs when a sensor has a target entering the sensing field, as previously stated. A similar delay results in order for the respective load to operate. The time between when the sensor conducts and the load operates is the *load response time*. Together these delays make up the *system response time* T_o.

Similarly, delays occur when the target reaches the release point in the sensing field caused by the sensor release time and the corresponding *load release time*. In order to ensure that the sensor will operate reliably and repeatedly, the target must stay in the field long enough to allow the load to respond. This is called *dwell time*. Figure 2.89 illustrates the time functions for reliable repeatable sensor operation. Figure 2.90 illustrates the dwell range.

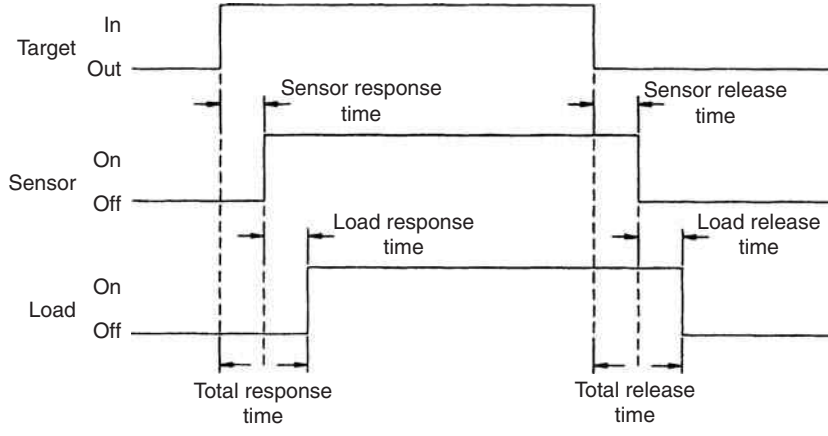
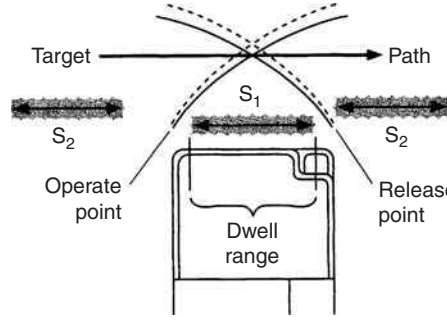


Figure 2.89 Maximum target length.

Figure 2.90
Dwell range.



2.9.7.4 Target Duty Cycle

Response (turn-on) times for the sensor and the controlled load may be considerably different from the release (turn-off) times for the same devices. Conditions for the target duty cycles are illustrated in Fig. 2.91. Note that the target is not out of the sensing field long enough to allow the sensor to turn off the controlled load. The application

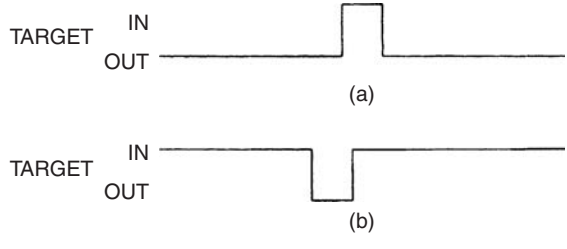


Figure 2.91 Target duty cycle. (*a*) Critical response time (turn-on). (*b*) critical release time (turn-off).

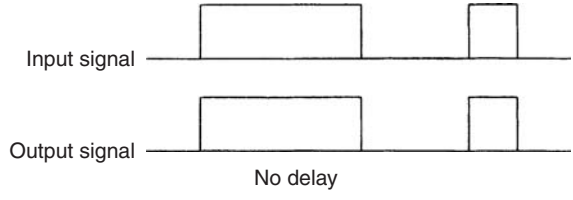


Figure 2.92 No delay.

must be arranged so that both sensor and load turn ON and OFF reliably and repeatedly.

2.9.7.5 Timing Functions

When an inductive control is operating a logic function, an output is generated for the length of time an object is detected (Fig. 2.92).

2.9.7.6 ON Delay Logic

ON delay logic allows the output signal to turn on only after the object has been detected for a predetermined period of time. The output will turn off immediately after the object is no longer detected. This logic is useful if a sensor must avoid false interruption from a small object. ON delay is useful in bin fill or jam detection, since it will not false-trigger in the normal flow of objects going past (Fig. 2.93).

2.9.7.7 OFF Delay Logic

OFF delay logic holds the output on for a predetermined period of time after an object is no longer detected. The output is turned on as soon as the object is detected. OFF delay ensures that the output will not drop out despite a short period of signal loss. If an object is once again detected before the output time out, the signal will remain ON. OFF delay logic is useful in applications susceptible to periodic signal loss (Fig. 2.94).

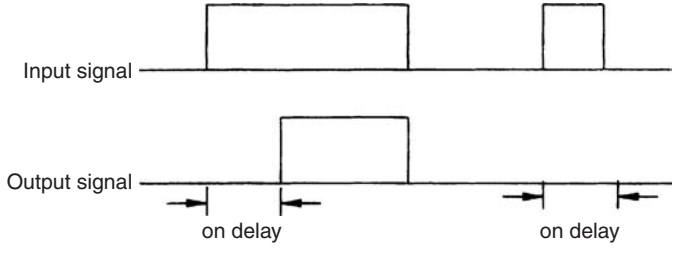


Figure 2.93 ON delay.

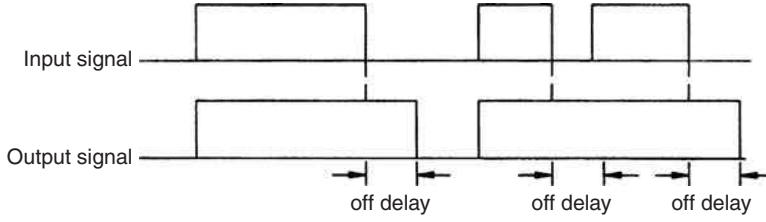


Figure 2.94 OFF delay.

2.9.7.8 ON/OFF Delay Logic

ON/OFF delay logic combines ON and OFF delay so that the output will be generated only after the object has been detected for a predetermined period of time, and will drop out only after the object is no longer detected for a predetermined period of time. Combining ON and OFF delay smoothes the output of the inductive proximity control (Fig. 2.95).

2.9.7.9 One-Shot Delay

One-shot logic generates an output of predetermined length no matter how long an object is detected. A standard one-shot must time out before it can be retriggered. One-shot logic is useful in applications that require an output of specified length (Fig. 2.96).

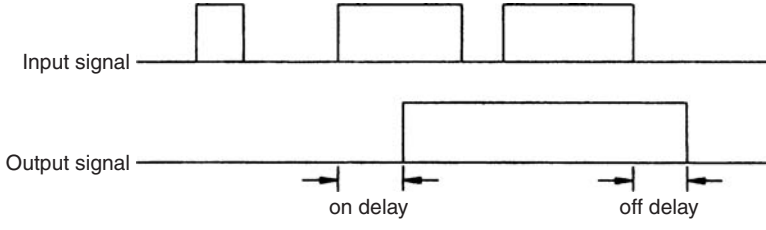


Figure 2.95 ON/OFF delay.

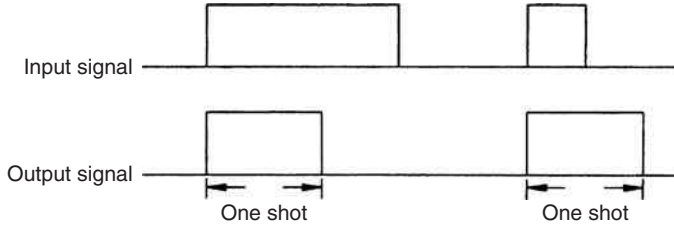


Figure 2.96 One-shot.

2.10 Understanding Microwave Sensing Applications

Microwave sensors are valuable tools in the industrial environment for measuring motion, velocity, direction of movement, and range. They are rugged devices capable of operating in hostile environments, and are intrinsically safe, since they have no moving parts and require low power. They will not cause any harm to operators and will function effectively in explosive environments. They can successfully measure large military and industrial objects over large distances and can provide a great deal of information about these objects, as observed during the Persian Gulf War in 1991, Iraq war in 2003, and current Afghanistan war on terror.

Microwave technology has long been an effective method of measuring the parameters of motion and presence. Applications range from simple intrusion alarms that merely indicate an object has entered its field of view to complex military radar systems that define the existence, location, and direction of motion in military maneuvers.

Microwave sensing technology can be classified into five categories:

- *Motion sensing.* Sensing a moving object in a defined domain—for example, detecting an intruder in a prohibited area.
- *Presence sensing.* Sensing that an object exists in a defined domain at a given time. This concept is vital in industrial control systems where the arrival of an object may not be noticed.
- *Velocity sensing.* Sensing the linear speed of an object in a specified direction. This concept is used by police to detect speeding cars.
- *Direction-of-motion sensing.* Determining whether a target is moving away from or toward the microwave sensor device. This concept is particularly important for manufacturers of automated guided vehicle systems for obstacle avoidance. It is also used to detect whether objects or personnel are approaching or departing from automatic doors.
- *Range sensing.* Measuring the distance from the sensor to an object of interest. Applications include sensing the level of oil or chemical solutions in tanks and containers.

2.10.1 Characteristics of Microwave Sensors

Microwave sensors' general characteristics important in industrial and commercial applications are:

- *No contact.* Microwave sensors operate without actually contacting the object. This is particularly important if the object is in a hostile environment or sensitive to wear. They can

monitor the speed of power-plant generator shafts, continuously monitoring acceleration and deceleration in order to maintain a constant rotational speed. Microwave sensors can effectively penetrate nonmetallic surfaces, such as fiberglass tanks, to detect liquid levels. They can also detect objects in packaged cartons.

- *Rugged.* Microwave sensors have no moving parts and have proven their reliability in extensive military use. They are packaged in sealed industrial enclosures to endure the rigors of the production environment.
- *Environmental reliability.* Microwave sensors operate reliably in harsh inhospitable environments. They can operate from –55°C to +125°C in dusty, dirty, gusty, polluted, and poisonous areas.
- *Intrinsically safe.* Industrial microwave sensors can be operated in an explosive atmosphere because they do not generate sparks due to friction or electrostatic discharge. Microwave energy is so low that it presents no concern about hazard in industrial applications.
- *Long range.* Microwave sensors are capable of detecting objects at distances of 25 to 45,000 mm or greater, depending on the target size, microwave power available, and the antenna design.
- *Size of microwave sensors.* Microwave sensors are larger than inductive, capacitive, and limit switch sensors. However, use of higher microwave frequencies and advances in microwave circuit development allow the overall package to be significantly smaller and less costly.
- *Target size.* Microwave sensors are better suited to detect large objects than smaller ones, such as a single grain of sand.

2.10.2 Principles of Operation

Microwave sensors consist of three major parts: (1) transmission source, (2) focusing antenna, and (3) signal processing receiver.

Usually the transmission and receiver are combined together in one module, which is called a *transceiver*. A typical module of this type is used by intrusion alarm manufacturers for an indoor alarm system. The transceiver contains a Gunn diode mounted in a small precession cavity that, upon application of power, oscillates at microwave frequencies. A special cavity design will cause this oscillation to occur at 10.525 GHz, which is one of the few frequencies that the U.S. Federal Communications Commission (FCC) has set aside for motion detectors. Some of this energy is coupled through an iris into an adjoining waveguide. Power output is in the 10- to 20-mW range.

The DC input power for this stage (8 V at 150 mA) should be well-regulated, since the oscillator is voltage-sensitive. The sensitivity of the system can be significantly reduced by noise (interference).

At the end of the waveguide assembly, a flange is fastened to the antenna. The antenna focuses the microwave energy into a beam, the characteristics of which are determined by the application. Antennas are specified by beam width or gain. The higher the gain, the longer the range and the narrower the beam. An intrusion alarm protecting a certain domain would require a wide-beam antenna to cover the area, while a traffic control microwave sensor would require a narrow-beam high-gain antenna to focus down the road.

Regardless of the antenna selection, when the beam of microwave energy strikes an object, some of the microwave energy is reflected back to the module. The amount of energy will depend on the composition and shape of the target. Metallic surfaces will reflect a great deal, while Styrofoam and plastic will be virtually transparent. A large target area will also reflect more than a small one.

The reflected power measured at the receiver decreases by the fourth power of the distance to the target. This relationship must be taken into consideration when choosing the transmitted power, antenna gain, and signal processing circuitry for a specific application.

When the reflected energy returns to the transceiver, the mixer diode will combine it with a portion of the transmitted signal. If the target is moving toward or away from the module, the phase relationships of these two signals will change and the signal out of the mixer will be an audio frequency proportional to the speed of the target. This is called the *Doppler frequency*. This is of primary concern in measuring velocity and direction of motion. If the target is moving across in front of the module, there will not be a Doppler frequency, but there will be sufficient change in the mixer output to allow the signal processing circuitry to detect it as unqualified motion in the field.

The signal from the mixer will be in the microvolt to millivolt range so that amplification will be needed to provide a useful level. This amplification should also include 60-Hz and 120-Hz notch filters to eliminate interference from power lines and fluorescent light fixtures, respectively. The remaining bandwidth should be tailored to the application.

Besides amplification, a comparator and output circuitry relays are added to suit the application (Fig. 2.97).

2.10.3 Detecting Motion with Microwave Sensors

The presence of an object in the microwave field disturbs the radiated field. There may be a Doppler frequency associated with the disturbance. The signal from the mixer to the signal processing circuitry may vary with a large amplitude and long duration so it can be detected. The amplitude gain and the delay period are of specific importance in tailoring the device for particular application, such as

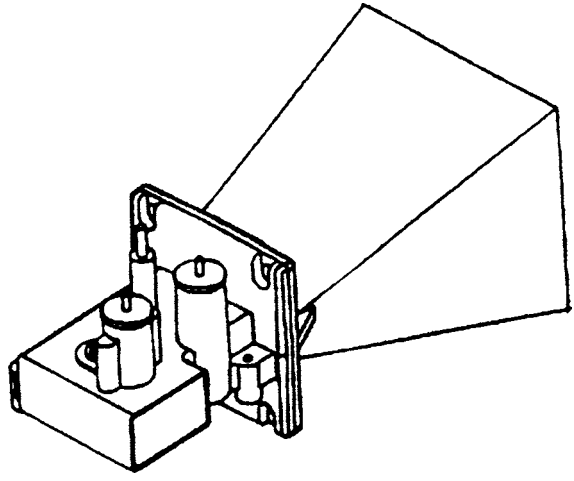


Figure 2.97 Typical microwave motion sensor module.

motion detection. These sensors are primarily used in intrusion alarm applications where it is only necessary to detect the movement rather than derive further information about the intruder. The sensitivity would be set to the minimum necessary level to detect a person-sized object moving in the protected domain in order to prevent pets or other nonhostile moving objects from causing a false alarm. In addition, some response delay would be introduced for the same reason, requiring continuous movement for some short period of time.

Other applications include parts counting on conveyer belts; serial object counting in general; mold ejection monitoring, particularly in hostile environments; obstacle avoidance in automated guided vehicle systems; fill indication in tanks; and invisible protection screens. In general, this type of sensor is useful where the objects to be sensed are moving in the field of interest (Fig. 2.98).

Other devices that compete for the same applications are ultrasonic, photoelectric, and infrared sensors.

In the intrusion alarm manufacturing industry, microwave sensors have the advantages of longer range and insensitivity to certain environmental conditions. Ultrasonic sensors are sensitive to drafts and high-frequency ambient noise caused by bells and steam escaping from radiators. Infrared sensors are sensitive to thermal gradients caused by lights turning on and off. The effectiveness of infrared sensors is severely reduced at high ambient temperatures. However, utilizing dual technologies is recommended to minimize false alarms—combining microwave technology with infrared technology, for example. It is necessary for the intruder to be sensed by both technologies before an alarm is given. In other applications, microwave sensors can show advantages over photoelectric sensors in the areas

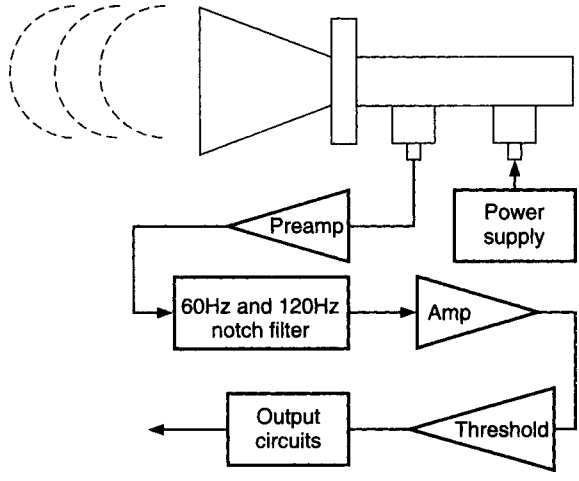


Figure 2.98 Microwave motion sensor.

of longer range, increased area of coverage, operation in hostile environments, and in applications where it is necessary to see through one medium (such as a cardboard box or the side of a nonmetallic tank) to sense the object on the other side.

If the target is moving toward or away from the transceiver, there will be an audio-frequency (Doppler) signal out of the mixer diode that is proportional to the velocity of the target. The frequency of this signal is given by the formula:

$$F_d = 2V\,(F_t/c)$$

where F_d = Doppler frequency
V = velocity of the target
F_t = transmitted microwave frequency
c = speed of light

If the transmitted frequency is 10.525 GHz (the motion detector frequency), this equation simplifies to:

$$F_d = 31.366 \text{ Hz} \times V \quad \text{in miles/hour}$$

or

$$F_d = 19.490 \text{ kHz} \times V \quad \text{in kilometers/hour}$$

or

$$F_d = 84.313 \text{ kHz} \times V \quad \text{in furlongs/fortnight}$$

This assumes that the target is traveling directly at or away from the transceiver. If there is an angle involved, then the equation becomes

$$F_d = 2V(F_t/c)\cos\Theta$$

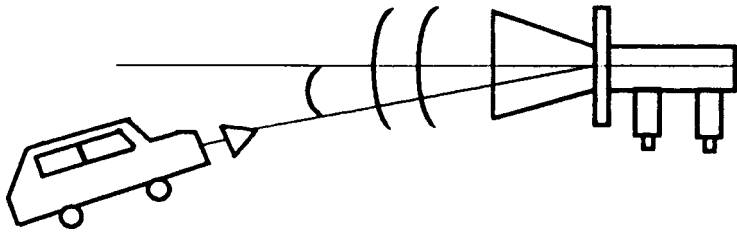


Figure 2.99 Angular velocity measurement.

where Θ is the angle between the transceiver and the line of movement of the target. Evidently, as the target is moving across the face of the transceiver, cos Θ = 0, and the frequency is 0. If the angle is kept below 18°, however, the measured frequency will be within 5 percent of the center frequency (Fig. 2.99).

Signal processing for this module must include amplification, a comparison network to shape the signal into logic levels, and a timing and counting circuit to either drive a display device or compare the frequency to certain limits. If more than one moving object is in the microwave field, it may be necessary to discriminate on the basis of amplitude or frequency bandwidth, limiting it to exclude unwanted frequencies. Velocities near 3 km/h and 6 km/h are also difficult to measure with this system since the corresponding Doppler frequencies are 60 and 120 Hz, which are prime interference frequencies from power lines and fluorescent fixtures. Extra shielding or isolation will be necessary in this case. False alarm rate may also be reduced by counting a specific number of cycles before triggering an output. This will actually correspond to the target moving a defined distance.

Microwave sensors are well-suited for measuring the velocity of objects, which most other sensors cannot do directly. Inductive, photoelectric, and other sensors can measure radial velocity. For example, inductive photoelectric sensors measure radial velocity when configured as a tachometer, and if the rotating element is configured as a trailing wheel, then linear velocity can be defined. Photoelectric sensors can also be set up with appropriate signal processing to measure the time that a moving object takes to break two consecutive beams. This restricts the measurement to a specific location. Multiple beams would be needed to measure velocity over a distance, whereas a single microwave sensor could accomplish the same result.

Aside from their use in police radars, microwave sensors can measure the speed of baseball pitches. These sensors have many industrial applications as well. Microwave sensors are an excellent means of closed-loop speed control of a relatively high-speed rotating shaft (3600 r/min). Other applications include autonomous-vehicle speed monitoring and independent safety monitoring equipment

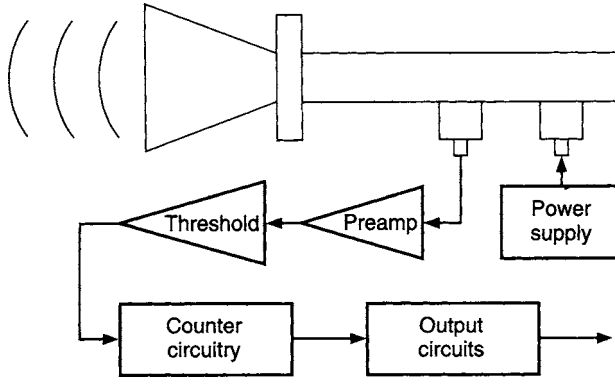


Figure 2.100 Velocity sensing module.

for heavy and high-speed machine tools. Also, a microwave sensor will detect an overvelocity condition (Fig. 2.100).

A microwave sensor, mounted on a tractor or other farm equipment to measure ground speed, will play an important role in reducing excessive distribution of seeds and fertilizer per acre. The ordinary wheel driven–speedometer is not sufficiently accurate because of wheel slippage. Accurate speed measurement is necessary in these vehicles so seeds and fertilizer are spread at a specific rate by the accessory equipment. An over- or underestimate of the speed will result in the wrong density per acre.

2.10.4 Detecting Presence with Microwave Sensors

A static object can be detected in the field of a microwave sensor. The purpose of this detection is to determine that the object is still in the field of interest and has not departed. This is particularly desirous in control systems where the controller is performing other tasks and then accesses the sensor to determine whether there is a sensed object at that particular time. In this situation, presence sensing is especially advantageous since the output can be verified by further interrogations to eliminate false sensing.

To detect the presence of an object, a microwave sensor with a separate transmitter and receiver must be used. A transceiver in this application is not adequate, although the transmitter and the receiver can be mounted in the same enclosure. The receiver must not sense any energy unless the object is present in the field. A means to modulate the transmitter is needed, and the receiver should be narrowband to amplify and detect the modulated reflection. The sensitivity of the receiver must be adjustable to allow for ambient reflections.

Microwave sensors have been extensively and successfully tested at various fast-food drive-through vending locations. Other types of sensors, such as ultrasonic and photoelectric sensors, were also tested,

but less successfully because they were too sensitive to the environment. It was discovered that frost heaving of the ground would eventually cause their buried loop to fail, and the cost of underground excavation to replace the loop was exorbitant.

Another application of the microwave sensor is the door-opening market. The microwave sensor will check, for safety reasons, the area behind a swinging door to detect whether there is an individual or an object in the path way. Ultrasonic sensors may perform the same task, yet range and environmental conditions often make a microwave sensor more desirable.

A microwave sensor can check boxes to verify that objects actually have been packed therein. The sensor has the ability to see through the box itself and triggers only if an object is contained in the box. This technology relies on the sensed object being more reflective than the package, a condition that is often met.

2.10.5 Measuring Velocity with Microwave Sensors

Microwave sensors are ideally suited to measuring linear velocity. Police radar is a simple example of a Doppler-frequency-based velocity sensor. This technology can be applied wherever it is necessary to determine velocity in a noncontact manner.

2.10.6 Detecting Direction of Motion with Microwave Sensors

Direction of motion—whether a target is moving toward or away from the microwave sensor—can be determined by the use of the Doppler-frequency concept (Fig. 2.101), by adding an extra mixer

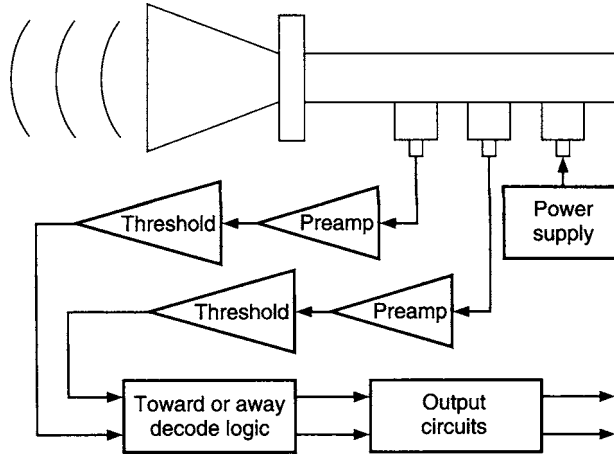


Figure 2.101 Direction of motion sensor schematic.

diode to the module. A discriminating effect is generated by the additional diode, which is located in the waveguide such that the Doppler outputs from the two mixers differ in phase by one-quarter wavelength, or 90°. These outputs will be separately amplified and converted into logic levels. The resulting signals can then be fed into a digital phase-discrimination circuit to determine the direction of motion. Such circuits are commonly found in motion control applications in conjunction with optical encoders. Figure 2.102 shows the phase relationships of the different directions.

Outputs from this module can vary widely to suit the application. The simplest is two outputs, one for motion and the other for direction (toward or away). These outputs can be added to a third, which provides the velocity of the target. The combination of signals could be analyzed to provide a final output when specific amplitude, direction, distance, and velocity criteria are met (Fig. 2.101).

In the door-opening field, using the amplitude, direction, distance, and velocity information reduces the number of false openings. This extends the life of the door mechanism, besides saving heat if the door is an entrance to a heated building.

In this case, the measurements by circuitry indicate the following characteristics:

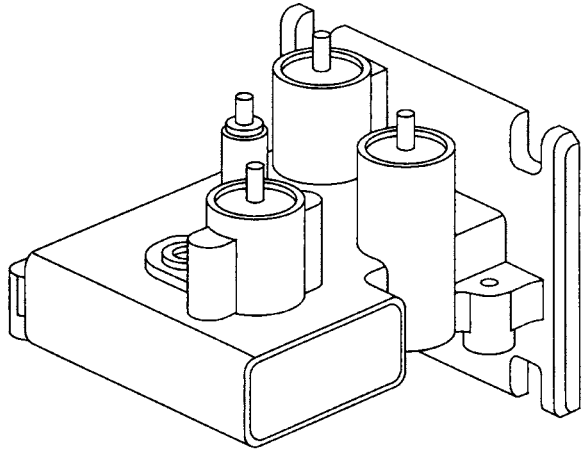
Characteristic	Measurement
Person-sized object	Amplitude of return
Moving at walking pace	Velocity
Toward or away	Direction
Specific time before opening	Distance

2.10.7 Detecting Range with Microwave Sensors

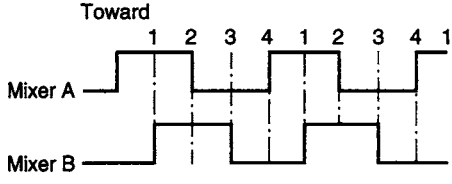
An early-warning military radar system depends on costly microwave sensors. A small yacht may use a microwave sensor selling for less than $1000 to detect targets at ranges up to 5 mi.

Regardless of their cost, microwave range sensors for commercial, industrial, and military applications employ essentially the same measuring technique. They transmit a narrow pulse of energy and measure the time required for the return from the target. Since microwave energy propagates at the speed of light, the time for the pulse to reach the target and return is 2 ns per foot of range. If the range to the target is 1 mi, the time required is 10.56 μs.

Although the microwave power needed is sufficient to raise the sensor temperature to 500°F, the design of the signal processing circuitry to measure the response is not difficult. However, if the target is very close to the transmitter, then the short response time may pose a real problem. At 3 ft, the time response is 6 ns. For 1-in resolution,

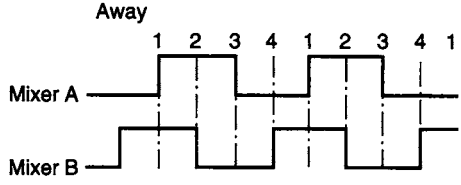


(a)



	A	B
1	1	↑
2	↓	1
3	0	↓
4	↑	0

(b)



	A	B
1	↑	1
2	1	↓
3	↓	0
4	0	↑

(c)

Figure 2.102 (*a*) Direction of motion sensor device. (*b*) Motion logic "away." (*c*) Direction logic "toward."

the circuitry must be able to resolve 167 ps. This may pose a significant problem.

The alternative method to resolve a target at a short range involves changing the frequency of a continuous oscillator. This method is better-suited to industrial applications. An oscillator starting at 10.525 GHz and sweeping at 50 MHz in 10 ms in the 6 ns mentioned earlier will have changed its frequency by:

$$(6 \text{ ns} \times 50 \text{ MHz}/0.01 \text{ s}) = 30 \text{ Hz}$$

The returning wave will still be 10.525 GHz. The output from the mixer diode as the sweep continues will be the 30-Hz difference. If this frequency is averaged over time, it is not difficult to resolve a range to 0.001 in.

The preceding calculation indicates that the frequency is high for faraway objects and low for targets that are close. This leads to two conclusions:

- The closer the object is, the lower the frequency, and therefore the longer the measurement will take.
- The signal processing amplifier should have a gain that increases with frequency.

Problems can arise when the target is moving or there are multiple targets in the area of interest. Movement can be detected by comparing consecutive readings and can be used as a discrimination technique. Multiple targets can be defined by narrow-beam antennas to reduce the width of the area of interest. Gain adjustments are also required to eliminate all but the largest target. Audio-bandwidth filters may be used to divide the range into sectors for greater accuracy.

Other sensor types, such as photoelectric and inductive sensors, may be utilized to measure distances. Inductive sensors are used in tank level measurements. They must be coupled with a moving component that floats on the surface of the substance to be measured.

Photoelectric sensors measure position by focusing a beam on a point in space and measuring the reflection on a linear array. This can give very precise measurements over a limited range but is subject to adverse environmental conditions. Photoelectric sensors can focus a camera on a target for a better picture. Ultrasonic sensors may perform the same function, but their range is limited and they can be defeated by a hostile environment.

Microwave sensors for measuring range have an impressive array of applications, including measurement of the level of liquid or solid in a tank, sophisticated intrusion alarms, autonomous guided vehicle industrial systems, and noncontact limit switching. In tank level

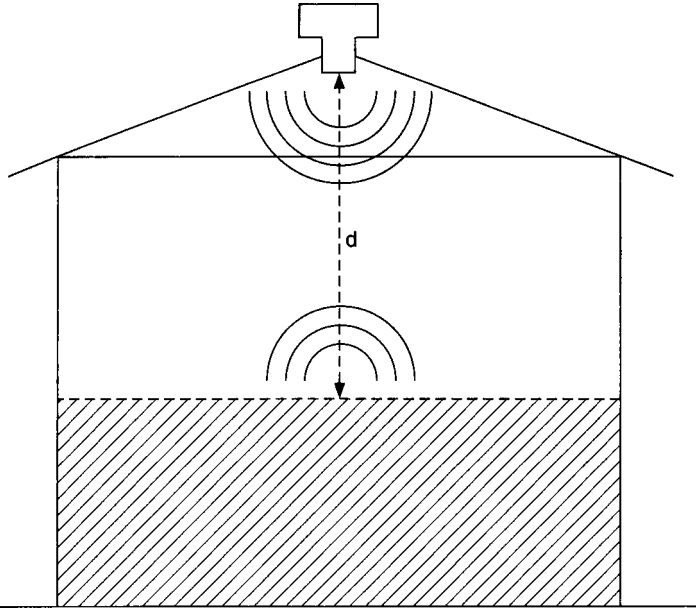


Figure 2.103 Tank level sensor.

sensing in the chemical industry (Fig. 2.103), the microwave sensor is mounted at the top of the tank and measures the distance from that position to the surface of the contents. Since the electronic circuitries can be isolated from the tank contents by a sealed window, it is intrinsically safe. It has the advantage of being a noncontact system, which means there are no moving parts to break or be cleaned. This allows the microwave sensor to be used on aggressive chemicals, liquids, liquefied gases, highly viscous substances, and solids such as grain and coal.

2.10.8 Microwave Technology Advancement

Advances in technology have opened up important new applications for microwave sensors. The expected governmental permission to utilize higher frequencies and the decreasing size of signal processing circuitry will significantly reduce the cost of the sensors and will enable them to detect even smaller targets at a higher resolution. Microwave integrated circuit technology (MICT), presently developed for the Military Microwave Integrated Circuit (MIMIC) program will overflow into the industrial market, causing increases in performance and in analysis capabilities. Consequently, the utilization of computer-integrated manufacturing technology will be broadened.

2.11 Confocal Microscopy Sensors

Confocal Microscopy inspection equipment is currently used to inspect probe marks, which is designed to handle high throughput, while capturing process excursions with high probability. Once excursions are detected, the data provides probe marks in the anomalous areas.

Efficient defect review, characterization, and root-cause analysis require a flexible and interactive tool that can provide both lateral and height measurements, that can revisit and fully describe features in the areas of interest, and that can provide engineers and scientists with timely data so they can correct the problems. The tool should supply the data required for production meetings as well as images that will represent the data to management and remote customers. The tool should allow automated review and give engineers the option of full manual control to devise an effective solution.

2.11.1 Confocal Profiling Characterization Systems

To date, conventional optical review stations and scanning-electron microscopy have served this purpose, giving engineers good lateral measurements and pictures of the features but providing only a rough impression of their volumetric extent. Good depth measurement by these methods requires skillful cross sectioning to view the deepest part of the features, a process that generally delays results for long periods, often for days, with works in progress at a standstill.

In the span of only several seconds, advanced confocal profiling systems provide the same height information available from the most careful scanning electron microscope (SEM) cross-section measurement (Fig. 2.104). Cross-sectioning is carried out on the data rather than on the sample, so the untouched wafer can be returned to the process unharmed promptly after determining the disposition. Analysis may be performed under automated recipe control, interactively under recipe guidance, or manually, allowing the engineer to reach the root cause.

Images generated by Hyphenated Systems' advanced confocal profiling systems provide the same height information available from scanning-electron microscope (SEM) cross-section measurements (Fig. 2.105).

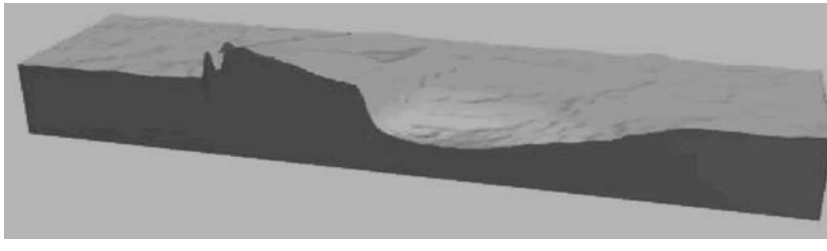


Figure 2.104 Advanced focal profiling providing highly accurate heights.

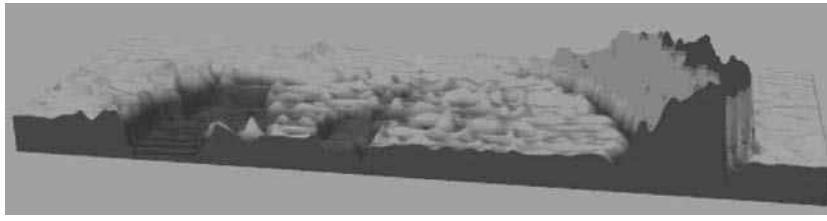


Figure 2.105 Images (generated by Hyphenated Systems) of advanced confocal profiling systems provide the same height information available from scanning-electron microscope (SEM) cross-section measurements.

The probe mark is relatively shallow, with the deepest portion less than 500 nm below the pad level.

The mark is deeper and clearly shows a breakthrough to a barrier layer and the underlying dielectric around 1 μm beneath the pad surface. (*Courtesy of Hyphenated Systems*.)

2.11.2 Driving Inspection Requirements

Wire bonding sensor application is by far the predominant interconnect technology for current multichip packages. As manufacturers have gained experience with multichip packages, they have realized the importance of probe-pad integrity. In particular, damage caused by contact between the pad and the electrical test probe (Fig. 2.106) is a primary cause of failure in wire bond interconnects. Failure modes

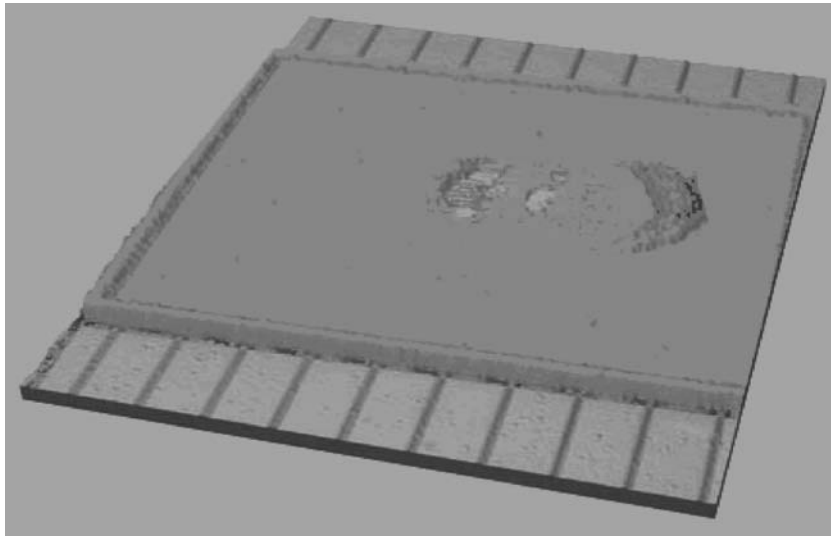


Figure 2.106 Damage caused by contact between the pad and the electrical test probe.

include excessive marking, which interferes with the bond reliability; punch-through, which reduces bond reliability and exposes underlying circuitry to the bonding process; and etch defects, in which overlying passivation is not completely cleared from the bond pad.

Another issue of growing importance is vendor accountability. Many devices incorporate chips from different vendors. Corporate integrators do not desire to be held responsible for defects that occurred upstream in the supply chain, and many are implementing incoming probe-mark inspection or imposing requirements for outgoing inspection on their suppliers, or both.

Microscopy can provide for the detailed review and characterization of probe marks. In this example, the mark is well resolved in three dimensions, and a breakthrough to underlying layers is clearly shown. (*Courtesy of Hyphenated Systems*.)

Advanced technology nodes (65 nm and below) are also increasing the need for probe-mark inspection. This vulnerability is exacerbated by the relentless efforts to put more circuitry on smaller die and the resulting trend to locate active circuitry beneath bond pads.

Another offender in the category of new materials is the use of copper on bond pads. Copper is a highly corrosive material. Copper particulates created during probing can easily damage the circuit, as illustrated in Fig. 2.106.

Recent developments in testing technology are also contributing to the demand for probe-mark inspection. There is significant incentive to reduce any risk of damage to probe cards that might be caused by particulate contaminants on the probe pad. A number of manufacturers have implemented pre-probe inspections to minimize this risk. Probes wear during normal use but may be reconditioned or repaired up to a certain point. Careful analysis of probe marks can provide information that signals the need for reconditioning of a probe.

Another trend in test technology is the movement of electrical testing upstream, into the fabrication process, to allow faster detection, diagnosis, and correction of defective processes. Though faster feedback is always desirable for process control, the need has become more urgent as the fraction of defects that cannot be detected with conventional in-line inspection techniques has grown at each successive technology node. Limiting, or at least detecting, probe-induced damage in the earlier stages of wafer processing will play an important role in the successful integration of electrical tests in the fabrication process.

Probe-mark characterization, as distinct from probe-mark inspection, can help engineers determine appropriate probing parameters during the characterization phase of process development.

Probe-mark characterization (Fig. 2.105) relies on detailed three-dimensional inspection. Three-dimensional inspection techniques tend to be too slow and expensive for production applications, but

engineers can use them to carefully quantify the effects of variables such as probe force to determine the best operational window for their process.

2.11.3 Probe-Mark Inspection Requirements

The probe-mark inspection process must be capable of evaluating a number of quantitative and qualitative characteristics. Quantitative measures include size, location within the pad, proximity to the edge of the pad, orientation relative to the center of the pad, and orientation trends across the wafer. Qualitative indicators include over-etch and under-etch, pitting, corrosion, abrasion, roughness, and punch-through. Advanced automatic defect classification (ADC) that makes efficient use of all available input to provide fast accurate classification is a practical requirement for probe-mark inspection in a production environment.

2.11.4 Multi-Chip Packages

The use of multi-chip packages is increasing, driven primarily by consumer demand for larger memory, multifunction handheld devices, and increased electronic capabilities in automobiles. The primary first-level interconnect technology in multi-chip packages is wire bonding. Probe marks created by electrical testing are a primary cause of failure in wire bonds. Probe-mark inspection is playing an increasingly important role in maintaining the yield and reliability of wire.

2.12 Understanding Laser Sensors

Two theories about the nature of light have been recognized. The particle theory was the first presented to explain the phenomena that were observed concerning light. According to this theory, light is a particle with mass, producing reflected beams. It was believed that light sources actually generated large quantities of these particles. Through the years, however, many phenomena of light could not be explained by the particle theory, such as reflection of light as it passes through optically transparent materials.

The second theory considered that light was a wave, traveling with characteristics similar to those of water waves. Many, but not all, phenomena of light can be explained by this theory.

A dual theory of light has been proposed, and is presently considered to be the true explanation of light propagation. This theory suggests that light travels in small packets of wave energy called *photons*. Even though photons are *bundles* of wave energy, they have momentum like particles of mass. Thus, light is wave energy traveling with some of the characteristics of a moving particle. The total transmitted as light is the sum of energies of all the individual photons emitted.

Velocity, frequency, and wavelength are related by the equation:

$$c = f\lambda$$

where c = velocity of light, km/s
f = frequency, Hz
λ = wavelength, m

This equation shows that the frequency of a wave is inversely proportional to the wavelength—that is, higher-frequency waves have shorter wavelengths.

2.12.1 Properties of Laser Light

Laser stands for *light amplification by stimulated emission of radiation*. Laser light is monochromatic, whereas standard white light consists of all the colors in the spectrum and is broken into its component colors when it passes through a standard glass prism (Figs. 1.107 and 1.108).

2.12.2 Essential Laser Components

Laser systems consist of four essential components:

- The active medium
- The excitation mechanism
- The feedback mechanism
- The output coupler

2.12.2.1 The Active Medium

The active medium is the collection of atoms, ions, or molecules in which stimulated emission occurs. It is in this medium that laser light is produced. The active medium can be a solid, liquid, gas, or semiconductor material. Often, the laser takes its name from that of the

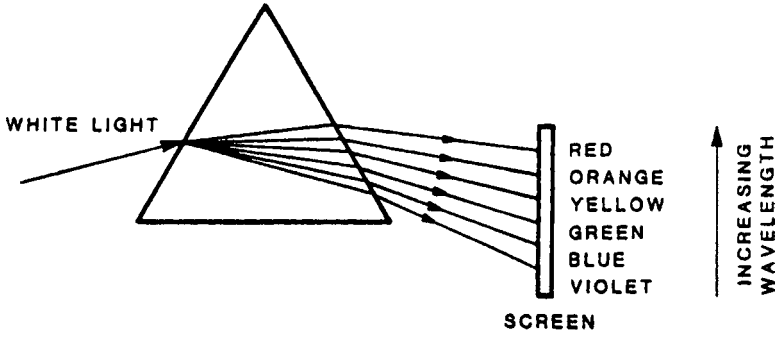


Figure 2.107 Standard white light.

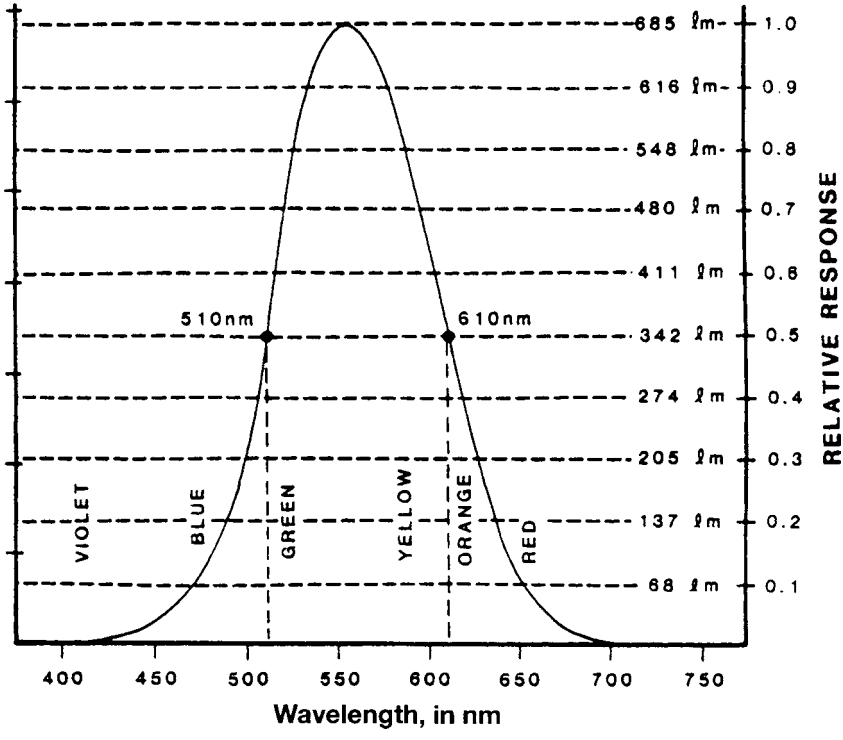


Figure 2.108 Spectrum of standard white light.

active medium. For example, the ruby laser has a crystal of ruby as its active medium while the CO_2 laser has carbon dioxide gas.

The wavelength emitted by a laser is a function of the active medium. This is because the atoms within the active medium have their own characteristic energy levels at which they release photons. It will be shown later that only certain energy levels within the atom can be used to enhance stimulated emission. Therefore, a given active medium can produce a limited number of laser wavelengths, and two different active media cannot produce the same wavelengths. Table 2.10 contains a list of materials commonly used in lasers and the corresponding wavelengths that these materials produce.

The active medium is the substance that actually lases. In the helium-neon laser, only the helium lases.

2.12.2.2 Excitation Mechanism

The excitation mechanism is the device used to put energy into the active medium. Three primary types of excitation mechanisms exist—optical, electrical, and chemical. All three provide the energy necessary to raise the energy state of the atom, ion, or molecule of the active

Type of Active Medium	Common Material	Wavelength Produced, nm
Solid	Ruby	694
	Nd:YAG	1,060
	Nd:glass	1,060
	Erbium	1,612
Liquid	Organic dyes	360–650
Gas	Argon (ionized)	488
	Helium-neon	632.8
	Krypton (ionized)	647
	CO_2	10,600
Semiconductor	Gallium arsenide	850
	Gallium antimonide	1,600
	Indium arsenide	3,200

Table 2.10 Wavelengths of Laser Materials

medium to an excited state. The process of imparting energy to the active medium is called *pumping the laser*.

2.12.2.2.1 Optical Excitation An optical excitation mechanism uses light energy of the proper wavelength to excite the active medium. The light may come from any of several sources, including a flash lamp, a continuous arc lamp, another laser, or even the sun. Although most of these use an electric power supply to produce the light, it is not the electrical energy that is used directly to excite the atoms of the active medium, but rather the light energy produced by the excitation mechanism.

Optical excitation is generally used with active media that do not conduct electricity (e.g., solid lasers like the ruby). Figure 2.109 is a schematic drawing of a solid laser with an optical pumping source.

The sun is considered a possible optical pumping source for lasers in space. The optical energy from the sun could be focused by curved mirrors onto the laser's active medium. Since the size and weight of an electric power supply is of concern in space travel, solar pumping of lasers is an interesting alternative.

2.12.2.2.2 Electrical Excitation Electrical excitation is most commonly used when the active medium will support an electric current. This is usually the case with gases and semiconductor materials.

When a high voltage is applied to a gas, current-carrying electrons or ions move through the active medium. As they collide with

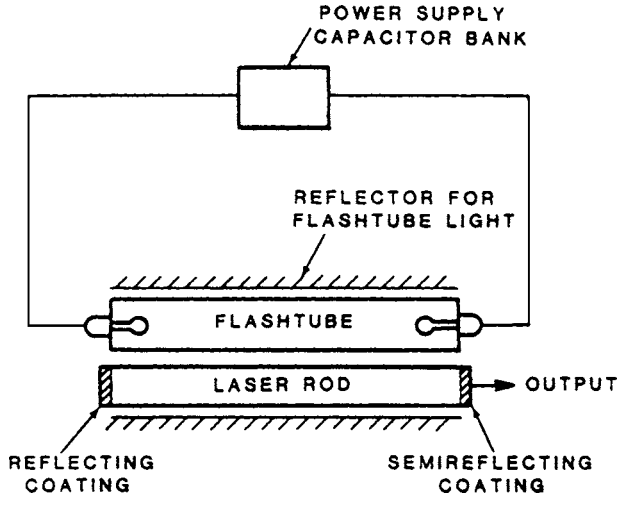


Figure 2.109 Solid laser with optical pumping source.

the atoms, ions, or molecules of the active medium, their energy is transferred and excitation occurs. The atoms, ions, and electrons within the active medium are called *plasma*.

Figure 2.110 is a schematic drawing of a gas laser system with electrical excitation. The gas mixture is held in a gas plasma tube and the power supply is connected to the ends of the plasma tube. When the power supply is turned on, electron movement within the tube is from the negative to the positive terminal.

2.12.2.2.3 Chemical Excitation Chemical excitation is used in a number of lasers. When certain chemicals are mixed, energy is released as chemical bonds are made or broken. This energy can be used as a pumping source. It is most commonly used in hydrogen-fluoride lasers, which are extremely high-powered devices used primarily in

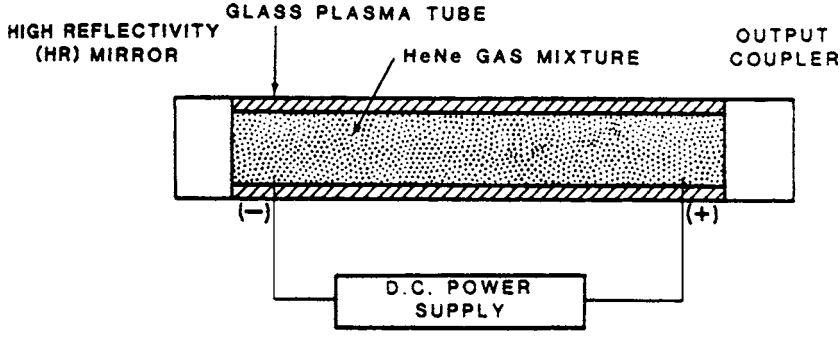


Figure 2.110 Gas laser with electrical excitation.

military weapons and research. These lasers are attractive for military applications because of the large power-to-weight ratio.

2.12.2.3 Feedback Mechanism

Mirrors at each end of the active medium are used as a *feedback mechanism*. The mirrors reflect the light produced in the active medium back into the medium along its longitudinal axis. When the mirrors are aligned parallel to each other, they form a resonant cavity for the light waves produced within the laser. They reflect the light waves back and forth through the active medium.

In order to keep stimulated emission at a maximum, light must be kept within the amplifying medium for the greatest possible distance. In effect, mirrors increase the distance traveled by the light through the active medium. The path that the light takes through the active medium is determined by the shape of the mirrors. Figure 2.111 shows some of the possible mirror combinations. Curved mirrors are often used to alter the direction in which the reflected light moves.

2.12.2.4 Output Coupler

The feedback mechanism keeps the light inside the laser cavity. In order to produce an output beam, a portion of the light in the cavity must be allowed to escape. However, this escape must be controlled. This is most commonly accomplished by using a partially reflective mirror in the feedback mechanism. The amount of reflectance varies with the type of laser. A high-power laser may reflect as little as 35 percent, with the remaining 65 percent being transmitted through the

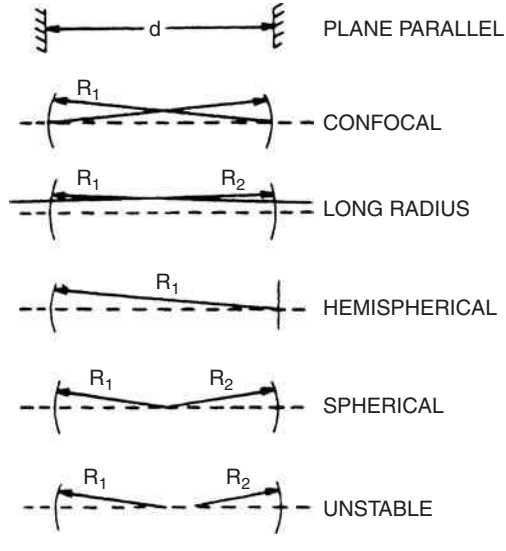


Figure 2.111 Mirror combinations for a feedback mechanism.

mirror to become the output laser beam. A low-power laser may require an output mirror reflectivity as high as 98 percent, leaving only 2 percent to be transmitted. The output mirror that is designed to transmit a given percentage of the laser light in the cavity between the feedback mirrors is called the *output coupler*.

2.12.3 Semiconductor Displacement Laser Sensors

Semiconductor displacement laser sensors, consisting of a light-metering element and a position-sensitive detector (PSD), detect targets by using triangulation. A light-emitting diode or semiconductor laser is used as the light source. A semiconductor laser beam is focused on the target by the lens. The target reflects the beam, which is then focused on the PSD, forming a beam spot. The beam spot moves on the PSD as the target moves. The displacement of the workpiece can then be determined by detecting the movement of the beam spot.

2.12.3.1 Industrial Applications of Semiconductor Displacement Lasers

The laser beam emitted from the laser diode in the transmitter is converged into a parallel beam by the lens unit. The laser beam is then directed through the slit on the receiver and focused on the light-receiving element. As the target moves through the parallel laser beam, the change in the size of the shadow is translated into the change in received light quantity (voltage). The resulting voltage is used as a comparator to generate an analog output voltage.

2.12.4 Industrial Applications of Laser Sensors*

Electrical and electronics industries:

- *Warpage and pitch of IC leads.* The visible beam spot facilitates the positioning of the sensor head for small workpieces. Warpage and pitch can be measured by scanning IC leads with the sensor head (Fig. 2.112).
- *Measurement of lead pitch of electronic components.* The sensor performs precise noncontact measurement of pitch using a laser beam (Fig. 2.113).
- *Measurement of disk head movement.* The laser sensor is connected to a computer in order to compare the pulse input to the disk head drive unit with actual movement. The measurement is done on-line, thus increasing productivity (Fig. 2.114).
- *Detection of presence/absence of resin coating.* The laser displacement sensor determines whether a resin coating was formed after wire bonding (Fig. 2.115).

*A few nonlaser optical sensors are included, as indicated.

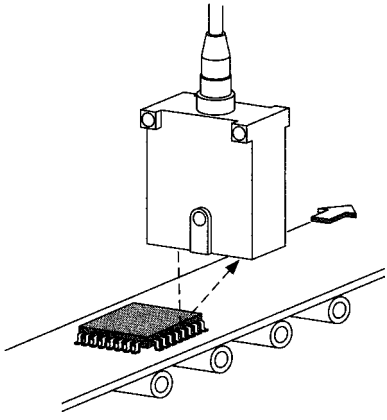


Figure 2.112 Warpage and pitch of IC lead.

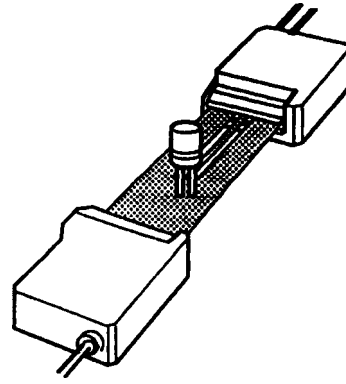


Figure 2.113 Measurement of lead pitch of electronic components.

Figure 2.114 Measurement of disk head movement.

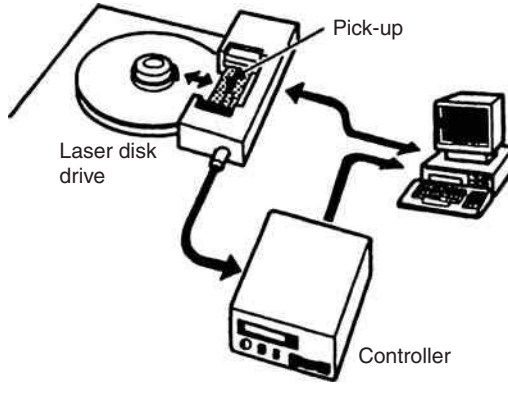
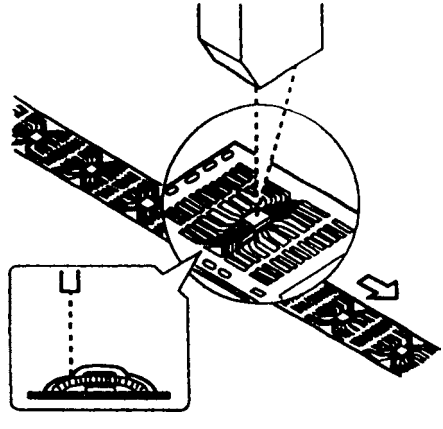


Figure 2.115 Detection of presence/absence of resin coating.



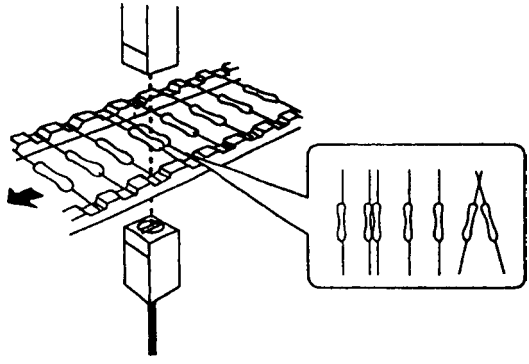


FIGURE 2.116 Detection of double-fed or mispositioned resistors.

- *Detection of double-fed or mispositioned resistors prior to taping.* Through-beam-type sensor heads are positioned above and below the resistors traveling on a transfer line. A variation on the line changes the quantity of light in the laser beam, thus signaling a defect (Fig. 2.116).
- *Detection of defective shrink wrapping of videocassette.* Defective film may wrap or tear during shrink wrapping. The laser sensor detects defective wrapping by detecting a change in the light quantity on the surface of the videocassette (Fig. 2.117).
- *Measurement of gap between roller and doctor blade.* Measures the gap between the roller and the doctor blade in submicrometer units. The sensor's automatic measurement operation eliminates reading errors (Fig. 2.118).
- *Measurement of surface run-out of laser disk.* The surface run-out of a laser disk is measured at a precision of 0.5 μm. The sensor head enables measurement on a mirror-surface object (Fig. 2.119).

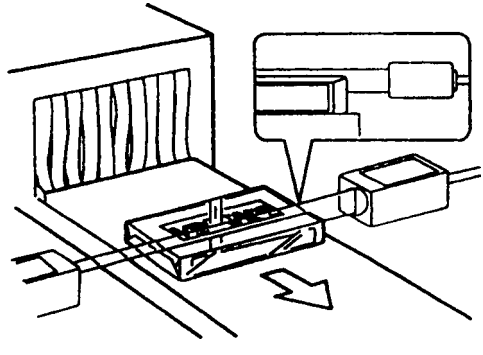


FIGURE 2.117 Detection of defective shrink wrapping of videocassette.

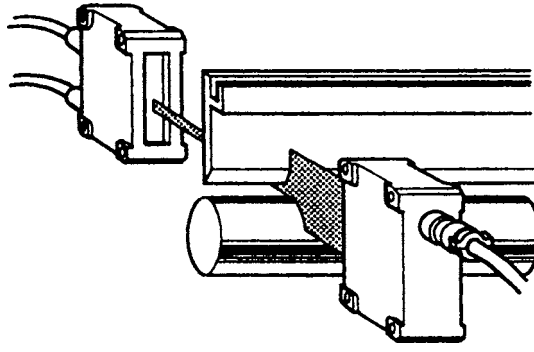


Figure 2.118 Measurement of gap between roller and doctor blade.

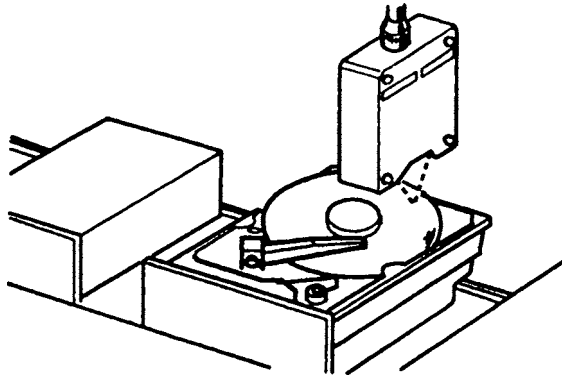


Figure 2.119 Measurement of surface run-out of laser disk.

- *Displacement of printer impact pins.* The visible beam spot facilitates positioning of the head of a pin-shaped workpiece, enabling measurement of the vertical displacement of impact pins (Fig. 2.120).

Automotive manufacturing industries:

- *Measurement of thickness of connecting rod.* Measures the thickness of the connecting rod by processing the analog inputs in the digital meter relay (Fig. 2.121).
- *Measurement of depth of valve recesses in piston head.* Measures the depth of the valve recesses in the piston head so that chamber capacity can be measured. Iron jigs are mounted in front of the sensor head, and the sensor measures the distance the jigs travel when they are pressed onto the piston head (Fig. 2.122).*

*Nonlaser sensor.

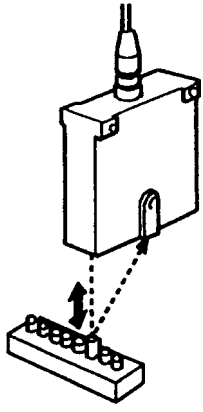


Figure 2.120 Displacement of printer impact pins.

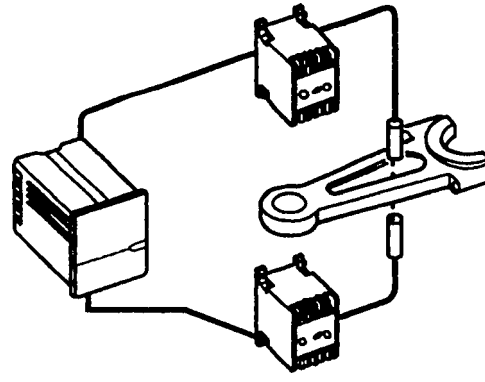
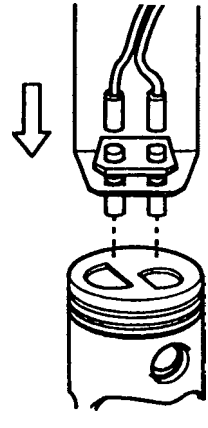


Figure 2.121 Measurement of thickness of connecting rod.

- *Measurement of height of radiator fin.* Detects improper radiator fin height by comparing the bottom value of the analog output with a stored pair of tolerances (Fig. 2.123).
- *Measurement of outer diameter of engine valve.* The laser scan micrometer allows on-line measurement of the outer diameter of engine valves simply by positioning a separate sensor head on either side of the conveyer (Fig. 2.124).
- *Positioning of robot arm.* The laser displacement sensor is used to maintain a specific distance between the robot arm and target. The sensor outputs a plus or minus voltage if the distance becomes greater or less, respectively, than the 100-mm reference distance (Fig. 2.125).

Figure 2.122 Measurement of depth of valve recesses in piston head.



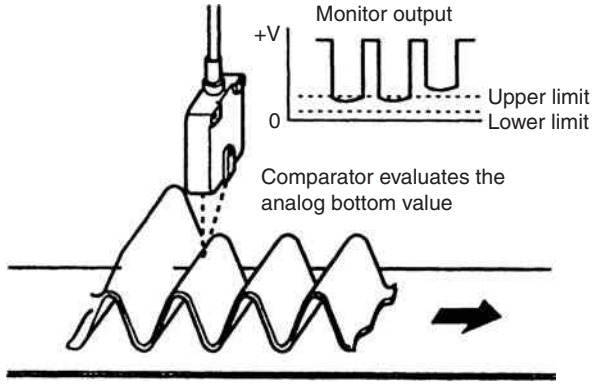


Figure 2.123 Measurement of height of radiator fin.

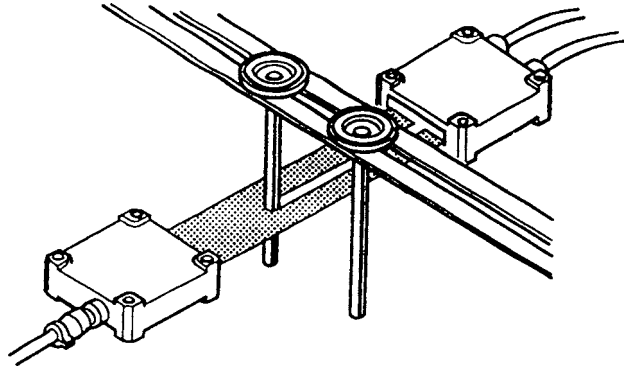


Figure 2.124 Measurement of outer diameter of engine valve.

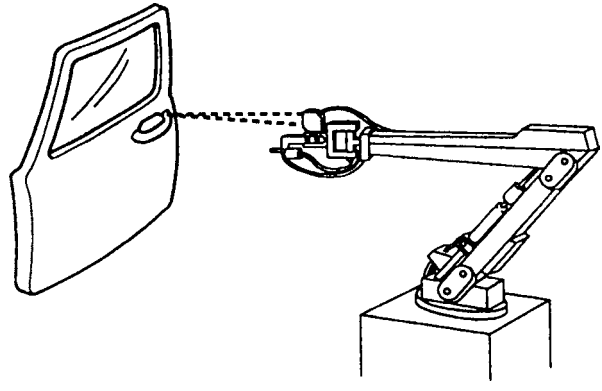


Figure 2.125 Positioning of robot arm.

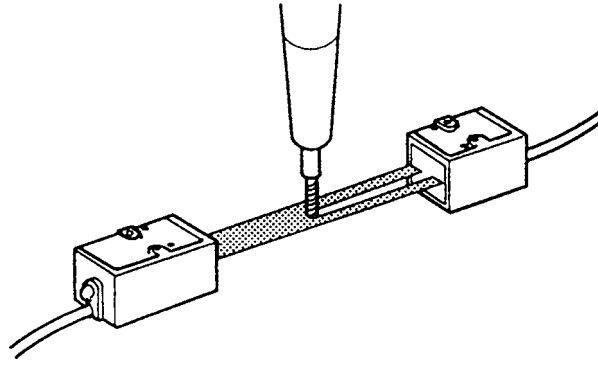


FIGURE 2.126 Detection of damage on microdiameter tool.

- *Detection of damage on microdiameter tool.* Detects a break, chip, or excess swarf from the variation of light quantity received (Fig. 2.126).

Metal/steel/nonferrous industries:

- *Detection of misfeeding in high-speed press.* The noncontact laser sensor, timed by a cam in the press, confirms the material feed by monitoring the pilot holes and then outputs the result to an external digital meter relay (Fig. 2.127).
- *Simultaneous measurement of outer diameter and eccentricity of ferrite core.* Simultaneously measures the outer diameter and eccentricity of a ferrite core with a single sensor system. The two measured values can then be simultaneously displayed on a single controller unit (Fig. 2.128).

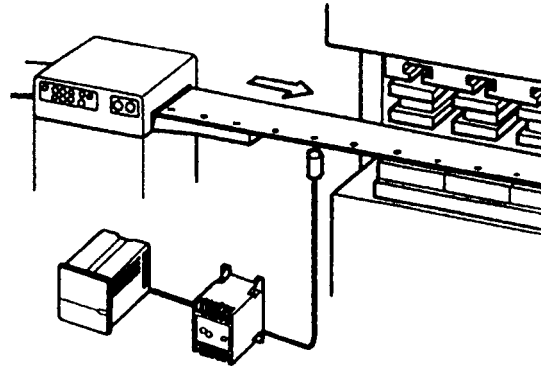


FIGURE 2.127 Detection of misfeeding in high-speed press.

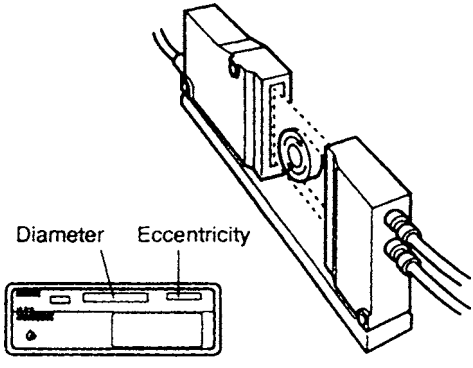


Figure 2.128 Simultaneous measurement of outer diameter and eccentricity of ferrite core.

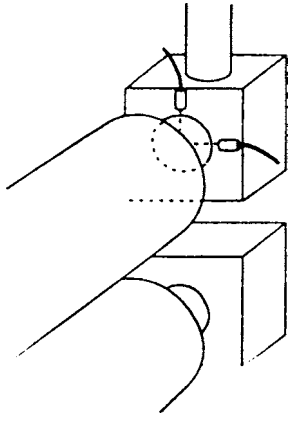


Figure 2.129 Confirmation of roller centering.

- *Confirmation of roller centering.* The analog output of the inductive displacement sensor is displayed as a digital value, thus allowing a numerical reading of the shaft position (Fig. 2.129).
- *Measurement of the height and inner diameter of sintered metal ring.* Determines the height and inner diameter of the metal ring by measuring the interrupted areas of the parallel laser beam (Fig. 2.130).
- *Measurement of outer diameter of wire in two axes.* Simultaneously measures in x axis to determine the average value of the outer diameter, thereby increasing dimensional stability (Fig. 2.131).

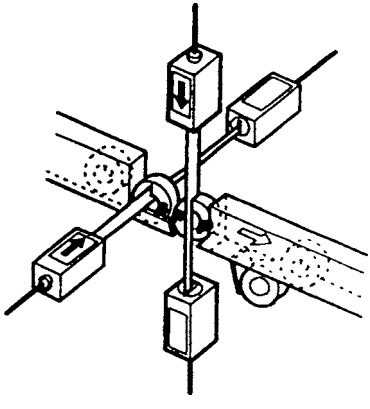


Figure 2.130 Measurement of the height and inner diameter of a sintered metal ring.

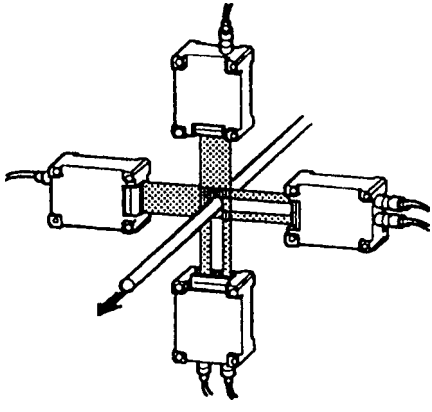


Figure 2.131 Measurement of the outer diameter of wire in two axes.

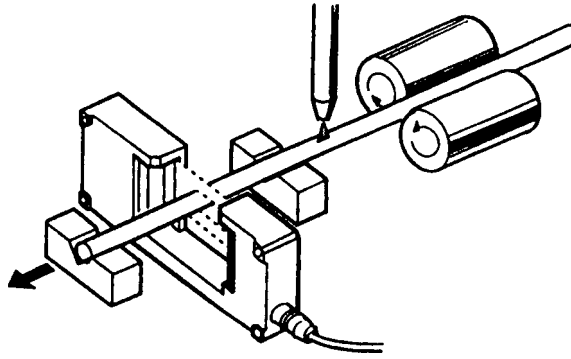


FIGURE 2.132 Measurement of the outer diameter after center-less grinding.

- *Measurement of outer diameter after centerless grinding.* The scanning head allows continuous noncontact measurement of a metal shaft immediately after the grinding process (Fig. 2.132).

Food processing and packaging:

- *Detection of material caught during heat sealing.* Detects material caught in the distance between rollers (Fig. 2.133).
- *Detection of missing or doubled packing ring in cap.* Detects a missing or doubled rubber ring in caps by using the comparator to evaluate sensor signals (Fig. 2.134).
- *Detection of incorrectly positioned small objects.* The transmitter and the receiver are installed to allow a parallel light beam to scan slightly above the tablet sheets. When a single tablet projects from a line of tablets, the optical axis is interrupted and the light quantity changes (Fig. 2.135).

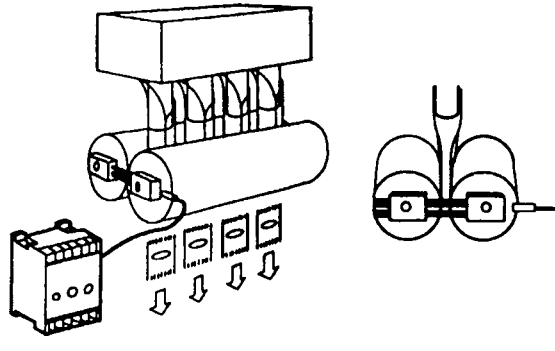


FIGURE 2.133 Detection of material caught during heat sealing.

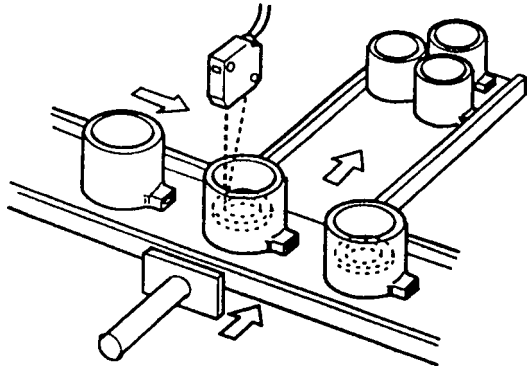


Figure 2.134 Detection of a missing or doubled packing ring in cap.

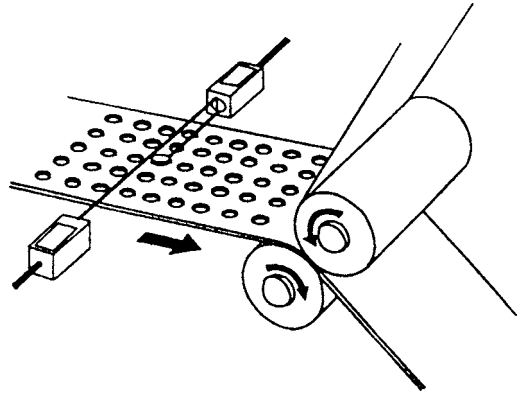


Figure 2.135 Detection of incorrectly positioned small objects.

- *Measurement of tape width.* Measures the width of running tape to the submicrometer level; 100 percent inspection improves product quality (Fig. 2.136).
- *Measurement of sheet thickness.* Use of two controllers enables thickness measurement by determining the distance in input values. Thus, thickness measurement is not affected by roller eccentricity (Fig. 2.137).

Automatic machinery:

- *Detection of surface run-out caused by clamped error.* Improper clamping due to trapped chips will change the rotational speed of the workpiece. The multifractional digital meter relay sensor calculates the surface run-out of the rotating workpiece, compares it with the stored value, and outputs a detection signal (Fig. 2.138).

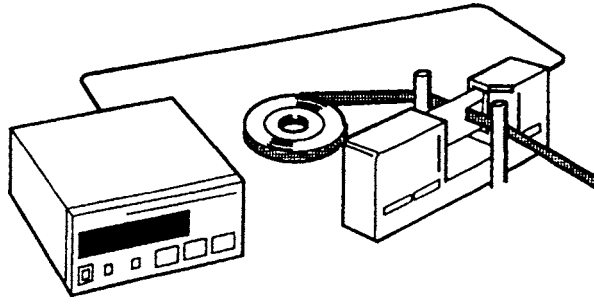


Figure 2.136 Measurement of tape width.

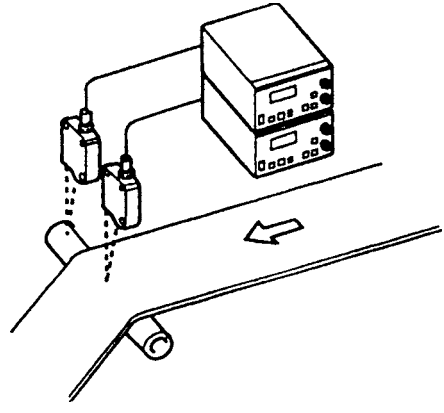


Figure 2.137 Measurement of sheet thickness.

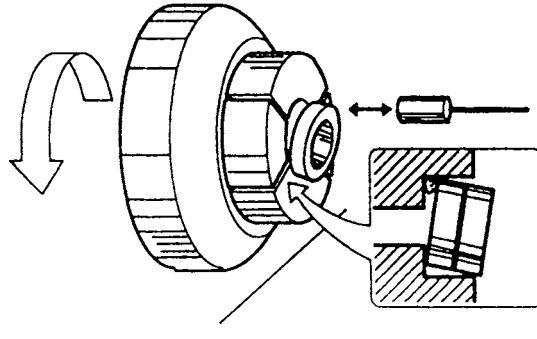


Figure 2.138 Detection of surface run-out caused by clamped error.

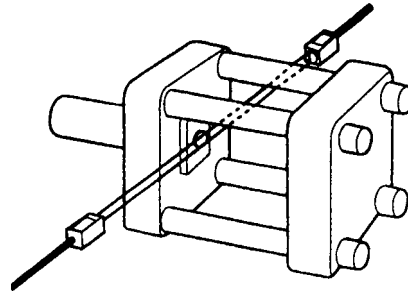


Figure 2.139 Detection of residual resin.

- *Detection of residual resin in injection molding machine.* When the sensor heads are such that the optical axis covers the surface of the die, any residual resin will interface with this axis (Fig. 2.139).
- *Measurement of travel of camera lens.* A separate sensor can be installed without interfering with the camera body, thus assuring a highly reliable reading of lens travel (Fig. 2.140).
- *Measurement of rubber sheet thickness.* With the segment function that allows the selection of measuring points, the thickness of a rubber sheet (i.e., the distance between the rollers) can be easily measured (Fig. 2.141).
- *Measurement of stroke of precision table.* Detects even minute strokes at a resolution of 0.05 μm. In addition, the AUTO ZERO function allows indication of relative movement (Fig. 2.142).
- *Measurement of plasterboard thickness.* A sensor head is placed above and below the plasterboard, and its analog outputs are fed into a digital meter relay. The meter relay indicates the absolute thickness value (Fig. 2.143).

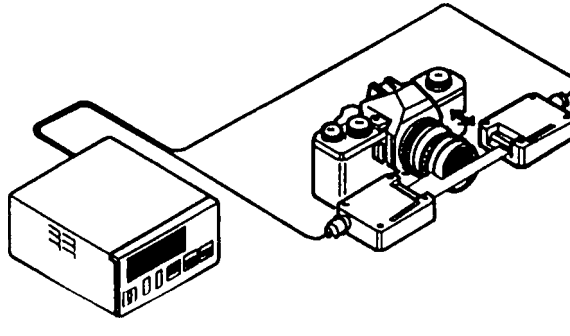


Figure 2.140 Measurement of the travel of a camera lens.

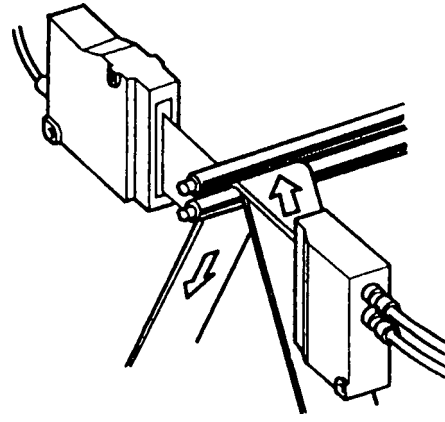


Figure 2.141 Measurement of rubber sheet thickness.

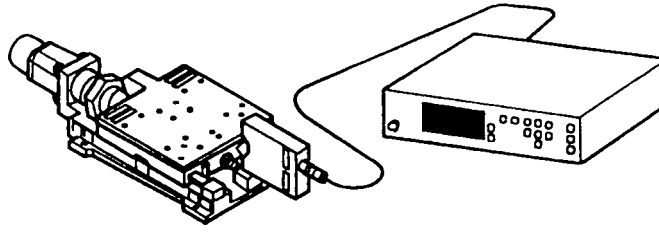


Figure 2.142 Measurement of the stroke of a precision table.

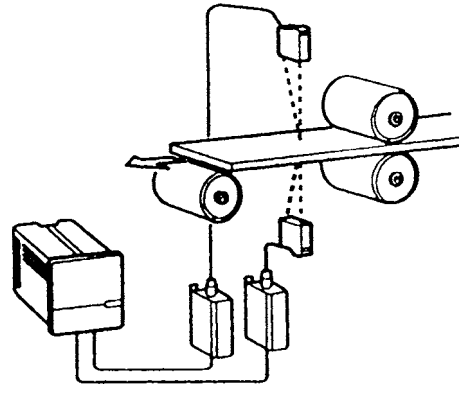


Figure 2.143 Measurement of plasterboard thickness.

Further Reading

Chappel, A. (ed.), *Optoelectronics: Theory and Practice*, McGraw-Hill, New York, 1978.
Doebelin, E. O., *Measurement Systems: Application and Design*, 4th ed., McGraw-Hill, New York, 1990.
Giallorenzi, T. G., "Optical Fiber Interferometer Technology and Hydrophones in Optical Fiber Sensors," NATO ASI Series E, No. 132, Martinus Nijhoff Dordrecht, 35–50 (1987).
Hocker, G. B., "Fiber Optic Sensing for Pressure and Temperature," *Appl. Optics*, 18(1445) (1979)
Holliday, D., and R. Resnick, *Physics*, Wiley, New York, 1975.
International Organization for Standardization, "Statistical Interpretation of Data: Comparison of Two Means in the Case of Paird Observations," ISO 3301-1975.
Jackson, D. A., A. D. Kersey, and A. C. Lewin, "Fiber Gyroscope with Passive Quadrature Demodulation," *Electron. Lett.*, 20(399) (1984).
Lion, K. L., *Elements of Electrical and Electronic Instrumentation*, McGraw-Hill, New York, 1975.
Neubert, H.K.P., *Instrument Transducers*, 2d ed., Clarendon Press, Oxford, 1975.
Ogata. K., *Modern Control Engineering*, 2d ed., Prentice-Hall, Englewood Cliffs, N.J, 1990.
Rock, I., *Lightness Constancy, Perception*, W. H. Freeman, New York, 1984.
Seippel, R. G., *Optoelectronics*, Reston Publishing Co., Reston, Va., 1981.
Shortley, G., and D. Williams, *Quantum Property of Radiation*, Prentice-Hall, Englewood Cliffs, N.J., 1971.
Todd, C. D. (Bourns Inc.), *The Potentiometer Handbook*, McGraw-Hill, New York, 1975.
White, R. M., "A Sensor Classification Scheme," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, March, 1987.

CHAPTER 3

Fiber Optics in Sensors and Control Systems

3.0 Introduction

Accurate position sensing is crucial to automated motion control systems in manufacturing. The most common components used for position sensing are photoelectric sensors, inductive proximity sensors, and limit switches. They offer a variety of options for manufacturing implementation from which highly accurate and reliable systems can be created. Each option has its features, advantages, and disadvantages that manufacturing personnel should understand for proper application. Three types of sensors are used in manufacturing applications:

- *Photoelectric sensors.* Long-distance detection.
- *Inductive proximity sensors.* Noncontact metal detection.
- *Limit switches.* Detection with traditional reliability.

3.1 Photoelectric Sensors—Long-Distance Detection

A photoelectric sensor is a switch that is turned on and off by the presence or absence of receiving light (Fig. 3.1). The basic components of a photoelectric sensor are a power supply, a light source, a photodetector, and an output device. The key is the photodetector, which is made of silicon, a semiconductor material that conducts current in the presence of light. This property is used to control a variety of output devices vital for manufacturing operation and control, such as mechanical relays, triacs, and transistors, which in turn control machinery.

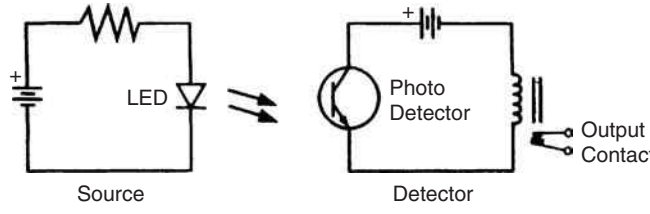


Figure 3.1 Photoelectric sensor.

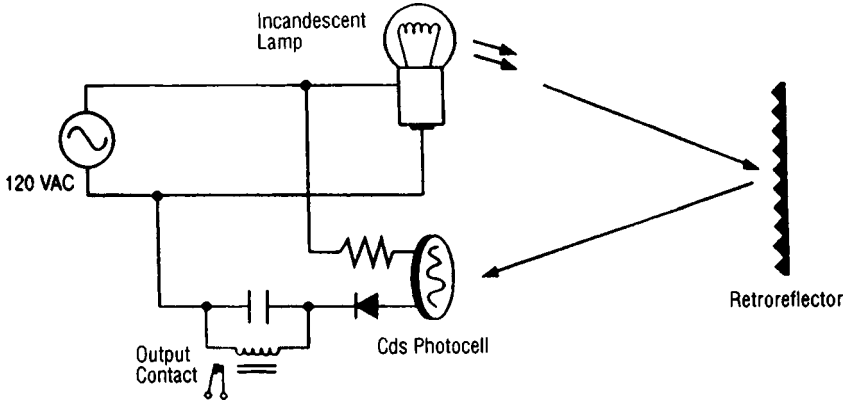


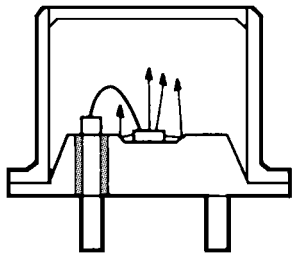
Figure 3.2 Early photoelectric control.

Early industrial photoelectric controls used focused light from incandescent bulbs to activate a cadmium sulfide photocell (Fig. 3.2). Since they were not modulated, ambient light such as that from arc welders, sunlight, or fluorescent light fixtures could easily false-trigger these devices. Also, the delicate filament in the incandescent bulbs had a relatively short life span, and did not hold up well under high vibration and the kind of shock loads normally found in an industrial environment. Switching speed was also limited by the slow response of the photocell to light/dark changes (Fig. 3.1).

3.1.1 Light-Emitting Diodes

Photoelectric sensors use an effective light source, light-emitting diodes (LEDs), which were developed in the early 1960s. LEDs are solid-state devices that emit light when current is applied (Fig. 3.3). This is the exact opposite of the photodetector, which emits current when light is received.

LEDs have several advantages over incandescent bulbs and other light sources. LEDs can be turned on and off very rapidly, are extremely small, consume little power, and last as long as 100,000 continuous hours. Also, since LEDs are solid-state devices, they are much more immune to vibration than incandescent bulbs.



Light Emitting Diodes (LED's)
- Fast turn-on and turn-off
- No warm-up
- Small
- Rugged
- Low power consumption
- High radiant efficiency
- Long life

FIGURE 3.3 Light-emitting diode.

LEDs emit light energy over a narrow wavelength (Fig. 3.4*a*). Infrared (IR) gallium arsenide LEDs emit energy only at 940 nm (Fig. 3.4*b*). As this wavelength is at the peak of a silicon photodiode's response, maximum energy transfer between source and detector is achieved.

A silicon photodetector's sensitivity to light energy also peaks in the infrared light spectrum. This contributes to the high efficiency and long range possible when silicon photodetectors are used in conjunction with gallium arsenide LEDs.

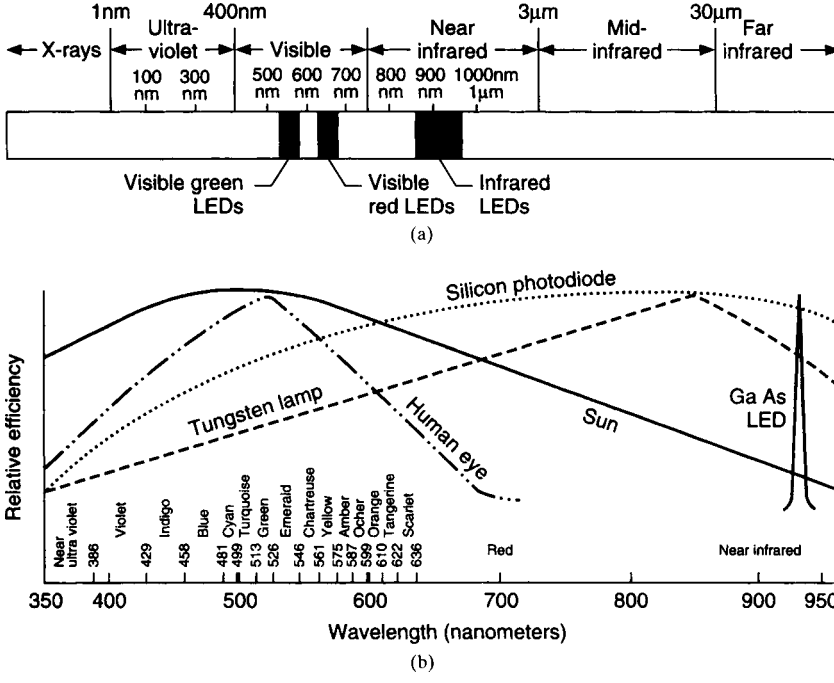


FIGURE 3.4 (a) LED emission wavelengths. (b) Infrared gallium arsenide LED emission.

In recent years, visible LEDs have been introduced as light sources in photoelectric controls. Because the beam is visible to the naked eye, the principle advantage of visible LEDs is ease of alignment. Visible beam photoelectric controls usually have lower optical performance than IR LEDs.

The modes of detection for optical sensors are (1) through-beam and (2) reflection (diffuse reflection and reflex detection).

3.1.2 Through-Beam Sensors

Through-beam sensors have separate source and detector elements aligned opposite each other, with the beam of light crossing the path that an object must cross (Fig. 3.5). The effective beam area is that of the column of light that travels straight between the lenses (Fig. 3.6).

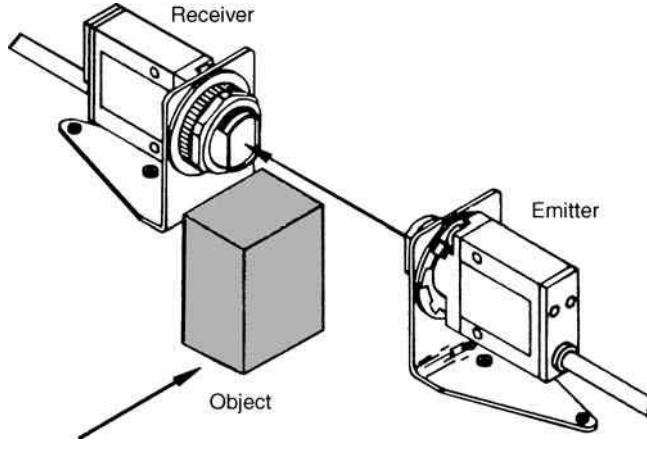


Figure 3.5 Through-beam sensor.

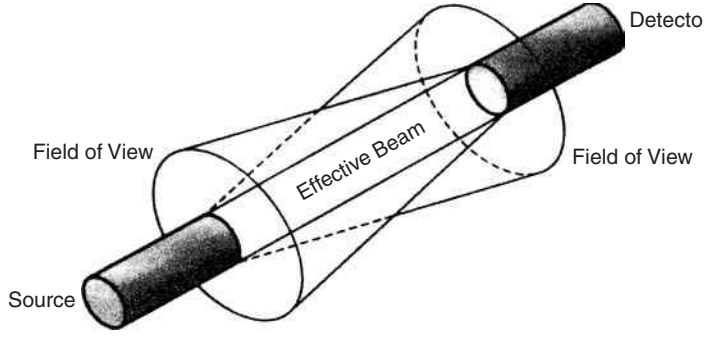


Figure 3.6 Effective beam area.

Because the light from the source is transmitted directly to the photodetector, through-beam sensors offer the following benefits:

- Longest range for sensing
- Highest possible signal strength
- Greatest light/dark contrast ratio
- Best trip point repeatability

The limitations of through-beam sensors are as follows:

- They require wiring of the two components across the detection zone.
- It may be difficult to align the source and the detector.
- If the object to be detected is smaller than the effective beam diameter, an aperture over the lens may be required (Fig. 3.7).

3.1.3 Reflex Photoelectric Controls

Reflex photoelectric controls (Fig. 3.8) position the source and detector parallel to each other on the same side of the target. The light is

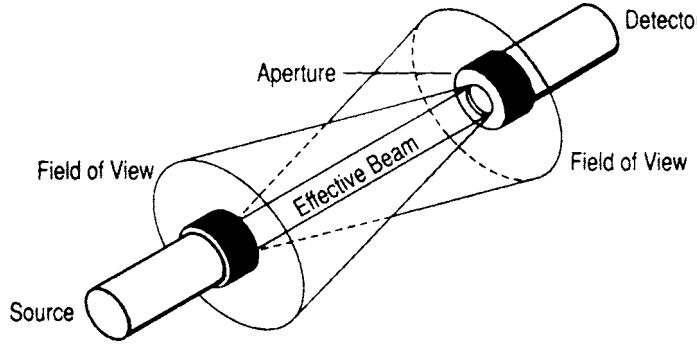


Figure 3.7 Sensor with aperture over lens for detecting small objects.

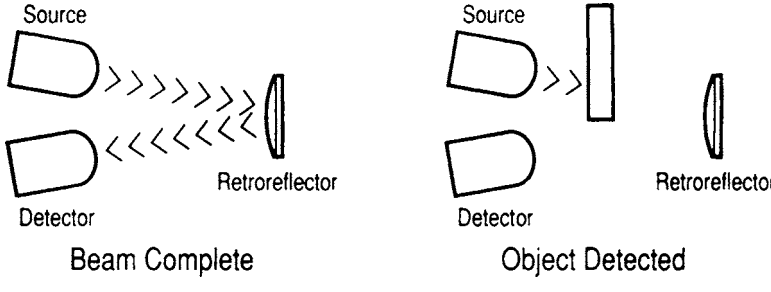


Figure 3.8 Reflex photoelectric controls.

directed to a retroreflector and returns to the detector. The switching and output occur when an object breaks the beam.

Since the light travels in two directions (hence twice the distance), reflex controls will not sense as far as through-beam sensors. However, reflex controls offer a powerful sensing system that is easy to mount and does not require that electrical wire be run on both sides of the sensing area. The main limitation of these sensors is that a shiny surface on the target object can trigger false detection.

3.1.4 Polarized Reflex Detection

Polarized reflection controls use a polarizing filter over the source and detector that conditions the light such that the photoelectric control sees only light returned from the reflector (Fig. 3.9). A polarized reflex sensor is used in applications where shiny surfaces such as metal or shrink-wrapped boxes may false-trigger the control.

Polarized reflex sensing is achieved by combining some unique properties of polarizers and retroreflectors. These properties are (1) polarizers pass light that is aligned along only one plane and (2) corner-cube reflectors depolarize light as it travels through the face of the retroreflector (Fig. 3.10).

Light from the source is aligned by a polarizer. When this light reflects off the retroreflector, it is depolarized. The returning light passes through another polarizing filter in front of the detector. The detector's polarizer is oriented at 90° to the source's polarizer. Only the light that has been rotated by the corner cube retroreflector can pass through the detector's polarizer. Light that bounces off other shiny objects, and has not been rotated 90°, cannot pass through the detector's polarizer, and will not trigger the control.

Polarized reflex sensors will not work with reflective tape containing glass beads. Also, shiny objects wrapped with clear plastic

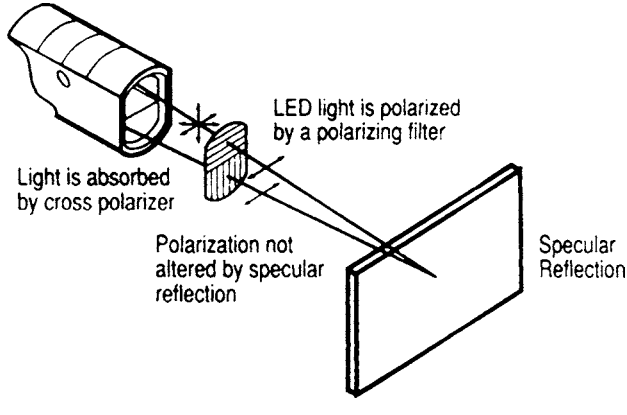


Figure 3.9 Polarization reflection controls.

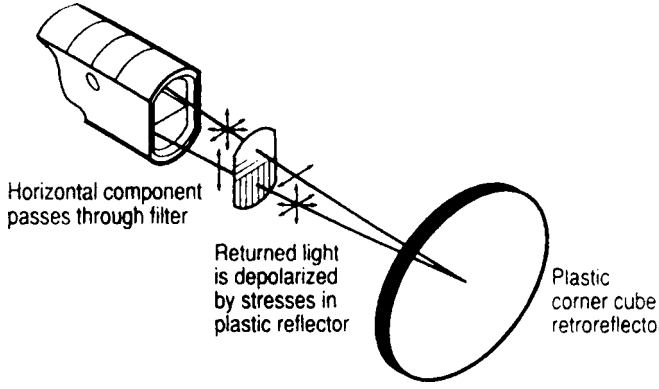


Figure 3.10 Corner-cube reflector.

shrink-wrap will potentially false-trigger a polarized reflex control, since under certain conditions these act as a corner-cube reflector.

The polarized reflex detection sensor offers the following advantages:

- It is not confused by the first surface reflections from target objects.
- It has a high dark/light contrast ratio.
- It is easily installed and aligned. One side of the sensing zone only need be wired.

It also has certain limitations:

- Its operating range is half that of a nonpolarized sensor since much of the signal is lost in the polarizing filters.
- The sensor can be fooled by shiny objects wrapped with shrink-wrap material.

3.1.5 Proximity (Diffuse-Reflection) Detection

Proximity detection is similar to reflex detection, because the light source and detector elements are mounted on the same side (Fig. 3.11). In this application, the sensors detect light that is bounced off the target object, rather than the breaking of the beam. The detection zone is controlled by the type, texture, and composition of the target object's surface.

Focused proximity sensors are a special type of proximity sensor where the source and detector are focused to a point in front of the sensor (Fig. 3.12). Focused proximity sensors can detect extremely small objects, or look into holes or cavities in special applications.

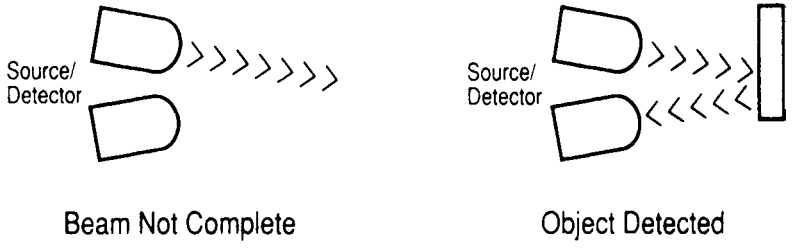


Figure 3.11 Proximity detection.

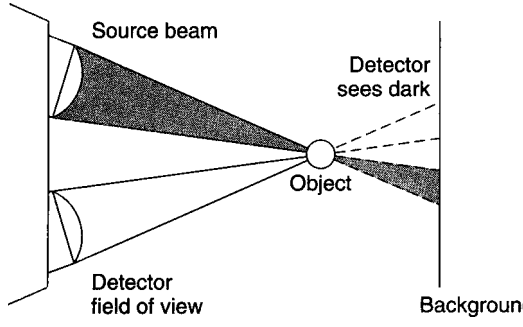


Figure 3.12 Focused proximity sensor.

Background objects will not false-trigger a focused proximity sensor since they are "cross-eyed" and cannot see past a certain point.

Advantages of the focused proximity sensor are:

- Its installation and alignment are simple. The control circuit can be wired through only one side of the sensing zone.
- It can detect differences in surface reflectivity.

It also has certain limitations:

- It has a limited sensing range.
- The light/dark contrast and sensing range depend on the target object's surface reflectivity.

3.1.6 Automated Guided Vehicle System

The wide sensing field of diffuse reflective photoelectric sensors makes them ideally suited for use as obstacle detection sensors. Sensing distance and field width are adjustable to form the obstacle detection zone vital for manufacturing operations.

Two outputs, *near* and *far*, provide switching at two separate sensing distances that are set by corresponding potentiometers.

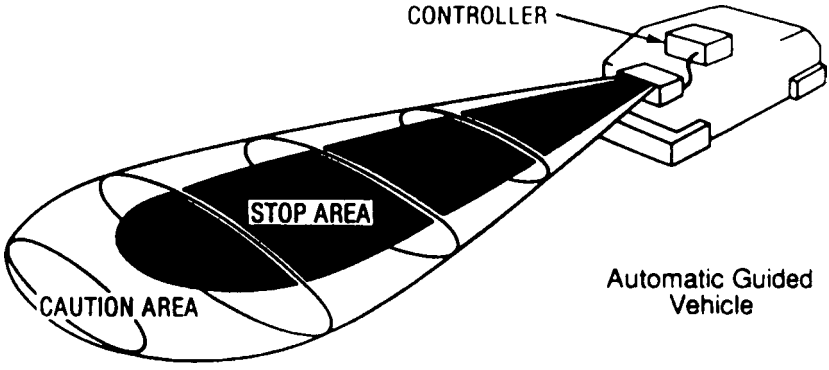


Figure 3.13 Application of diffuse reflective photoelectric sensor in automated guided vehicle system.

The sensing distance of the far output is adjustable up to 3 m maximum. The sensing distance of the near output is adjustable from 30 to 80 percent of the far output. Indicators include a red LED that glows with the near output ON, and a yellow LED that glows with the far output ON.

An ideal application for this family of sensors is the automated guided vehicle (AGV), which requires both slow-down and stop controls to avoid collisions when obstacles enter its path (Fig. 3.13).

A modulated infrared light source provides immunity to random operation caused by ambient light. Additionally, unwanted sensor operation caused by adjacent sensor interference (crosstalk) is also eliminated through the use of multiple-position modulated frequency adjustments.

3.2 Fiber Optics

Fiber optics has greatly expanded the applications of photoelectric sensors. Fiber optics uses bundles of thin plastic or glass fibers that operate on a principle discovered in 1854 by John Tyndahl. When Tyndahl shined a beam of light through a stream of water, instead of emerging straight from the stream of water as might be expected, the light tended to bend with the water as it arced towards the floor. Tyndahl discovered that the light was transmitted along the stream of water. The light rays inside the water bounced off the internal walls of the water and were thereby contained (Fig. 3.14). This principle has come to be known as *total internal reflection*.

Industry has since discovered that the principle of total internal reflection also applies to small-diameter glass and plastic fibers, and this has lead to rapid growth of applications throughout the industry. Because optical fibers are small in diameter and flexible, they can bend and twist in confined places. Also, because they contain no

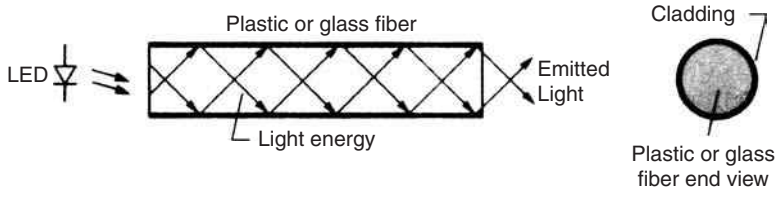


Figure 3.14 Total internal reflection.

electronics, they can operate in much higher temperatures—as high as 400°F—and in areas of high vibration. They are limited by sensing distances, which typically are 80 mm in the proximity mode and 400 mm in the through-beam mode. Also, because of their small sensing area, optical fibers can be fooled by a small drop of water or dirt over the sensing area.

Fiber optics is used to transmit data in the communication field and to transmit images or light in medicine and industry. Photoelectric controls use fiber optics to bend the light from the LED source and return it to the detector so sensors can be placed in locations where common photoelectric sensors cannot be applied.

Fiber optics used with photoelectric controls consists of a large number of individual glass or plastic fibers that are sheathed in suitable material for protection. The optical fibers used with photoelectric controls are usually covered by either PVC or stainless-steel jackets. Both protect the fibers from excessive flexing and the environment (Fig. 3.15).

Optical fibers are transparent fibers of glass or plastic used for conducting and guiding light energy. They are used in photoelectric sensors as "light pipes" to conduct sensing light into and out of a sensing area.

Glass optical-fiber assemblies consist of a bundle of 0.05-mm-diameter discrete glass optical fibers housed within a flexible sheath. Glass optical fibers are also able to withstand hostile sensing environments. Plastic optical-fiber assemblies consist of one or two acrylic monofilaments in a flexible sheath.

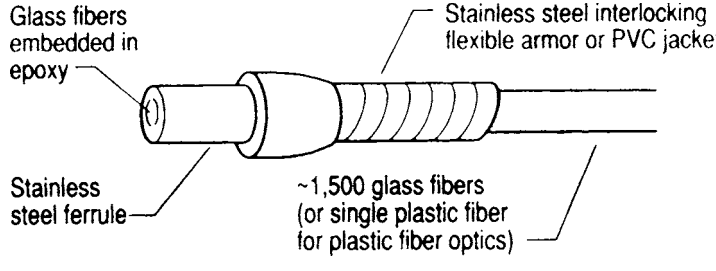


Figure 3.15 Jacketed glass fibers.

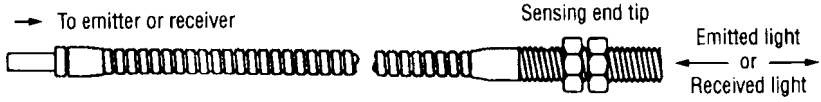


Figure 3.16 Individual fiber-optic assembly.

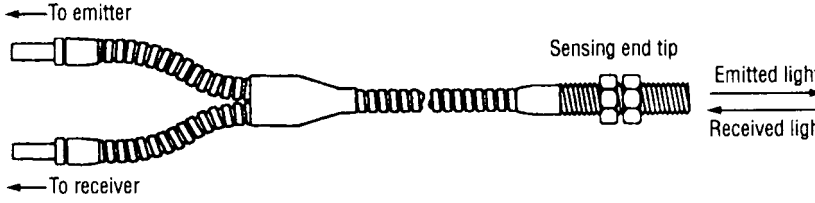


Figure 3.17 Bifurcated fiber-optic assembly.

There are two basic styles of fiber-optic assemblies: (1) individual fiber optics (Fig. 3.16) and (2) bifurcated fiber optics (Fig. 3.17).

Individual fiber-optic assemblies guide light from an emitter to a sensing location, or to a receiver from a sensing location. Bifurcated fibers use half their fiber area to transmit light and the other half to receive light.

3.2.1 Individual Fiber Optics

A fiber-optic assembly having one control end and one sensing end is used for piping photoelectric light from an emitter to the sensing location or from the sensing location back to a receiver. It is usually used in pairs in the opposed sensing mode, but can also be used side by side in the diffuse proximity mode or angled for the specular reflection or mechanical convergent mode.

3.2.2 Bifurcated Fiber Optics

A bifurcated fiber-optic assembly is branched to combine emitted light with received light in the same assembly. Bifurcated fibers are used for diffused (divergent) proximity sensing, or they may be equipped with a lens for use in the retroreflective mode.

Three types of sensing modes are used in positioning a sensor so the maximum amount of emitted energy reaches the receiver sensing element:

- Opposed sensing mode (Fig. 3.18)
- Retroreflective sensing mode (Fig. 3.19)
- Proximity (diffused) sensing mode (Fig. 3.20)

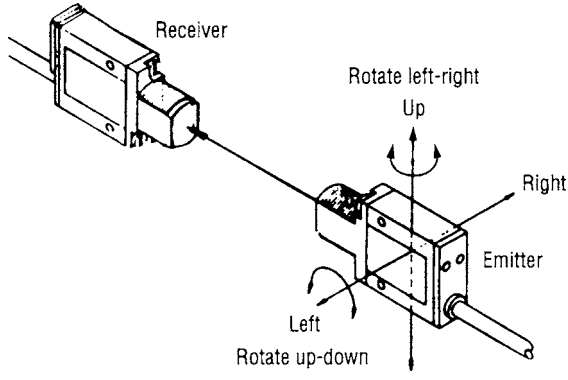


Figure 3.18 Opposed sensing mode. For alignment, move emitter or receiver up-down, left-right, and rotate.

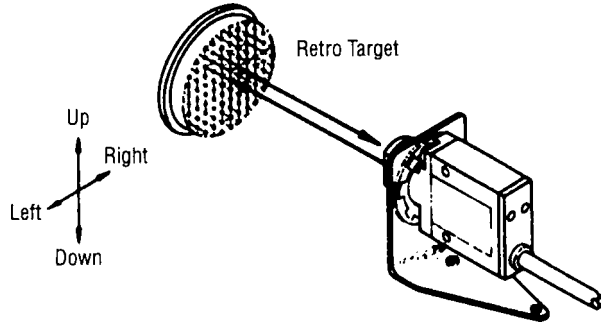


Figure 3.19 Retroreflective sensing mode. For alignment, move target up-down, left-right.

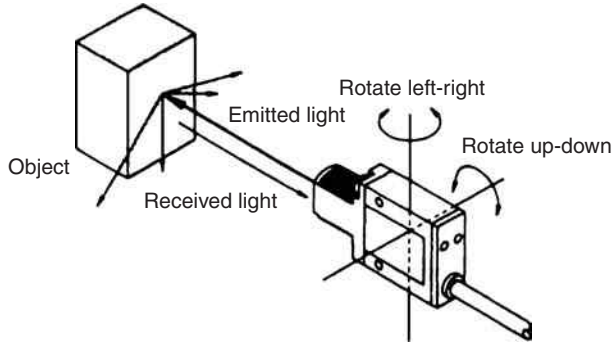


Figure 3.20 Proximity sensing mode. For alignment, rotate up-down, left-right.

Opposed sensing is the most efficient photoelectric sensing mode and offers the highest level of optical energy to overcome lens contamination, sensor misalignment, and long scanning ranges. It is also often referred to as *direct scanning* and sometimes called the *beam break* mode.

The addition of fiber optics to photoelectric sensing has greatly expanded the application of photoelectric devices. Because they are small in diameter and flexible, optical fibers can bend and twist into tiny places formerly inaccessible to bulky electronic devices.

Optical fibers operate in the same sensing modes as standard photoelectric controls—through-beam, proximity, and reflex. The sizes and shapes of sensing tips have been developed to accommodate many applications.

Optical fibers have a few drawbacks:

- Limited sensing distance. Typical sensing distance in the proximity mode is 80 mm; 380 mm for the through-beam mode.
- Typically more expensive than other photoelectric sensing controls.
- Easily fooled by a small drop of water or dirt over the sensing surface.

Optical fibers' advantages:

- Sensing in confined places.
- Ability to bend around corners.
- No electronics at sensing point.
- Operation at high temperatures (glass).
- Total immunity from electrical noise and interference.
- Easily cut to desired lengths (plastic).

3.3 Optical Fiber Parameters

The most important parameters affecting optical-fiber performance are:

- Excess gain
- Background suppression
- Contrast
- Polarization

3.3.1 Excess Gain

Excess gain is the measure of energy available between the source and the detector to overcome signal loss due to dirt or contamination.

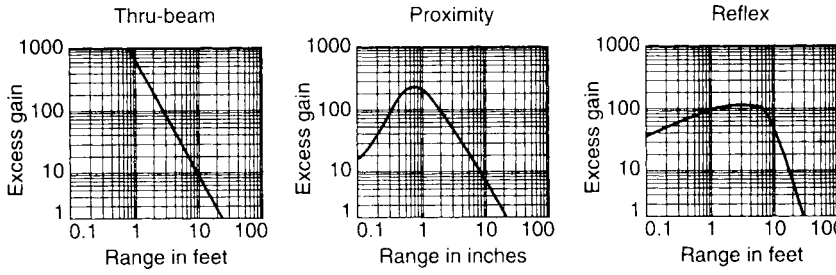


Figure 3.21 Excess gain curves.

Excess gain is the single most important consideration in choosing a photoelectric control in manufacturing. It is the extra punch that the sensor has available within its detecting region.

By definition, excess gain is the ratio of the amount of light the detector sees to the minimum amount of light required to trip the sensor. This ratio is depicted graphically for all photoelectric sensors. In Fig. 3.21, excess gain is plotted along the vertical logarithmic axis, starting at 1, the minimum amount of light required to trigger the detector. Every point above 1 represents the amount of light required to trigger the photoelectric control—the excess gain.

Often, the standard of comparison for choosing between different photoelectric sensors is range. Actually, more important to most applications is the excess gain. For a typical application, the higher the excess gain within the sensing region, the more likely the application will work. It is the extra margin that will determine whether the photoelectric control will continue to operate despite the buildup of dirt on the lens or the presence of contamination in the air.

Example. An application requires detecting boxes on a conveyer in a filthy industrial environment (Fig. 3.22). The boxes will pass about 2 to 5 mm from

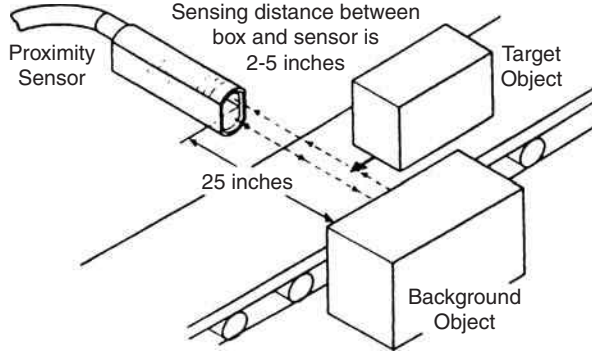


Figure 3.22 Box detection.

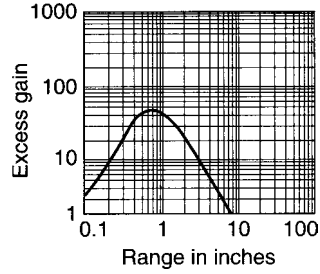


Figure 3.23
Excess gain curve for sensor 1.

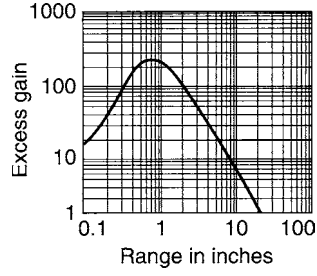


Figure 3.24
Excess gain curve for sensor 3.

the sensor as they move along the conveyer at the sensing location. Given a choice between the two proximity sensors (whose excess gain curves appear in Figs. 3.23 and 3.24), which photoelectric control should be selected for this application?

If the decision were based solely on specified range, the unit described in Fig. 3.23 would be selected. However, if units were installed in this application, it might fail after a short time in operation. Over time, contaminants from the environment would settle on the lens, decreasing the amount of light the sensor sees. Eventually, enough lens contamination would accumulate that the photoelectric control would not have enough excess gain to overcome the signal loss created by the coating, and the application would fail.

A better choice for this application would be the unit represented in Fig. 3.24. It delivers much more excess gain in the operating region required for this application and will therefore work much more successfully than the other unit.

3.3.2 Background Suppression

Background suppression enables a diffuse photoelectric sensor to have high excess gain to a predetermined limit and insufficient excess gain beyond that range, where it might pick up objects in motion and yield a false detection. By using triangular ranging, sensor developers have created a sensor that emits light that reflects on the detector from two different target positions. The signal received from the more distant target is subtracted from that of the closer target, providing high excess gain for the closer target.

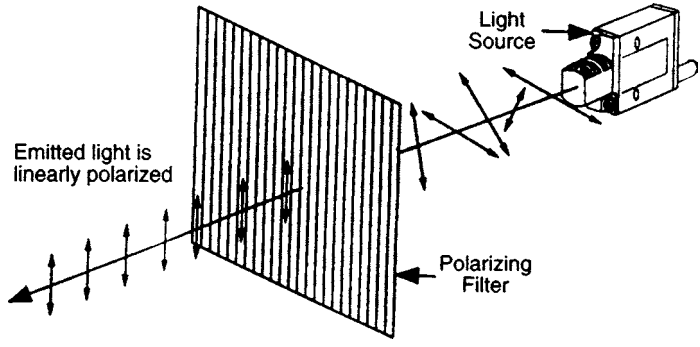


Figure 3.25 Polarization.

3.3.3 Contrast

Contrast measures the ability of a photoelectric control to detect an object; it is the ratio of the excess gain under illumination to the excess gain in the dark. All other things being equal, the sensor that provides the greatest contrast ratio should be selected. For reliable operation, a ratio 10:1 is recommended.

3.3.4 Polarization

Polarization is used in reflection sensors in applications where shiny surfaces, such as metal or shrink-wrapped boxes, may trigger the control falsely. The polarizer passes light along only one plane (Fig. 3.25), and the corner-cube reflectors depolarize the light as it passes through the plastic face of the retroreflector (Fig. 3.10). Only light that has been rotated by the corner-cube retroreflector can pass through the polarizer, whereas light that bounces off other shiny objects cannot.

Like regular reflex photoelectric sensors, polarized sensors have a high light/dark contrast ratio and are simple to install and align. However, the polarizers do limit the sensor's operating range because light is lost passing through them.

3.4 Inductive Proximity Sensors—Noncontact Metal Detection

Inductive proximity sensors are another common choice for position sensing. An inductive proximity sensor consists of four basic elements:

- The sensor, which comprises a coil and ferrous core
- An oscillator circuit
- A detector circuit
- A solid-state output

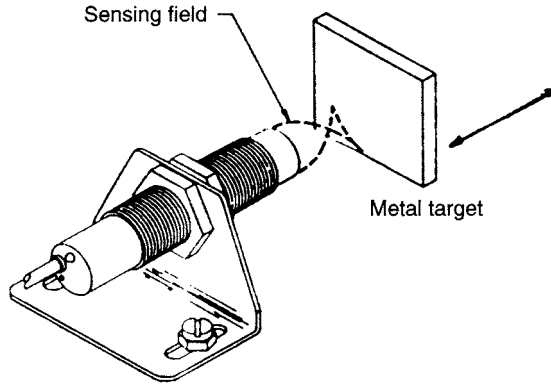


Figure 3.26 Inductive proximity sensor.

In the circuitry in Fig. 3.26, the oscillator generates an electromagnetic field that radiates from the sensor's face. This field is centered around the axis and detected by the ferrite core to the front of the sensor assembly. When a metal object enters the electromagnetic field, eddy currents are induced in the surface of the target. This loads in the oscillator circuit, which reduces its amplitude.

The detector circuit detects the change in the oscillator amplitude and, depending on its programming, switches ON and OFF at a specific oscillator amplitude. The sensing circuit returns to its normal state when the target leaves the sensing area and the oscillator circuit regenerates.

The nominal sensing range of inductive proximity sensors is a function of the diameter of the sensor and the power that is available to generate the electromagnetic field. This is subject to a manufacturing tolerance of ±10 percent, as well as a temperature drift tolerance of ±10 percent. The target size, shape, and material will have an effect on the sensing range. Smaller targets will reduce the sensing range, as will targets that are not flat or are made of nonferrous material.

Basically two types of inductive proximity sensors are used: (1) shielded and (2) nonshielded. The shielded version has a metal cover around the ferrite core and coil assembly. This focuses the electromagnetic field to the front of the sensor and allows it to be imbedded in metal without influencing the sensing range. The nonshielded sensor can sense on the side as well as in front of a sensor. It requires a nonmetallic area around the sensor to operate correctly.

Inductive proximity sensors have several benefits:

- *High repeatability*. Visibility of the environment is not an issue, since inductive proximity sensors can sense only electromagnetic fields. Therefore, environments from dirt to sunlight pose no problem for inductive proximity sensors. Also, because they are noncontact sensors, nothing wears.

- *Shock and vibration resistance*.
- *Versatile connections*. They can connect two or three wires with an AC, AC/DC, or DC power supply and up to four-wire DC connections.
- Wide operating temperature range. They operate between –20 to +70°C, ±10 percent.
- Very fast response, particularly in the DC models. These sensors can detect presence, send a signal, and reset in 50 μs (2000 times per second) in DC models.

Inductive proximity sensors are generally limited by their sensing distances and the material they sense. The effective range is limited to about 25 mm for most models and can be extended to only about 100 mm with the large models.

3.5 Limit Switches—Traditional Reliability

Limit switches are mechanical position-sensing devices that offer simplicity, robustness, and repeatability to processes. Mechanical limit switches are the oldest and simplest of all presence- or position-sensing devices: contact is made and a switch is engaged. This simplicity contributes generally to the cost advantage of limit switches. Yet, they can provide the control capabilities and versatility demanded in today's error-free manufacturing environment. The key to their versatility is the various forms they can take in the switch, actuating head, and lever operator. Two-step dual-pole limit switches can detect and count two products of different sizes and can provide direct power control to segregate or process the items differently. The lever operator will rotate 10° to activate one set of contacts and 20° to activate another set. Because of the high amperage they can handle, limit switches can control up to ten contacts from the movement of a single lever.

They are easy to maintain because the operator can hear the operation of the switch and can align it easily to fit the application. They are also robust. Limit switches are capable of handling an inrush current ten times that of their steady state current rating. They have rugged enclosures and also prewiring that uses suitable strain-relief bushings to enable the limit switch to retain cables with 500 to 600 pounds of force on them. Limit switches can also handle direct medium-power switching for varying power factors and inrush stresses. For example, they can control a multihorsepower motor without any interposing starter, relay, or contactor.

Reliability is another benefit. Published claims for repeat accuracy for standard limit switches vary from within 0.03 mm to within 0.001 mm over the temperature range of –4 to +200°F. Limit switches dissipate energy spikes and rarely break down under normal mode surges.

They will not be affected by electromagnetic interference (EMI), and there are no premature responses in the face of EMI. However, because they are mechanical devices, limit switches face physical limitations that can shorten their service life even though they are capable of several million operations. Also, heavy sludge, chips, or coolant can interfere with their operation.

3.6 Factors Affecting the Selection of Position Sensors

In selecting a position sensor, several key factors should be considered:

- *Cost.* Both initial purchase price and life-cycle cost must be considered.
- *Sensing distance.* Photoelectric sensors are often the best selection when sensing distances are longer than 25 mm. Photoelectric sensors can have sensing ranges as long as 300,000 mm for outdoor or extremely dirty applications, down to 25 mm for extremely small parts or for ignoring background. Inductive proximity sensors and limit switches, on the other hand, have short sensing distances. The inductive proximity sensors are limited by the distance of the electromagnetic field—less than 25 mm for most models—and limit switches can sense only as far as the lever operator reaches.
- *Type of material.* Inductive proximity sensors can sense only ferrous and nonferrous materials, whereas photoelectric and limit switches can detect the presence of any solid material. Photoelectric sensors, however, may require a polarizer if the target's surface is shiny.
- *Speed.* Electronic devices using DC power are the fastest—as fast as 2000 cycles per second for inductive proximity models. The fastest-acting limit switches can sense and reset in 4 ms or about 300 times per second.
- *Environment.* Proximity sensors can best handle dirty, gritty environments, but they can be fooled by metal chips and other metallic debris. Photoelectric sensors will also be fooled or left inoperable if they are fogged or blinded by debris.
- *Types of voltages, connections, and requirements of the device's housing.* All three types can accommodate varying requirements, but the proper selection must be made in light of the power supplies, wiring schemes, and environments.
- *Third-party certification.* Underwriters Laboratories (UL), National Electrical Manufacturers Association (NEMA), International Electrotechnical Commission (IEC), Factory Mutual, Canadian Standards Association (CSA), and other organizations impose requirements for safety, often based on

the type of application. The certification will ensure the device has been tested and approved for certain uses.

- *Intangibles.* These can include the availability of application support and service, the supplier's reputation, local availability, and quality testing statements from the manufacturer.

3.7 Wavelengths of Commonly Used Light-Emitting Diodes

An LED is a semiconductor that emits a small amount of light when current flows through it in the forward direction. In most photoelectric sensors, LEDs are used both as emitters for sensing beams and as visual indicators of alignment or output status for a manufacturing process. Most sensor manufacturers use visible red, visible green, or infrared (invisible) LEDs (Fig. 3.4*b*). This simple device plays a significant part in industrial automation. It provides instantaneous information regarding an object during the manufacturing operation. LEDs, together with fiber optics, allow a controller to direct a multitude of tasks, simultaneously or sequentially.

3.8 Sensor Alignment Techniques

A sensor should be positioned so the maximum amount of emitted energy reaches the receiver element in one of three different modes:

- Opposed sensing mode
- Retroreflective sensing mode
- Proximity (diffuse) sensing mode

3.8.1 Opposed Sensing Mode

In the opposed sensing mode, the emitter and receiver are positioned opposite each other so that the light from the emitter shines directly at the receiver. An object then breaks the light beam that is established between the two. Opposed sensing is always the most reliable mode.

3.8.2 Retroreflective Sensing Mode

Retroreflective sensing is also called the *reflex* mode or simply the *retro* mode. A retroreflective photoelectric sensor contains both emitter and receiver. A light beam is established between the sensor and a special retroreflective target. As in opposed sensing, an object is sensed when it interrupts this beam.

Retro is the most popular mode for conveyer applications where the objects are large (boxes, cartons, etc.), where the sensing environment

is relatively clean, and where scanning ranges are typically a few meters in length. Retro is also used for code-reading applications. Automatic storage and retrieval systems and automatic conveyer routing systems use retroreflective code plates to identify locations and/or products.

3.8.3 Proximity (Diffuse) Sensing Mode

In the proximity (diffuse) sensing mode, light from the emitter strikes a surface of an object at some arbitrary angle and is diffused from the surface at all angles. The object is detected when the receiver captures some small percentage of the diffused light. Also called the *direct reflection* mode or simply the photoelectric *proximity* mode, this method provides direct sensing of an object by its presence in front of a sensor. A variation is the ultrasonic proximity sensor, in which an object is sensed when its surface reflects a sound wave back to an acoustic sensor.

3.8.4 Divergent Sensing Mode

The divergent sensing mode is a variation of the diffuse photoelectric sensing mode in which the emitted beam and the receiver's field of view are both very wide. Divergent mode sensors (Fig. 3.27) have loose alignment requirements, but have a shorter sensing range than diffuse mode sensors of the same basic design. Divergent sensors are particularly useful for sensing transparent or translucent materials or for sensing objects with irregular surfaces (e.g., webs that flutter). They are also used effectively to sense objects with very small profiles, such as small-diameter thread or wire, at close range.

All unlensed bifurcated optical fibers are divergent. The divergent mode is sometimes called the *wide-beam diffuse* (or *proximity*) mode.

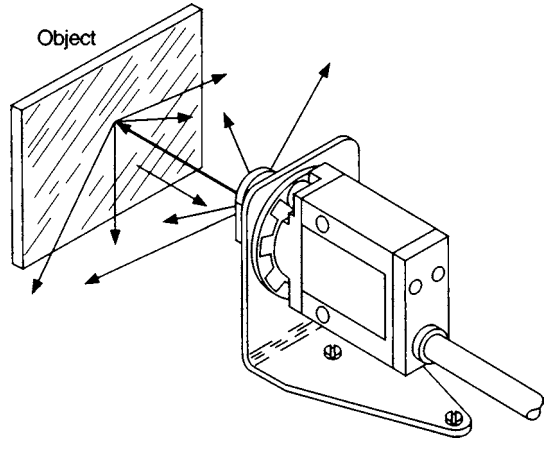


Figure 3.27 Divergent sensing mode.

3.8.5 Convergent Sensing Mode

The convergent sensing mode is a special variation of diffuse mode photoelectric proximity sensing that uses additional optics to create a small, intense, and well-defined image at a fixed distance from the front surface of the sensor lens (Fig. 3.28). Convergent beam sensing is the first choice for photoelectric sensing of transparent materials that remain within a sensor's depth of field. It is also called the *fixed-focus proximity* mode.

3.8.6 Mechanical Convergence

In mechanical convergence (Fig. 3.29), an emitter and a receiver are simply angled toward a common point ahead of the sensor. Although less

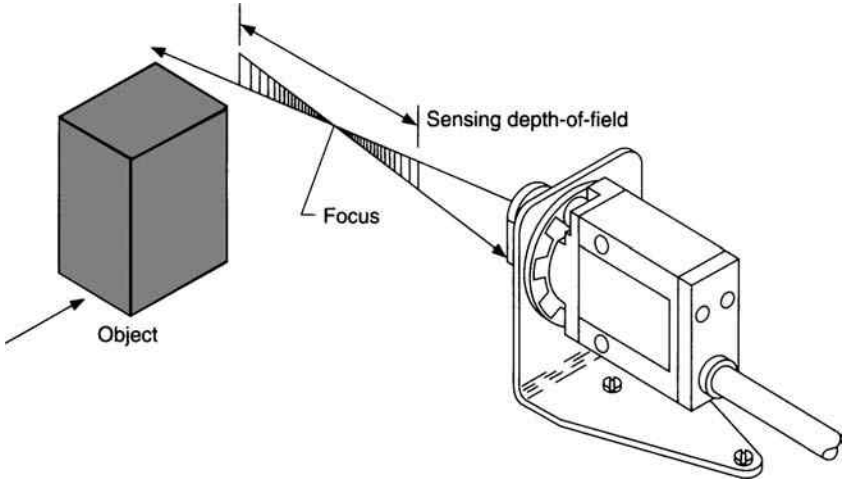


Figure 3.28 Convergent sensing mode.

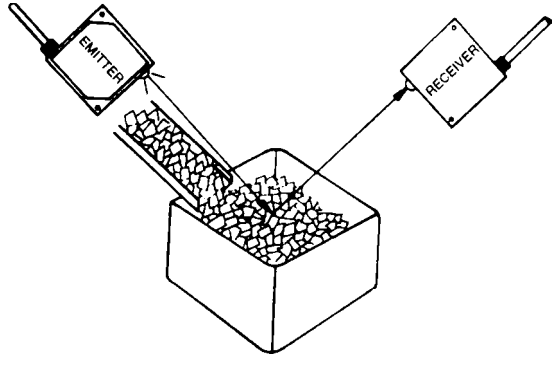


Figure 3.29 Mechanical convergence.

precise than the optical convergent-beam sensing mode, this approach to reflective sensing uses light more efficiently than diffuse sensing and gives a greater depth of field than true optical convergence.

Mechanical convergence may be customized for an application by mounting the emitter and the receiver to converge at the desired distance. Depth of field is controlled by adjusting the angle between the emitter and the receiver.

3.9 Fiber Optics in Industrial Communication and Control

The application of fiber optics to industrial information transfer is a natural extension of the current commercial uses of this technology in high-data-rate communications. While the primary advantage of fiber optics in traditional application areas has been extremely reliable communication at high rates, exceeding 1 Gbit/s over distances exceeding 100 km, other intrinsic features of the technology are more important than data rate and distance capability in industrial uses.

The physical mechanism of light propagating through a glass fiber has significant advantages that enable sensors to carry data and plant communications successfully and in a timely manner—a fundamental condition that must be constantly maintained in a computer-integrated manufacturing environment:

- The light signal is completely undisturbed by electrical noise. This means that the fiber-optic cables can be laid wherever convenient without special shielding. Fiber-optic cables and sensors are unaffected by electrical noise when placed near arc welders, rotating machinery, electrical generators, and so on, whereas in similar wired applications, even the best conventional shielding methods are often inadequate.
- Fiber-optic communication is devoid of any electrical arcing or sparking, and thus can be used successfully in hazardous areas without danger of causing an explosion.
- The use of a fiber link provides total electrical isolation between terminal points on the link. Over long plant distances, this can avoid troublesome voltage or ground differentials and ground loops.
- A fiber-optic system can be flexibly configured to provide additional utility in existing hardware.

3.10 Principles of Fiber Optics in Communications

An optical fiber (Fig. 3.30) is a thin strand composed of two layers: an inner core and an outer cladding. The core is usually constructed of

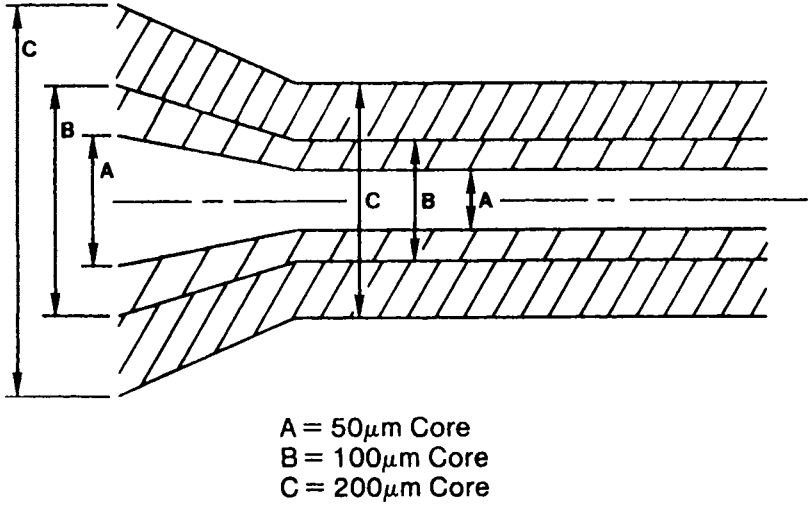


Figure 3.30 Structure of optical fiber.

glass, and the cladding structure composed of glass or plastic. Each layer has a different index of refraction, the core being higher. The difference between the index of refraction of the core material n_1 and that of the surrounding cladding material n_2 causes rays of light injected into the core to continuously reflect back into the core as they propagate down the fiber.

The light-gathering capability of the fiber is expressed in terms of its numerical aperture NA, the sine of the half angle of the acceptance cone for that fiber. Simply stated, the larger the NA, the easier it is for the fiber to accept light (Fig. 3.31).

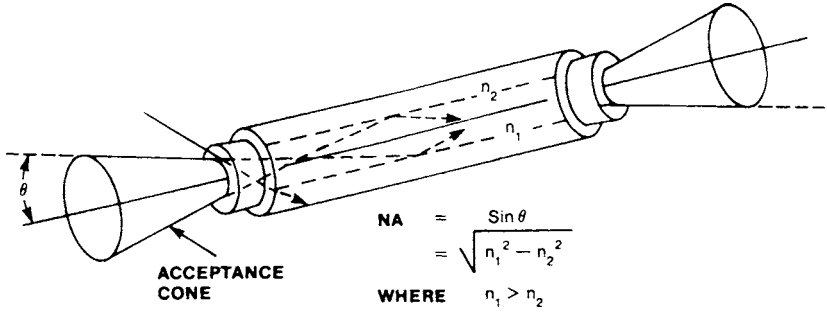


Figure 3.31 Numerical aperture (NA) of an optical fiber.

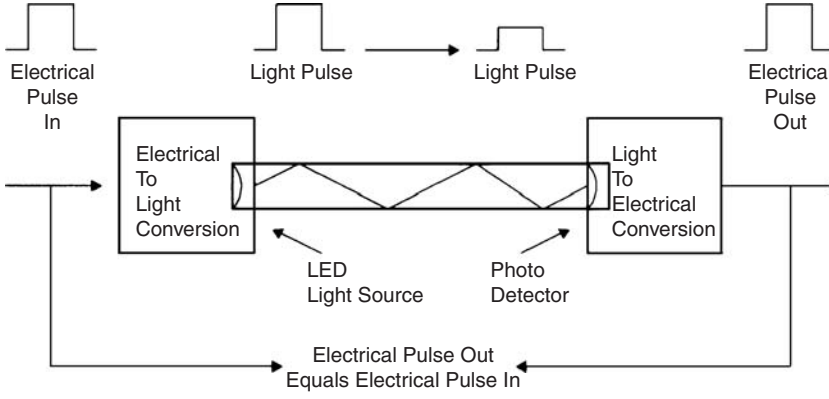


FIGURE 3.32 Fiber-optic communication system.

3.11 Fiber-Optic Information Link

A fiber-optic communication system (Fig. 3.32) consists of:

- A light source (LED or laser diode) pulsed by interface circuitry and capable of handling data rates and voltage levels of a given magnitude.
- A detector (photodiode) that converts light signals to electrical signals and feeds interface circuitry to re-create the original electrical signal.
- Fiber-optic cables between the light source and the detector (called the *transmitter* and the *receiver*, respectively).

It is usual practice for two-way communication to build a transmitter and receiver into one module and use a duplex cable to communicate to an identical module at the end of the link. The 820- to 850-nm range is the most frequent for low-data-rate communication, but other wavelengths (1300 and 1550 nm) are used in long-distance systems. Fiber selection must always take transmission wavelength into consideration.

3.12 Configurations of Fiber Optics

The selection of optical fiber plays a significant part in sensor performance and information flow. Increased fiber diameter results in higher delivered power, but supports a lower bandwidth. A fiber with a 200-μm core diameter (glass core, silicone plastic cladding), by virtue of its larger diameter and acceptance angle, can transmit five to seven times more light than a 100-μm core fiber, or up to 30 times

more light than a 40-μm fiber, the historical telecommunication industry standard.

Computer data communications systems commonly employ 62.5-μm fiber because of its ability to support very high data rates (up to 100 Mbaud) while offering increased power and ease of handling. Factory bandwidth requirements for most links are typically one or more orders of magnitude. Therefore, fiber core size may be increased to 200 μm to gain the benefits of enhanced power and handling ease, the decrease in bandwidth being of no consequence (Fig. 3.31).

3.12.1 Optical Power Budget

An optical power budget examines the available optical power and how it is used and dissipated in a fiber-optic system. It is important to employ the highest possible optical power budget for maximum power margin over the detector requirement. A budget involves four factors:

- Types of light source
- Optical fiber acceptance cone
- Receiver sensitivity
- Splice, coupling, and connector losses

Laser-diode sources are generally not economically feasible or necessary in industrial systems. Light-emitting diodes are recommended for industrial applications. Such systems are frequently specified with transmitted power at 50 μW or greater if 200-μm-core fiber is used.

Successful communication in industry and commercial applications is determined by the amount of light energy required at the receiver, specified as *receiver sensitivity*. The higher the sensitivity, the less light required from the fiber. High-quality systems require power only in the hundreds of nanowatts to low microwatts range.

Splice losses must be low so that as little light as possible is removed from the optical-fiber system. Splice technology to repair broken cable is readily available, permitting repairs in several locations within a system in a short time (minutes) and causing negligible losses. Couplers and taps are generally formed through the process of glass fusion and operate on the principle of splitting from one fiber to several fibers. New active electronic couplers replenish light as well as distribute the optical signal.

An example of an optical power budget follows:

1. The optical power injected into a 200-μm-core fiber is 200 μW (the same light source would inject approximately 40 μW into a 100-μm core fiber).

2. The receiver sensitivity is 2 μW.
3. The receiver budget (dB) is calculated as:

$$\begin{aligned} \text{dB} &= 10 \log [(\text{available light input})/(\text{required light output})] \\ &= 10 \log [(200\ \mu\text{W})/(2\ \mu\text{W})] \\ &= 10 \log 100 \\ &= 20\ \text{dB} \end{aligned}$$

4. Three major sources of loss are estimated as:
   - 2 to 3 dB loss for each end connector
   - 1 to 2 dB loss for each splice
   - 1 dB/150 m loss for fiber of 200 μm diameter

Most manufacturers specify the optical power budget and translate this into a recommended distance.

3.12.2 Digital Links—Pulsed

The one-for-one creation of a light pulse for an electrical pulse is shown in Fig. 3.32. This represents the simplest form of data link. It does not matter what format and signal level the electrical data takes (e.g., whether IEEE RS-232 or RS-422 standard format or CMOS or TTL logic level), as long as the interface circuitry is designed to accept them at its input or reproduce them at the output. Voltage conversion may be achieved from one end of the link to the other, if desired, through appropriate interface selection.

The light pulses racing down the fiber are independent of electrical protocol. Several design factors are relevant to these and other types of data links as follows:

- *Minimum output power.* The amount of light, typically measured in microwatts, provided to a specific fiber size from the data link's light source
- *Fiber size.* Determined from the data link's light source.
- *Receiver sensitivity.* The amount of light, typically measured in microwatts or nanowatts, required to activate the data link's light detector.
- *Data rate.* The maximum rate at which data can be accurately transmitted.
- *Bit error rate (BER).* The frequency with which a light pulse is erroneously interpreted (for example, 10^{-9} BER means no more than one of 10^9 pulses will be incorrect).
- *Pulse-width distortion.* The time-based disparity between input and output pulse widths. The simple pulse link is also

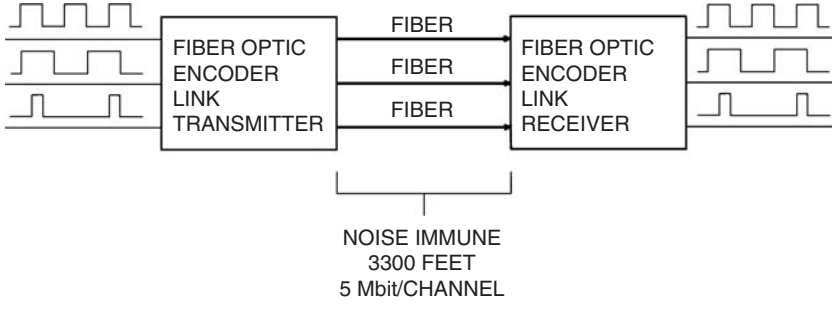


Figure 3.33 Three-channel optical link.

the basic building block for more complex links. Figure 3.33 provides an example of a 5-Mbit three-channel link used to transmit encoder signals from a servomotor to a remote destination.

3.12.3 Digital Links—Carrier-Based

A carrier-based digital link is a system in which the frequency of the optical carrier is varied by a technique known as *frequency-shift keying* (*FSK*). Figure 3.34 illustrates the modulation concept; two frequencies are employed to create the logic 0 and 1 states. This scheme is especially useful in systems where electrical "handshaking" (confirmation of reception and acceptance) is employed. Presence of the optical carrier is the equivalent of the handshake signal, with the data signal presented by frequency.

Figure 3.35 illustrates a system where the logic of the fiber-optic line driver recognizes the optical carrier to create a handshake between terminal and processor.

Additionally, since the processor is capable of recognizing only one terminal, the carrier is controlled to deny the handshake to all other terminals once one terminal is actively on-line to the processor.

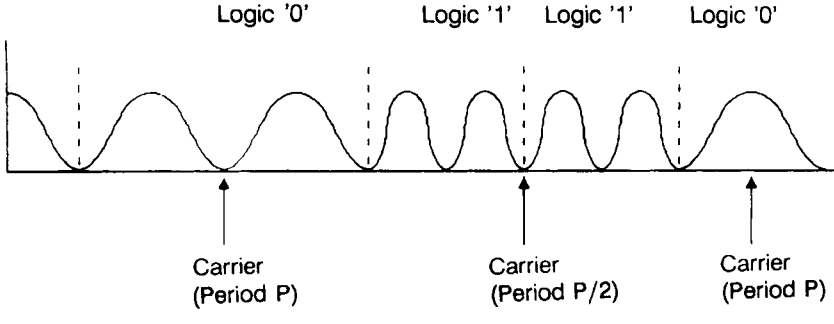


Figure 3.34 Modulation concept.

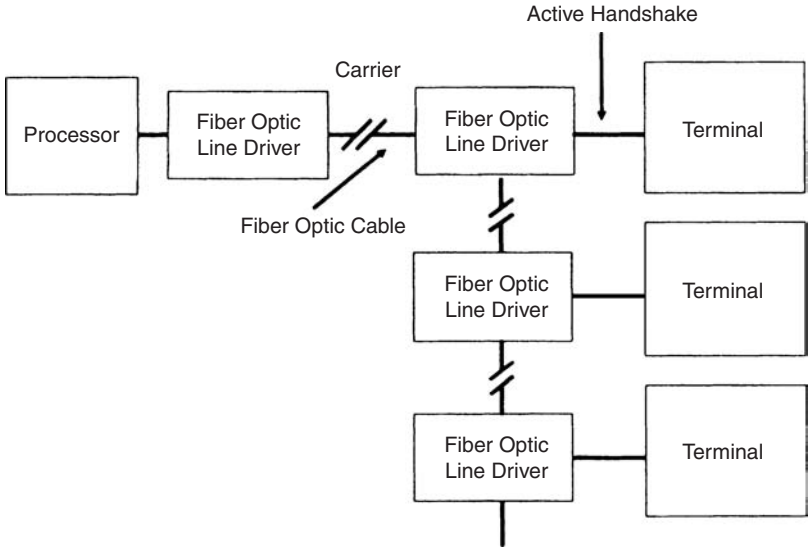


Figure 3.35 System employing optical carrier.

3.12.4 Analog Links

It is well-recognized that, in motion control and process measurement and control, transmitting analog information without distortion is important. Analog information can be dealt with in several ways using fiber optics.

Analog data cannot easily be transmitted through light intensity variation. A number of external factors—such as light source variation, bending losses in cable, and connector expansion with temperature—can affect the amount of raw light energy reaching the detector. It is not practical to compensate for all such factors and deliver accurate analog data. A viable method of transmitting data is to use an unmodulated carrier whose frequency depends on the analog signal level. A more advanced means is to convert the analog data to digital data, where accuracy also is determined by the number of bits used, multiplex the digital bits into one stream, and use the pulsed digital link approach.

Figure 3.36 illustrates a link in which this last approach is used to produce both digital and analog forms of the data at the output.

3.12.5 Video Links

Long-distance video transmission in industrial situations is easily disrupted by radiated noise and lighting. Repeaters and large-diameter coaxial cables are often used for particularly long runs. The use of fiber optics as a substitute for coaxial cable allows propagation of noise-free video over long distances. Either an intensity- or

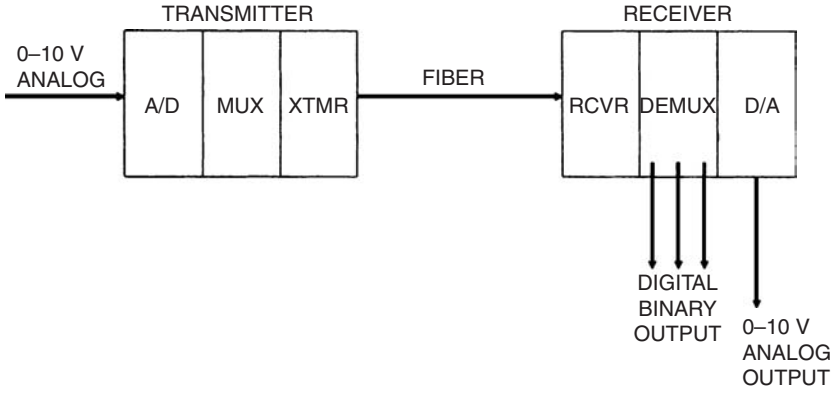


Figure 3.36 Analog and digital data transmission.

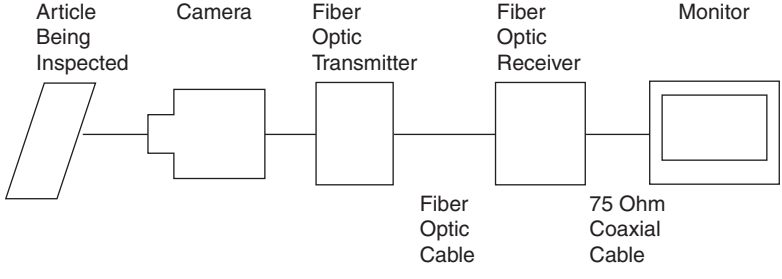


Figure 3.37 Fiber-optic video link.

frequency-modulated optical carrier signal is utilized as the transmission means over fiber. With intensity-modulated signals, it is mandatory that some sort of automatic gain control be employed to compensate for light degradation due to varying cable losses, splices, and so on. Figure 3.37 illustrates a typical fiber-optic video link in a machine-vision application.

3.12.6 Data Bus Networks

Wiring a system often causes serious problems for designers and communication system integrators regarding the choice of topology. The basic difference between fiber and wiring is that one normally does not splice or tap into fiber as one would with coaxial or twin axial cable to create a drop point.

3.12.6.1 Daisy Chain Data Bus

The simplest extension of a point-to-point data link is described in Fig. 3.38. It extends continuously from one drop point (node) to the next by using each node as a repeater. The fiber-optic line driver

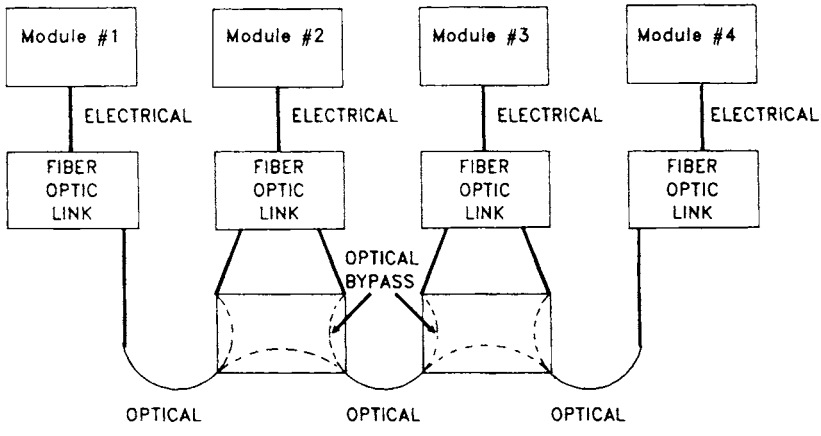


Figure 3.38 Point-to-point data link.

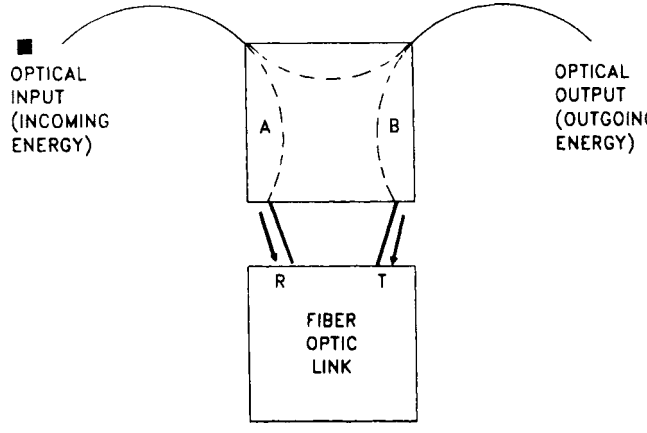


Figure 3.39 Daisy chain data bus.

illustrated in Fig. 3.35 is such a system, providing multiple access points from remote terminals to a programmable controller processor. A system with several repeater points is vulnerable to the loss of any repeater, and with it all downstream points, unless some optical bypass scheme is utilized. Figures 3.38 and 3.39 exhibit such a scheme.

3.12.6.2 Ring Coupler

A preferred choice among several current fiber-optic system designs is the token-passing ring structure.

Signals are passed around the ring, with each node serving to amplify and retransmit. Care must be taken to provide for a node becoming nonoperational. This is usually handled by using some type of bypass switching technique, given that the system provides

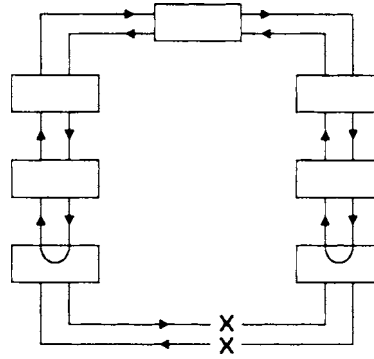


Figure 3.40
Ring coupler.

sufficient optical power to tolerate a bypassed repeater. Another contingency method is to provide for the transmitting node to read its own data coming around the ring, and to retransmit in the other direction if necessary, as illustrated in Fig. 3.40. Yet another is to provide for a second pair of fibers paralleling the first, but routed on a physically different path.

3.12.6.3 The Passive Star Coupler

Certain systems have attempted to utilize a fiber-optic coupling technology offered from the telecommunications and data communications applications areas. When successful, this technique allows tapping into fiber-optic trunk lines, a direct parallel with coaxial or twin axial systems. Light entering into the tap or coupler is split into a given number of output channels. The amount of light in any output channel is determined by the total amount of light input, less system losses, divided by the number of output channels. Additional losses are incurred at the junction of the main data-carrying fibers with the fiber leads from the tap or star. As such, passive couplers are limited to systems with few drops and moderate distances. Also, it is important to minimize termination losses at the coupler caused by the already diminished light output from the coupler. A partial solution is an active in-line repeater, but a superior solution, the active star *coupler*, is described next.

3.12.6.4 The Active Star Coupler

The basic principle of the active star coupler is that any light signal received as an input is converted to an electrical signal, amplified, and reconverted to optical signals on all other output channels. Figure 3.41 illustrates an eight-port active star coupler, containing eight sets of fiber-optic input/output (I/O) ports. A signal received on the channel 1 input will be transmitted on the channel 2 to 8 output ports. One may visualize the use of the active star coupler as aggregating a number of taps into one box. Should the number of

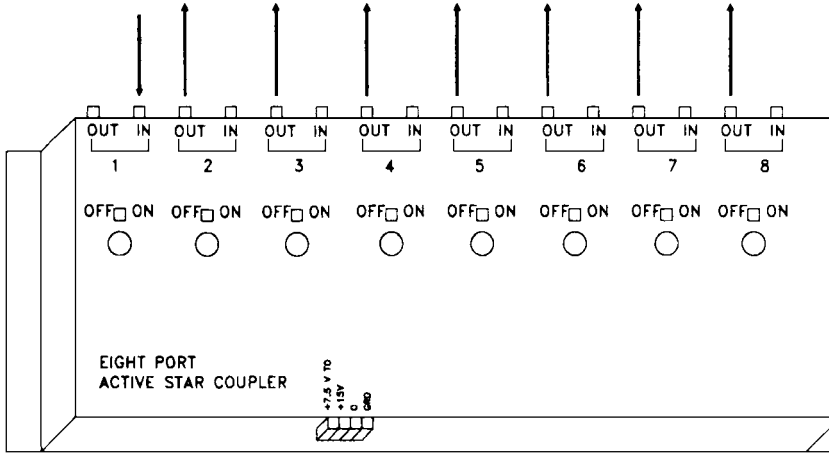


FIGURE 3.41 Eight-port active star coupler.

required taps exceed the number of available I/O ports, or should it be desirable to place these tap boxes at several locations in the system, the active star couplers may be jumpered together optically by tying a pair of I/O ports on one coupler to that on another in a hub-and-spoke system.

With the active star coupler serving as the hub of the data bus network, any message broadcast by a unit on the network is retransmitted to all other units on the network. A response of these other units is broadcast back to the rest of the network through the star, as in an electrical wired data bus network.

3.13 Configurations of Fiber Optics for Sensors

Fiber-optic sensors for general industrial use have largely been restricted to applications in which their small size has made them convenient replacements for conventional photoelectric sensors. Until recently, fiber-optic sensors have almost exclusively employed standard bundle technology, whereby thin glass fibers are bundled together to form flexible conduits for light.

Recently, however, the advances in fiber optics for data communications have introduced an entirely new dimension into optical sensing technology. Combined with novel but effective transducing technology, they set the stage for a powerful class of fiber-optic sensors.

3.13.1 Fiber-Optic Bundles

A typical fiber-optic sensor probe, often referred to as a *bundle* (Fig. 3.42), is 1.25 to 3.15 mm in diameter and made of individual fiber elements approximately 0.05 mm in diameter. An average

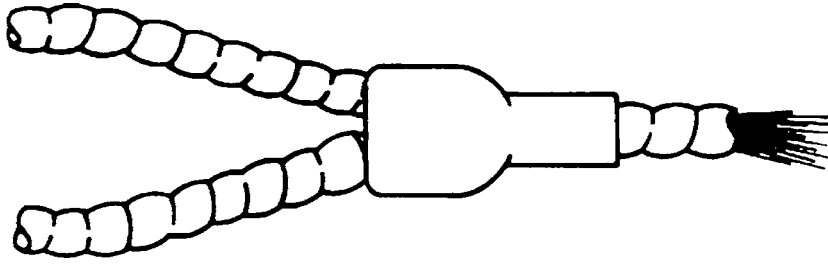


Figure 3.42 Fiber-optic sensor probe.

bundle will contain up to several thousand fiber elements, each working on the conventional fiber-optic principle of total internal reflection.

Composite bundles of fibers have an acceptance cone of the light based on the numerical aperture of the individual fiber elements.

$$\begin{aligned} \text{NA} &= \sin\Theta \\ &= \sqrt{n_1^2 - n_2^2} \end{aligned}$$

where $n_1 > n_2$ and Θ = half the cone angle.

Bundles normally have NA values in excess of 0.5 (an acceptance cone full angle greater than 60°), contrasted with individual fibers for long-distance, high-data-rate applications, which have NA values approaching 0.2 (an acceptance cone full angle of approximately 20°).

The ability of fiber-optic bundles to readily accept light, as well as their large total cross-sectional surface area, have made them an acceptable choice for guiding light to a remote target and from the target area back to a detector element. This has been successfully accomplished by using the pipe as an appendage to conventional photoelectric sensors, proven devices conveniently prepackaged with adequate light source and detector elements.

Bundles are most often used in either opposed beam or reflective mode. In the opposed beam mode, one fiber bundle pipes light from the light source and illuminates a second bundle—placed on the same axis at some distance away—which carries light back to the detector. An object passing between the bundles prevents light from reaching the detector.

In the reflective mode, all fibers are usually contained in one probe but divided into two legs at some junction point in an arrangement known as *bifurcate*. One bifurcate leg is then tied to the light source and the other to the detector (Fig. 3.43). Reflection from a target provides a return path to the detector for the light. The target may be fixed so it breaks the beam, or it may be moving so that, when present in the probe's field of view, it reflects the beam.

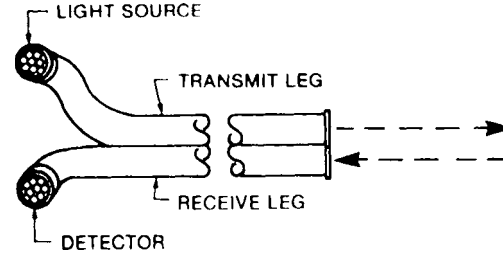


Figure. 3.43 Reflective mode bifurcate fiber optics.

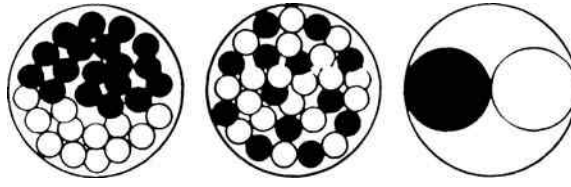


Figure 3.44 Bundle construction.

Typical bundle construction in a bifurcate employs one of two arrangements of individual fibers. The sending and receiving fibers are arranged either randomly or hemispherically (Fig. 3.44). As a practical matter, there is little, if any, noticeable impact on the performance of a photoelectric system in any of the key parameters such as sensitivity and scan range.

Application areas for bundle probes include counting, break detection, shaft rotation, and displacement/proximity sensing.

3.13.2 Bundle Design Considerations

Microscopic fiber flaws (Fig. 3.45) such as impurities, bubbles, voids, material absorption centers, and material density variations all diminish the ability of rays of light to propagate down the fiber, causing a

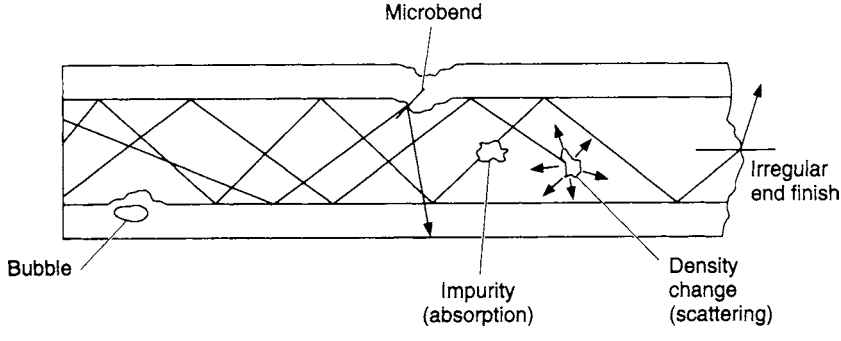


Figure 3.45 Microscopic fiber flaws.

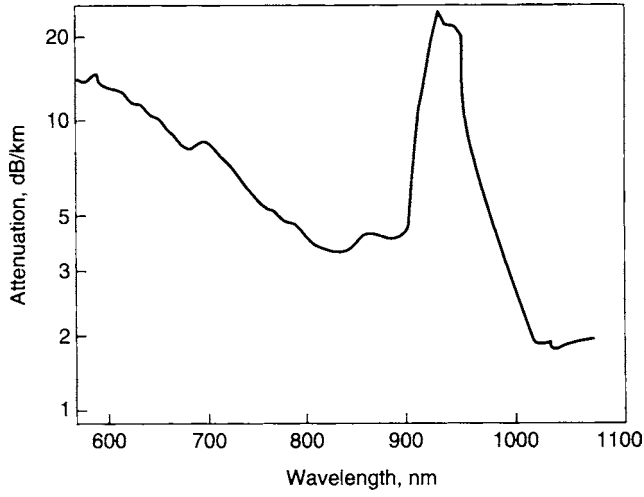


Figure 3.46 Characteristic absorption curve.

net loss of light from one end of the fiber to the other. All these effects combine to produce a characteristic absorption curve, which graphically expresses a wavelength-dependent loss relationship for a given fiber (Fig. 3.46). The fiber loss parameter is expressed as attenuation in dB/km as follows:

$$\text{Loss} = -10 \log (p_2/p_1)$$

where p_2 = light power output and p_1 = light power input.

Therefore, a 10-dB/km fiber would produce only 10 percent of the input light at a distance of 1 km.

Because of their inexpensive lead silicate glass composition and relatively simple processing techniques, bundles exhibit losses in the 500-dB/km range. This is several orders of magnitude greater than a communications-grade fiber, which has a loss of 10 dB/km. The maximum practical length for a bundle is thus only about 3 m. Further, the absence of coating on individual fibers and their small diameter make them susceptible to breakage, especially in vibratory environments.

Also, because of fiber microflaws, it is especially important to shield fibers from moisture and contaminants. A fiber exposed to water will gradually erode to the point of failure. This is true of any optical fiber, but is especially true of uncoated fibers in bundles.

3.13.3 Fiber Pairs for Remote Sensing

A viable solution to those design problems that exist with fiber bundles is to use large-core glass fibers that have losses on a par with

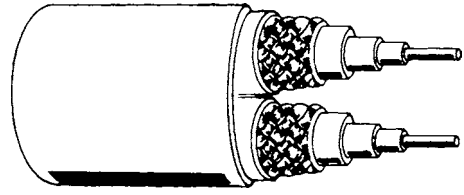


Figure 3.47
Pair of fibers.

telecommunication fibers. Although its ability to accept light is less than that of a bundle, a 200- or 400-μm core diameter plastic clad silica (PCS) fiber provides the ability to place sensing points hundreds of meters away from corresponding electronics. The fiber and the cable construction shown in Fig. 3.47 lend themselves particularly well to conduit pulling, vibratory environments, and general physical abuse. These fibers are typically proof-tested for tensile strength to levels in excess of 50,000 lb/in^2. A pair of fibers (Fig. 3.44) is used much like a bundle, where one fiber is used to send light to the sensing point and the other to return light to the detector. The performance limitation of a fiber pair compared to a bundle is reduced scan range; however, lenses may be used to extend the range. A fiber pair may be used in one of two configurations: (1) a single continuous probe (i.e., an unbroken length of cable from electronics to sensing point), or (2) a fiber-optic extension cord to which a standard probe in either a bundle or a fiber pair is coupled mechanically. This allows the economical replacement, if necessary, of the standard probe, leaving the extension cord intact.

The typical application for a fiber pair is object detection in explosive or highly corrosive environments (e.g., ammunition plants). In such cases, electronics must be remote by necessity. Fiber pairs also allow the construction of very small probes for use in such areas as robotics, small object detection, thread break detection, and small target rotation.

3.13.4 Fiber-Optic Liquid Level Sensing

Another technique for interfacing with fiber-optic probes involves the use of a prism tip for liquid sensing (Fig. 3.48). Light traveling down one leg of the probe is totally internally reflected at the prism-air interface. The index of refraction of air is 1. Air acts as a cladding material around the prism. When the prism contacts the surface of a liquid, light is stripped from the prism, resulting in a loss of energy at the detector. A properly configured system can discriminate between liquid types, such as gasoline and water, by the amount of light lost from the system, a function of the index of refraction of the liquid.

This type of sensor is ideal for set point use in explosive liquids, in areas where electronics must be remote from the liquid by tens or hundreds of meters, and where foam or liquid turbulence make other level-sensing techniques unusable.

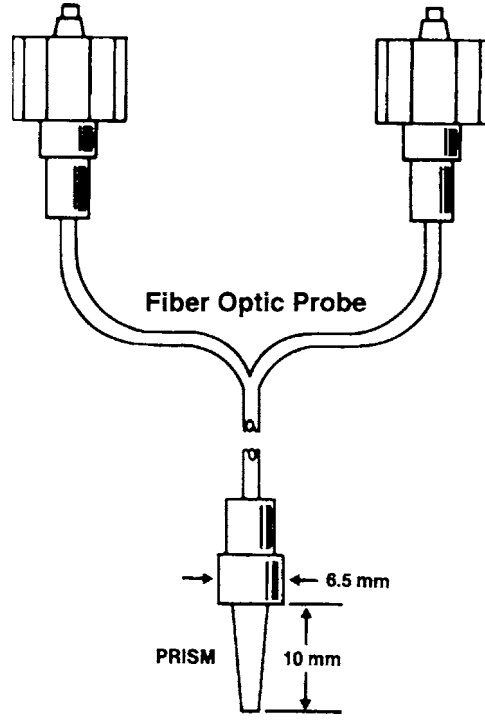


Figure 3.48
Prism tip for liquid sensing.

3.14 Flexibility of Fiber Optics

The power of fiber optics is further shown in the flexibility of its system configurations. A master industrial terminal (Fig. 3.49) can access any of a number of remote processors. The flexibility of switching, distance capability, and noise immunity of such a system are its primary advantages.

Figure 3.50 illustrates a passive optical coupler with a two-way fiber-optic link communicating over a single fiber through an on-axis rotary joint. Such a system allows a simple uninterrupted communication link through rotary tables or other rotating machinery. This is a true challenge for high-data-rate communication in a wire system.

3.14.1 Fiber-Optic Terminations

Optical fibers are becoming increasingly easier to terminate as rapid advances in termination technology continue to be made. Several manufacturers have connector systems that require no polishing of the fiber end, long a major objection in fiber optics. Products that eliminate epoxy adhesives are also being developed. Field installation times now typically average less than ten minutes for large-core fibers (100 and 200 μm) with losses in the 1- to 3-dB range. Further,

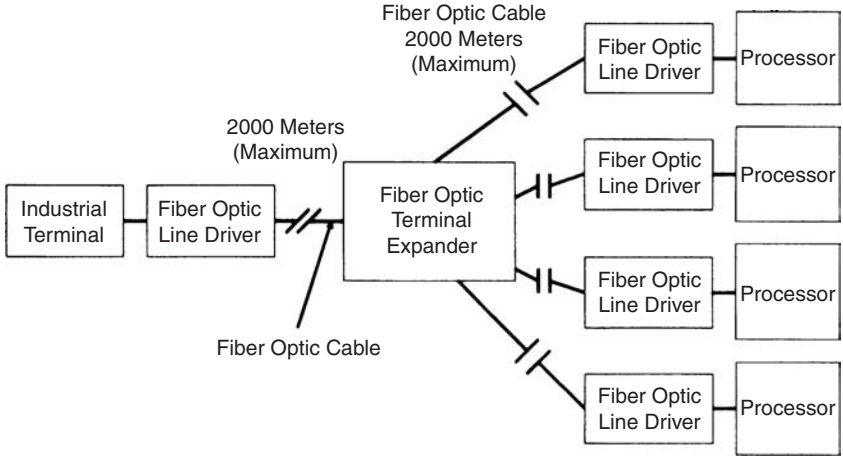


Figure 3.49 Master industrial terminal and remote processors.

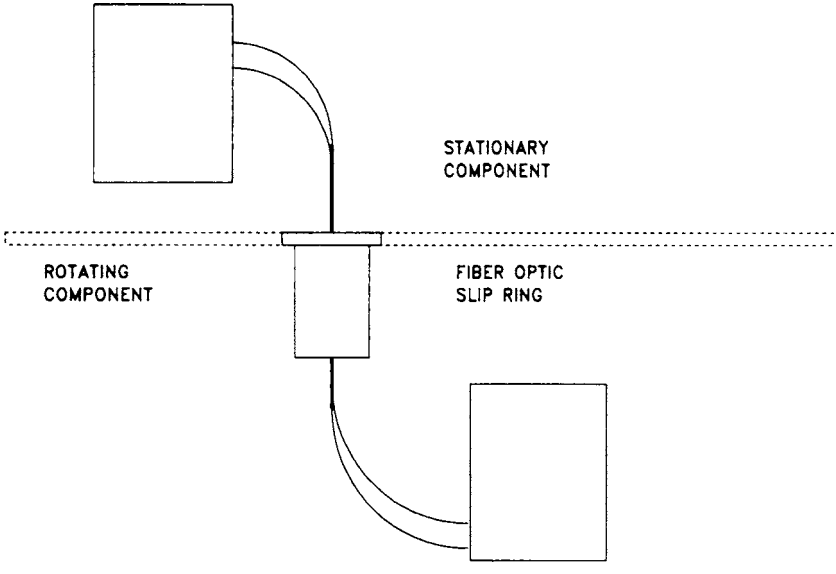


Figure 3.50 Passive optical coupler.

power budgets for well-designed industrial links normally provide a much greater latitude in making a connection. A 5- to 6-dB-loss connection, while potentially catastrophic in other types of systems, may be quite acceptable in short-haul systems with ample power budgets.

The most popular connector style for industrial communications is the SMA style connector, distinguished by its nose dimensions and

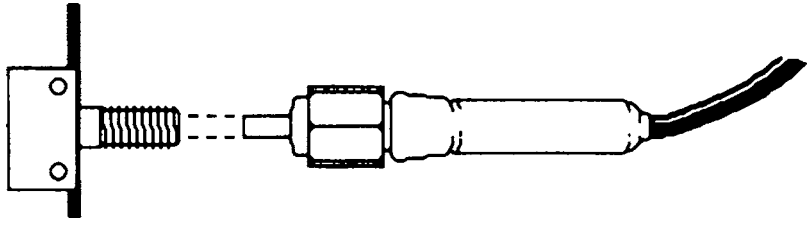


Figure 3.51 SMA connection to an information link.

configuration, as well as the thread size on the coupling nut. The coupling nut is employed to mechanically join the connector to a mating device on the data link or to a thread splice bushing. Figure 3.51 illustrates an SMA connection to an information link.

3.15 The Testing of Fiber Optics

Optical measurements, perhaps among the most difficult of all physical measurements, are fundamental to the progress and development of fiber-optic technology. Recently, various manufacturers have offered lines of fiber-optic test equipment for use in field and laboratory. Typical field measurement equipment determines the average optical power emitted from the system source, the component and overall system loss, the bit error rate, and the location of breaks in the fiber. Laboratory equipment measures loss through connectors and splicing, characterizes transmitters and receivers, and establishes bit error rate.

The testing of fiber-optic cables or systems is normally done with a calibrated light source and companion power meter. The light source is adjusted to provide a 0-dB reading on the power meter with a short length of jumper cable. The cable assembly being tested is then coupled between the jumper and the power meter to provide a reading on the meter, in decibels, that corresponds to the actual loss in the cable assembly.

Alternatively, the power through the cable from the system's transmitter can be read directly and compared with the system's receiver sensitivity specification. In the event of a cable break in a long span, a more sophisticated piece of test equipment, an optical time-domain reflectometer (OTDR), can be employed to determine the exact position of the break.

3.16 Testing Light Sources

The Photodyne 9XT optical source driver (Fig. 3.52) is a handheld unit for driving LED and laser fiber-optic light sources. The test equipment is designed to take the shock and hard wear of the typical

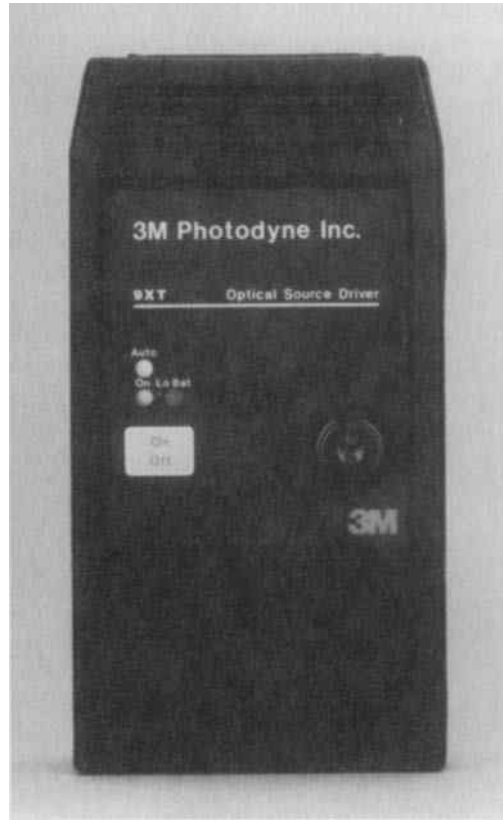


Figure 3.52
Photodyne 9XT optical source driver. (*Courtesy 3M Corporation*)

work-crew environment. The unit is powered from two rechargeable nicad batteries or from line voltage. The LED series is suited for measurement applications where moderate dynamic range is required or where coherent light should be avoided. The laser modules are used for attenuation measurements, requiring extreme dynamic range or where narrow spectral width and coherent light are required. The laser modules are the most powerful source modules available.

3.16.1 Power Meters

Optical power meters are commonly used to measure absolute light power in dBm, Figure 3.53. For dBm measurement of light transmission power, traceable calibration of fiber-optic power meters is essential. A fiber-optic power meter is also used with an optical light source for measuring relative power levels in dB. Alternatively, some users may prefer an integrated two-way Test Loss Set (TLS), or a simple Loss Test Set (LTS).

The Photodyne 2285XQ fiber-optic power meter (Fig. 3.54) is a low-cost optical power meter for general-purpose average-power



Figure 3.53 Hand-held optical power meter.



Figure 3.54 Photodyne 2285XQ fiber-optic power meter. (*Courtesy 3M Corporation*)

measurement, but particularly for fiber-optic applications, in both manual and computer-controlled test setups. A unique and powerful feature of the 2285XSQ is its ratio function. This allows the user to make ratio (A/B) measurements by stacking several instruments together via the interface without the need for a controller.

Another very powerful feature is the built-in data logger. With this function, data may be taken at intervals from 1 to 9999 s. This feature is useful in testing optical devices for short- and long-term operation.

At the heart of the instrument is a built-in large-area high-sensitivity indium gallium arsenide (InGaAs) sensor. All industry connectors may be interfaced to the sensor using any of the Photodyne series 2000 connector adapters. The sensor is calibrated for the all the fiber-optic windows: 820, 850, 1300, and 1550 nm.

3.17 Dual Laser Test Sets

The Photodyne 2260XF and 2260XFA are switchable dual laser test sets, with a transmit and receive section in one unit. They measure and display loss in fiber links at both 1300 and 1550 nm simultaneously in one pass. They are designed for use in installing, maintaining, and troubleshooting single-mode wavelength-division multiplexed (WDM) fiber-optic links operating at 1300- and 1550-nm wavelengths. They may also be used for conventional links operating at either 1300 or 1550 nm.

The essential differences between the two models are that the XF version (Fig. 3.55) has a full complement of measurement units, log and linear (dBm, dB, *m*W, μW, *n*W), whereas the XFA version has log units only (dBm, dB). In the XF version, laser power output is

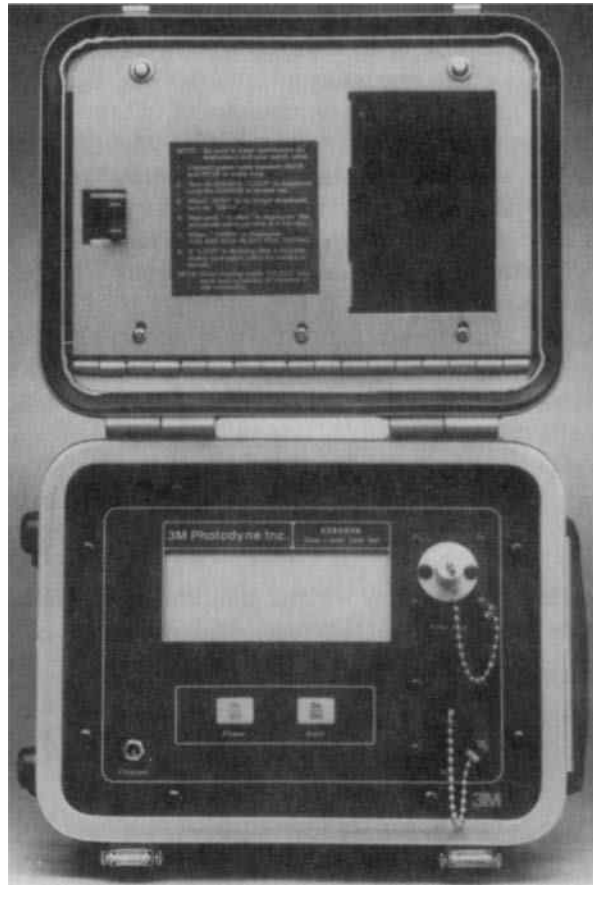


Figure 3.55 Photodyne 2260XF switchable dual laser test set. (*Courtesy 3M Corporation*)

adjustable over the range 10 to 20 dBm (100 μW to 21 mW). In the XFA version, it is automatically set to a fixed value of 10 dBm. The XF version allows the user to access wavelengths over the ranges 1250 to 1350 nm and 1500 to 1600 nm in 10-nm steps. The XFA version is fixed at 1300 nm to 1550 nm. To control all functions, the XF version has six keys; the XFA has only two. Although both instruments perform identical tasks equally well, the XF may be seen as the more flexible version, and the XFA as the simpler-to-use version.

3.17.1 Test Sets/Talk Sets

The handheld Photodyne 21XTL fiber-optic test set/talk set (Fig. 3.56) is for use in installation and maintenance of fiber cables. This instrument functions as a power meter and test set, as well as a talk set. For maintenance purposes, the user may establish voice communication over the same fiber pair being measured.

The 21XTL, as a power meter, covers an exceptionally wide dynamic range (–80 to +3 dBm). As an option, the receiver may include a silicon or an enhanced InGaAs photodiode. With the InGaAs version, the user can perform measurements and voice communication at short and long wavelengths. The silicon version achieves a superior dynamic range at short wavelengths only.

The highly stabilized LED ensures repeatable and accurate measurements. Precision optics couples the surface-emitter LED to the fiber core. With this technique, the fiber end will not wear out or scratch. The transmitter is interchangeable, providing complete flexibility of wavelengths and connector types.



Figure 3.56 Photodyne 21XTL fiber-optic test set/talk set. (*Courtesy 3M Corporation*)

3.17.2 Attenuators

The 19XT optical attenuator (Fig. 3.57) is a handheld automatic optical attenuator. It provides continuous attenuation of both short- and long-wave optical signals in single-mode and multimode applications. The calibrated wavelengths are 850 nm/1300 nm multimode and/or 1300 nm/1550 nm single mode. An attenuation range of 70 dB is offered with 0.1-dB resolution and an accuracy of ±0.2 dB typical (±0.5 dB maximum). Unique features allow scanning between two preset attenuation values and includes the insertion loss in the reading.

The 19XT allows simple front panel operation or external control of attenuation through analog input and output connections.

3.17.3 Fault Finders

The Photodyne 5200 series optical fault finders (Fig. 3.58) offer flexible alternatives for localizing faults or trouble areas on any fiber-optic network. The 5200 series fault finders can easily be integrated into the troubleshooting routines of fiber-optic crews. They offer a variety of features, such as:

- Fast accurate analysis of faults
- Autoranging for greatest accuracy

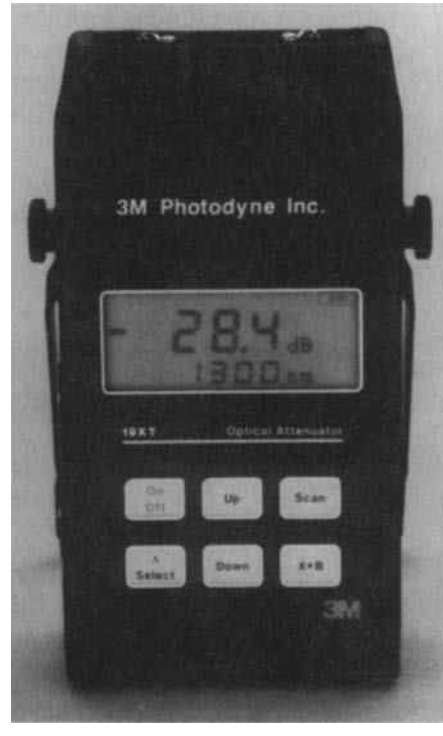


Figure 3.57 Photodyne 19XT optical attenuator. (*Courtesy 3M Corporation*)

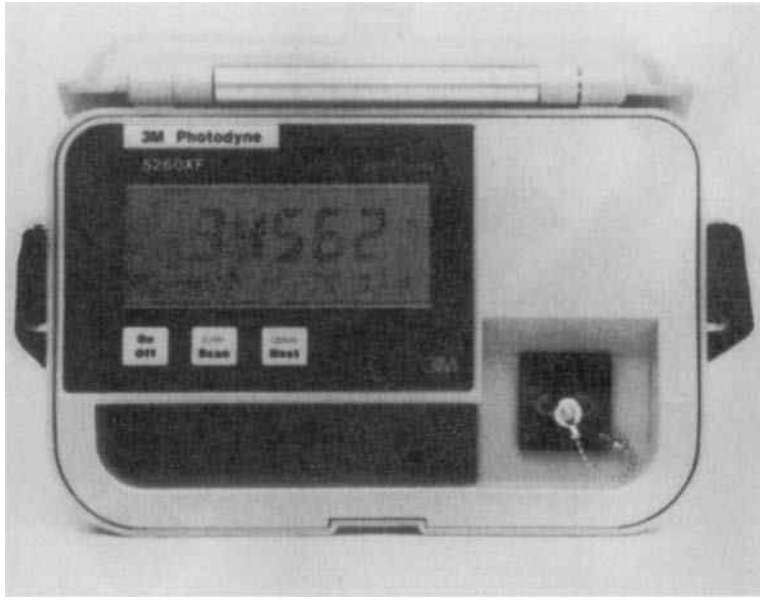


Figure 3.58 Photodyne 5200 series optical fault finder. (*Courtesy 3M Corporation*)

- Reflective and nonreflective fault detection
- Multiple-fault detection capabilities
- Go/no-go splice qualification
- Variable fault threshold settings (0.5 to 6.0 dB)
- Fault locations up to 82 km
- Automatic self-test and performance checks
- AC or rechargeable battery operation
- A large easy-to-read liquid-crystal display (LCD)

3.17.4 Fiber Identifiers

The Photodyne 8000XG fiber identifier (Fig. 3.59) is designed for fast accurate identification and traffic testing of fiber-optic lines without cutting the fiber line or interrupting normal service. Ideal for use during routine maintenance and line modification, this small handheld unit can be used to locate any particular fiber line, identify live fibers, and determine whether or not traffic is present. Features of the 8000XG include:

- Light weight, portability, and battery operation
- Automatic self-test after each fiber insertion

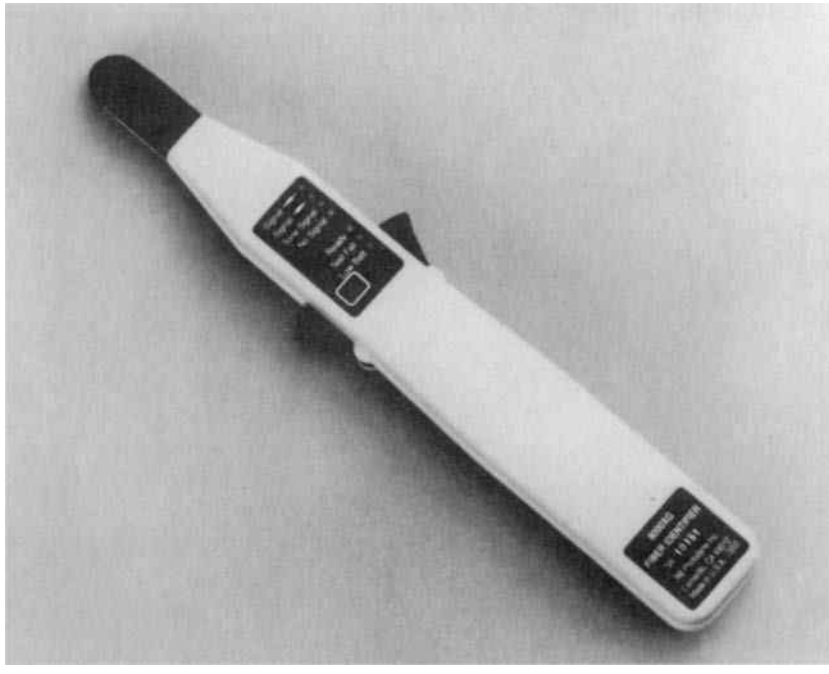


FIGURE 3.59 Photodyne 8000XG fiber identifier. (*Courtesy 3M Corporation*)

- Mechanically damped fiber action
- Operation over 850- to 1550-nm range
- Transmission direction indicators
- 1- and 2-kHz tone detection
- Low insertion loss at 1300 and 1550 nm
- Completely self-contained operation

3.18 Networking with Electrooptic Links

The following examples describe a number of products utilizing communication through fiber electrooptic modules. The function of the fiber is to replace wire. This can be achieved by interconnecting the electrooptic modules in a variety of ways. Figure 3.60 shows a programmable controller communication network through a coaxial cable bus branched to four remote input/output stations. The programmable controller polls each of the input/output stations in sequence. All input/output stations hear the programmable controller communication, but only the one currently being addressed responds.

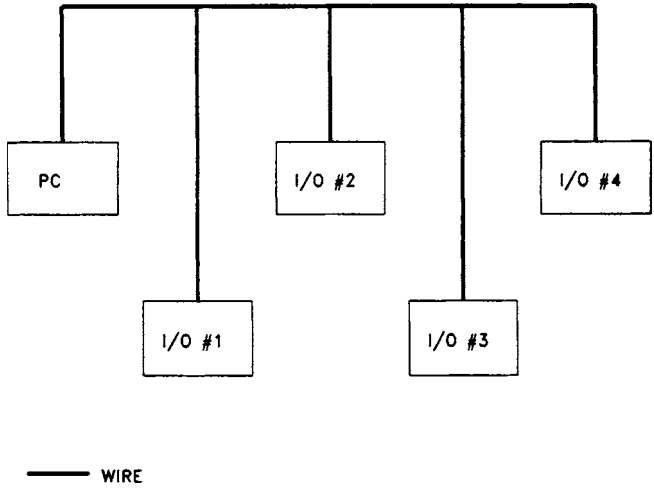


Figure 3.60 Programmable controller communication network.

3.18.1 Hybrid Wire/Fiber Networks

Figure 3.61 shows an electrooptic module used to replace a troublesome section of coaxial cable subject to noise, grounding problems, lightning, or hazardous environments. When fiber is used in this mode, the electrooptic module should be placed as close to the input/output drop as possible in order to minimize the effect of potential electrical noise problems over the section of coax cable connecting them.

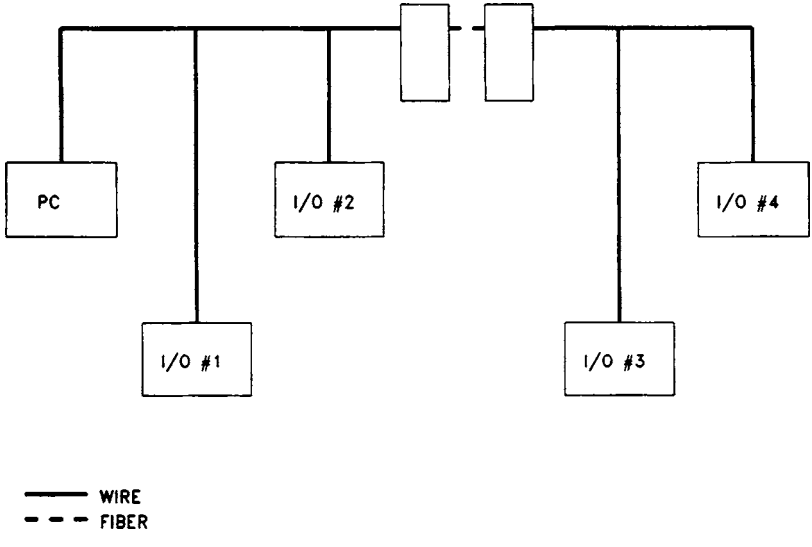


Figure 3.61 Hybrid wire/fiber network.

3.18.2 Daisy Chain Networks

The daisy chain configuration (Fig. 3.62) is an economical choice for long straight-line installations (e.g., conveyer or mine shaft applications). The signal generated at the programmable controller is converted to light and transmitted outward. At each transmitted section, the electrical signal is reconstructed and a light signal is regenerated and sent down the chain.

3.18.3 Active Star Networks

The electrooptical programmable controller links may be joined to an electrooptic module with four-port and eight-port active star couplers for a hub-and-spoke type system. Figure 3.63 shows two active star couplers joined via fiber, a configuration that might be appropriate where clusters of input/output racks reside in several locations

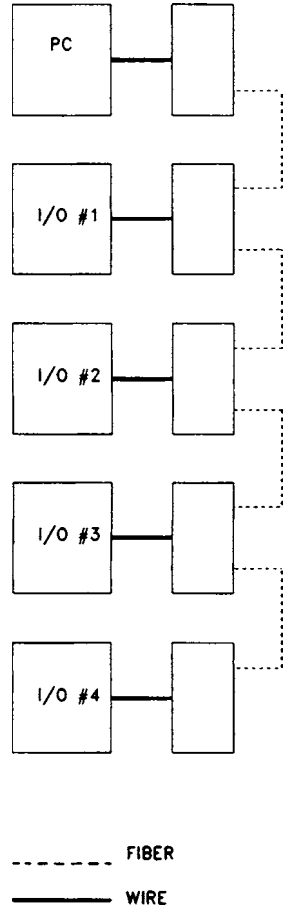


Figure 3.62 Daisy chain network.

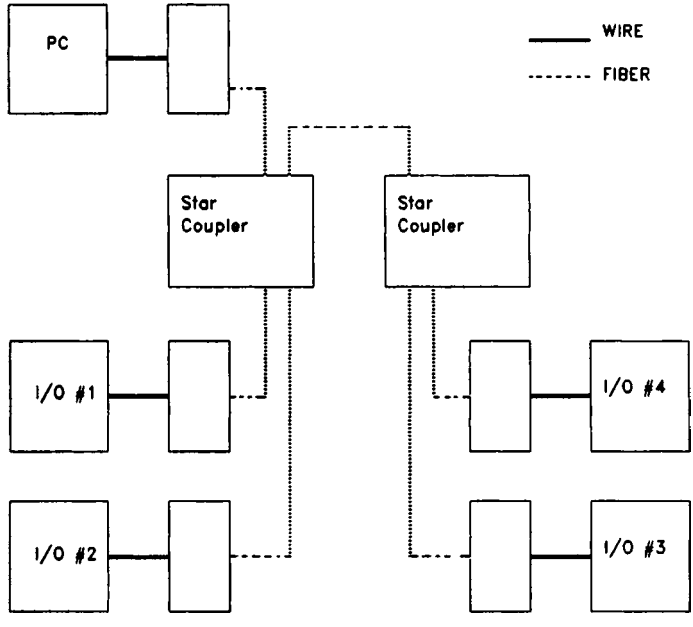


Figure 3.63 Two active star couplers.

separated by some distance. The star coupler then becomes the distributor of the light signals to the racks in each cluster, as well as a potential repeater to a star in another cluster.

3.18.4 Hybrid Fiber Networks

Star and daisy chain network structures can be combined to minimize overall cabling costs (Fig. 3.64). The fiber network configuration exactly duplicates the coax network. The final decision on exactly which configuration to choose depends on the following criteria:

- Economical cabling layout
- Length of cable runs versus cost of electronics
- Location of end devices and power availability
- Power-out considerations

Other considerations specific to the programmable controller being used can be summarized as follows:

- *Pulse-width.* Total acceptable pulse-width distortion, which may limit the number of allowable repeater sites, is

$$\text{Allowable distortion (ns)} = \text{allowable percent distortion} \times \text{period of signal (ns)}$$

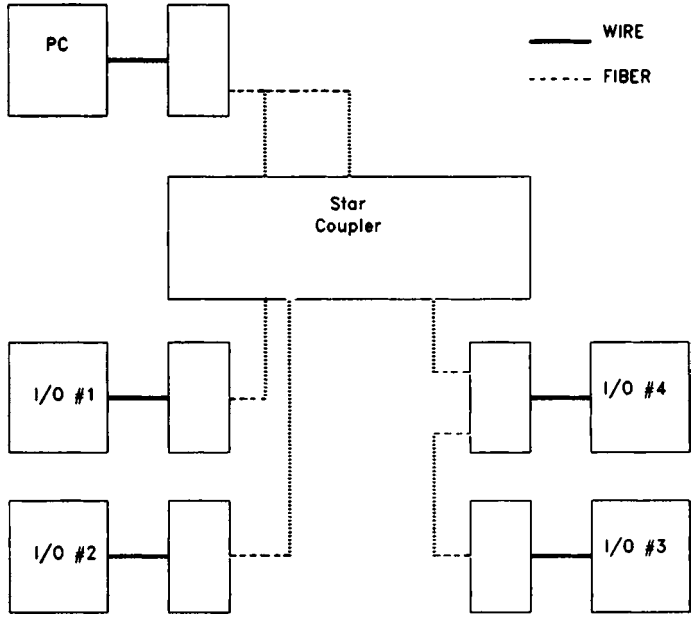


FIGURE 3.64 Hybrid fiber network.

- *Signal propagation delay.* Allowable system signal propagation delay may limit the overall distance of the fiber network.

Example.

$$\begin{aligned}\text{Electrooptic module distortion} &= 50\% \text{ allowable distortion} \\ &\quad \times 1 \text{ Mbaud transmission} \\ &= 50 \text{ ns}\end{aligned}$$

Calculate:

$$\text{Allowable distortion} = 50\% \text{ allowable distortion} \times 10^{-6} = 500 \text{ ns}$$

$$\text{Maximum number of repeater sites} = 500 \text{ ns}/50 \text{ ns} = 10$$

NOTE: *A star coupler counts as a repeater site.*
Distance Calculation (Signal Propagation Delay)
Delay of light in fiber-optic regeneration 1.5 ns/ft
Delay in module 50 ns

Manufacturers of programmable controllers will provide the value of the system delay. This value must be compared with the calculated allowable delay. If the overall fiber system is longer than the published maximum length of the wired system, the system must be reconfigured.

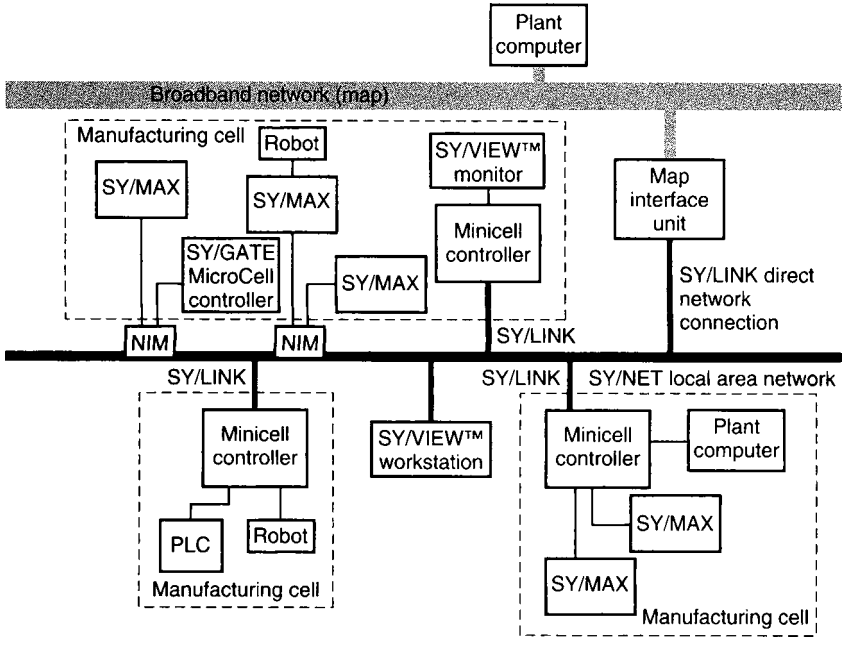


Figure 3.65 Minicell controller.

3.18.5 Fiber-Optic Sensory Links for Minicell Controllers

A minicell controller is typically used to coordinate and manage the operation of a manufacturing cell, consisting of a group of automated programmable machine controls (programmable controllers, robots, machine tools, etc.) designed to work together and perform a complete manufacturing or process-related task. A key benefit of a minicell controller is its ability to adjust for changing products and conditions. The minicell controller is instructed to change data or complete programs within the automation work area. A minicell controller is designed to perform a wide variety of functions such as executing programs and data, uploading/downloading from programmable controllers, monitoring, data analysis, tracking trends, generating color graphics, and communicating in the demanding plant floor environment. Successful minicell controllers use fiber-optic links that can interface with a variety of peripheral devices. A minicell controller can be used in a variety of configurations, depending on the optical-fiber lengths, to provide substantial system design and functional flexibility (Fig. 3.65).

3.19 Versatility of Fiber Optics in Industrial Applications

A constant concern in communication is the ever-increasing amount of information that must be sent with greater efficiency over a medium requiring less space and less susceptibility to outside interferences.

As speed and transmission distance increase, the problems caused by electromagnetic interference, radio frequency interference, cross talk, and signal distortion become more troublesome. In terms of signal integrity, just as in computer-integrated manufacturing data acquisition and information-carrying capacity, fiber optics offers many advantages over copper cables. Furthermore, optical fibers emit no radiation and are safe from sparking and shock. These features make fiber optics the ideal choice for many processing applications where safe operation in hazardous or flammable environments is a requirement.

Accordingly, fiber-optic cables offer the following advantages in industrial applications:

- Wide bandwidth
- Low attenuation
- Electromagnetic immunity
- No radio frequency interference
- Small size
- Light weight
- Security
- Safety in hazardous environment

3.19.1 High-Clad Fiber-Optic Cables

Large-core, multimode, step-index high-clad silica fiber-optic cables make fiber-optic technology user-friendly and help designers of sensors, controls, and communications realize substantial cost savings. The coupling efficiency of high-clad silica fibers allows the use of less expensive transmitters and receivers. High-clad silica polymer technology permits direct crimping onto the fiber cladding. Field terminations can be performed in a few minutes or less, with minimal training. The following lists describe the structure and characteristics of several fiber-optic cables used in industry.

Simplex fiber-optic cables (Fig. 3.66) are used in:

- Light-duty indoor applications
- Cable trays
- Short conduits

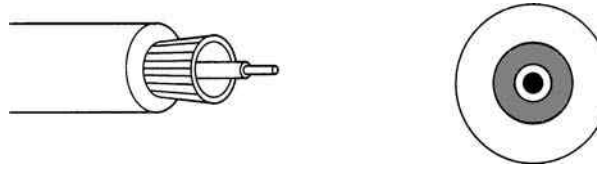


Figure 3.66 Simplex fiber-optic cable.

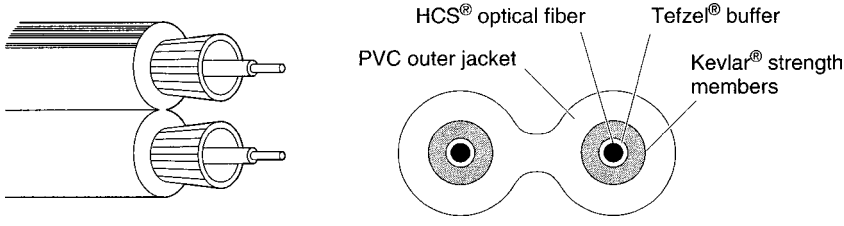


Figure 3.67 Zipcord fiber-optic cable.

- Loose tie wrapping
- Subchannels for breakout cables

Zipcord fiber-optic cables (Fig. 3.67) are used in:

- Light-duty (two-way) transmission
- Indoor runs in cable trays
- Short conduits
- Tie wrapping

Multichannel fiber-optic cables (Fig. 3.68) are used in:

- Outdoor environments
- Multifiber runs where each channel is connectorized and routed separately
- Aerial runs
- Long conduit pulls
- Two, four, and six channels (standard)
- Eight to eighteen channels

Heavy-duty duplex fiber-optic cables (Fig. 3.69) are used in:

- Rugged applications
- Wide-temperature-range environments

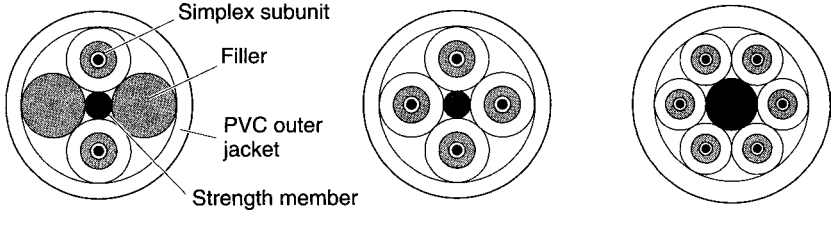


Figure 3.68 Multichannel fiber-optic cable.

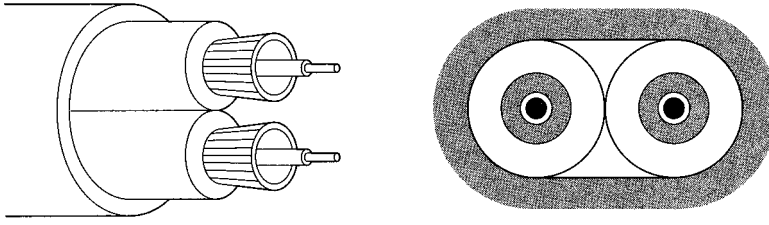


Figure 3.69 Heavy-duty fiber-optic cable.

- Direct burial
- Loose tube subchannel design
- High-tensile-stress applications

Table 3.1 summarizes fiber-optic cable characteristics.

				Channels		
Characteristics	**Simplex**	**Zipcord**	**Duplex**	**2**	**4**	**6**
Cable diameter, mm	2.5	2.5 × 5.4	3.5 × 6	8	8	10
Weight, kg/km	6.5	11	20	41	41	61
Jacket type	PVC	PVC	PVC	PE	PE	PE
Jacket color	Orange	Orange	Orange	Black	Black	Black
Pull tension	330 N	490 N	670 N	890 N	1150 N	2490 N
Max. long-term tension	200 N	310 N	400 N	525 N	870 N	1425 N
Max. break strength	890 N	1340 N	1780 N	2370 N	4000 N	11,000 N
Impact strength at 1.6 N · m	100	100	150	200	200	200
Cyclic flexing, cycles	>5000	>5000	>5000	>2000	>2000	>2000
Minimum bend radius, mm	25	25	25	50	50	75
Cable attenuation	0.6 dB/km at 820 nm					
Operating temperature	–40 to + 85°C					
Storage temperature	–40 to + 85°C					

Table 3.1 Fiber-Optic Cable Characteristics

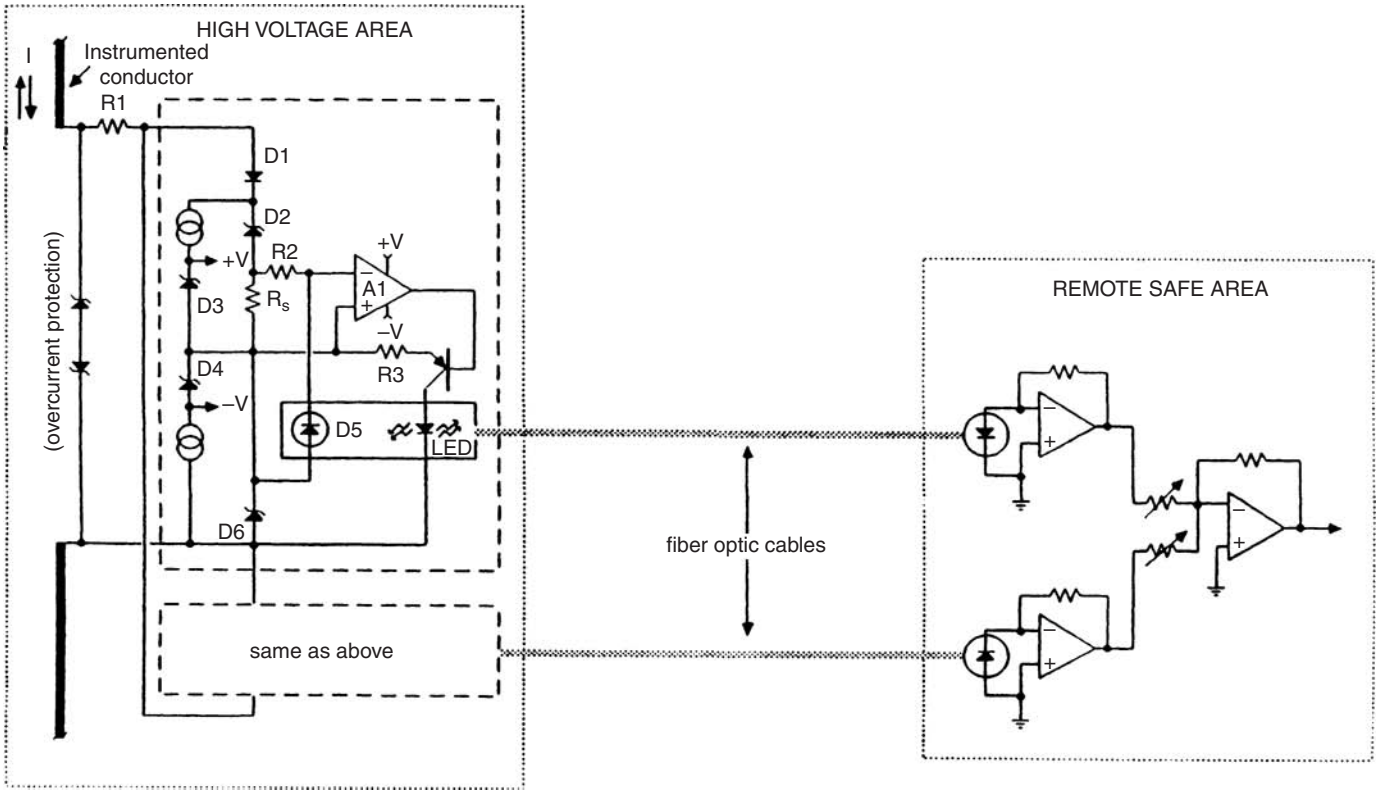


Figure 3.70 Fiber-optic current monitor.

3.19.2 Fiber-Optic Ammeter

In many applications, including the fusion reactors, radio frequency systems, and telemetry systems, it is often necessary to measure the magnitude and frequency of current flowing through a circuit in which high DC voltages are present. A fiber-optic current monitor (Fig. 3.70) has been developed at the Princeton Plasma Physics Laboratory (PPPL) in response to a transient voltage breakdown problem that caused failures of Hall-effect devices used in the Tokamak fusion test reactor's natural-beam heating systems.

The fiber-optic current monitor measures low current in a conductor at a very high voltage. Typical voltages range between tens of kilovolts and several hundred kilovolts. With a dead band of approximately 3 mA, the circuit derives its power from the conductor being measured and couples information to a (safe) area by means of fiber optics. The frequency response is normally from direct current to 100 kHz, and a typical magnitude range is between 5 and 600 mA.

The system is composed of an inverting amplifier, a current regulator, transorbs, diodes, resistors, and a fiber-optic cable. Around an inverting amplifier, a light-emitting diode and a photodiode form an optical closed feedback loop. A fraction of the light emitted by the LED is coupled to the fiber-optic cable.

As the current flows through the first diode, it splits between the 1.5-mA current regulator and the sampling resistor. The voltage across the sampling resistor causes a small current to flow into the inverting amplifier summing junction and is proportional to the current in the sampling resistor. Since photodiodes are quite linear, the light power from the LED is proportional to the current through the sampling resistor. The light is split between the local photodiode and the fiber cable. A photodiode, located in a remote safe area, receives light that is linearly proportional to the conductor current (for current greater than 5 mA and less than 600 mA).

To protect against fault conditions, the design utilizes two back-to-back transorbs in parallel with the monitor circuit. The transorbs are rated for 400 A for 1 ms. The fiber-optic ammeter is an effective tool for fusion research and other applications where high voltage is present.

Further Reading

Berwick, M., J.D.C. Jones, and D. A. Jackson, "Alternating Current Measurement and Non-Invasive Data Ring Utilizing the Faraday Effect in a Closed Loop Fiber Magnetometer," *Optics Lett.*, 12(294) (1987).

Cole, J. H., B. A. Danver, and J. A. Bucaro, "Synthetic Heterodyne Interferometric Demodulation," *IEEE J. Quant. Electron.*, QE-18(684) (1982).

Dandridge, A., and A. B. Tveten, "Phase Compensation in Interferometric Fiber Optic Sensors," *Optics Lett.*, 7(279) (1982).

Desforges, F. X., L. B. Jeunhomme, Ph. Graindorge, and G. L. Baudec, "Fiber Optic Microswitch for Industrial Use," presented at SPIE O-E Fiber Conf., San Diego, no. 838–41 (1987).

Favre, F., and D. LeGuen, "High Frequency Stability in Laser Diode for Heterodyne Communication Systems," *Electron. Lett.*, 16(179) (1980).
Giallorenzi, T. G., "Optical Fiber Interferometer Technology and Hydrophones in Optical Fiber Sensors," NATO ASI Series E, No. 132, Martinus Nijhoff Dordrecht, 35–50 (1987).
Hocker, G. B., "Fiber Optic Sensing for Pressure and Temperature," *Appl. Optics*, 18(1445) (1979).
Jackson, D. A., A. D. Kersey, and A. C. Lewin, "Fiber Gyroscope with Passive Quadrature Demodulation," *Electron. Lett.*, 20(399) (1984).
Optical Society of America, *Optical Fiber Optics.* Summaries of papers presented at the Optical Fiber Sensors Topical Meeting, January 27–29, 1988, New Orleans, La. (IEEE, Catalog No. 88CH2524–7).
Popovic, R. S., "Hall Effect Devices," *Sensors and Actuators*, 17, 39–53 (1989).
Rock, I., *Lightness Constancy, Perception*, W. H. Freeman, New York, 1984.
Seippel, R. G., *Optoelectronics*, Reston Publishing Co., Reston, Va., 1981.
Saxena, S. C., and S. B. Lal Seksena, "A Self-Compensated Smart LVDT Transducer," *IEEE Trans. Instrum. Meas.*, 38, 748–783 (1989).
Wong, Y. L., and W. E. Ott, *Function Circuits and Applications*, McGraw-Hill, New York, 1976.

CHAPTER 4

Networking of Sensors and Control Systems in Manufacturing

4.1 Introduction

Central to the development of any computer-integrated manufacturing facility is the selection of the appropriate automated manufacturing system and the sensors and control systems to implement it. The degree to which a CIM configuration can be realized depends on the capabilities and cost of available equipment and the simplicity of information flow.

When designing an error-free manufacturing system, the manufacturing design group must have an appreciation for the functional limits of the automated manufacturing equipment of interest and the ability of the sensors to provide effective information flow since these parameters will constrain the range of possible design configurations. Obviously, it is not useful to design a manufacturing facility that cannot be implemented because it exceeds the equipment's capabilities. It is desirable to match automated manufacturing equipment to the application. Although sensors and control systems are—by far—less costly than the automated manufacturing equipment, it is neither useful nor cost-effective to apply the most sophisticated sensors and controls, with the highest performance, to every possible application. Rather, it is important that the design process determines the preferred parameter values.

The preferred values must be compatible with available equipment and sensors and control systems, and should be those appropriate for

the particular factory. The parameters associated with the available equipment and sensors and control systems drive a functional process of modeling the manufacturing operation and facility. The parameters determine how the real-world equipment constraints will be incorporated into the functional design process. In turn, as many different functional configurations are considered, the cost-benefit relations of these alternatives can be evaluated and preferred parameter values determined. So long as these preferred values are within the limits of available automated manufacturing equipment and sensory and control systems, the design group is assured that the automated manufacturing equipment can meet its requirements. To the degree that optimum design configurations exceed present equipment capabilities, original equipment manufacturers (OEMs) are motivated to develop new equipment designs and advanced sensors and control systems.

Sensors and control systems, actuators/effectors, controllers, and control loops must be considered in order to appreciate the fundamental limitations associated with manufacturing equipment for error-free manufacturing. Many levels of factory automation are associated with manufacturing equipment; the objective at all times should be to choose the levels of automation and information flow that are appropriate for the facility being designed, as revealed through cost-benefit studies. Manufacturing facilities can be designed by describing each manufacturing system—and the sensors and controls to be used in it—by a set of functional parameters. These parameters are:

- The number of product categories for which the automated manufacturing equipment, sensors, and control systems can be used (with software downloaded for each product type)
- The mean time between operator interventions (MTOI)
- The mean time of intervention (MTI)
- The percentage yield of product of acceptable quality
- The mean processing time per product

An *ideal equipment unit* would be infinitely flexible so it could handle any number of categories desired, would require no operator intervention between setup times, would produce only product of acceptable quality, and would have unbounded production capabilities.

The degree to which real equipment containing sensors and control systems can approach this ideal depends on the physical constraints associated with the design and operation of the equipment and the ability to obtain instantaneous information about equipment performance through sensors and control systems. The performance of the equipment in each of the preceding five parameters is related

to details of the equipment's operation in an error-free environment. Relationships must be developed between the physical description of the equipment's operation and the functional parameters that will be associated with this operation. The objective is to link together the physical design of the equipment and its functional performance through sensory and control systems in the factory setting.

This concept provides insight into an area in which future manufacturing system improvements would be advantageous, and also suggests the magnitude of the cost-benefit payoffs that might be associated with various equipment designs. It also reveals the operational efficiency of such systems.

An understanding of the relationships between the equipment characteristics and the performance parameters based on sensors and control systems can be used to select the best equipment for the parameter requirements associated with a given factory configuration. In this way, the manufacturing design team can survey alternative types of available equipment and select the units most appropriate for each potential configuration.

4.2 The Number of Products in a Flexible System

The first parameter listed earlier, the number of product categories for which the manufacturing system can be used, represents the key concern in flexible manufacturing. A unit of automated manufacturing equipment is described in terms of the number of product categories for which it can be used with only a software download to distinguish among product types. A completely fixed automated manufacturing system that cannot respond to computer control might be able to accommodate only one product category without a manual setup. On the other hand, a very flexible manufacturing system would be able to accommodate a wide range of product categories with the aid of effective sensors and control systems. This parameter will thus be defined by the breadth of the processes that can be performed by an automated manufacturing equipment unit and the ability of the unit to respond to external control data to shift among these operations.

The most effective solution will depend on the factory configuration that is of interest. Thus, OEMs are always concerned with anticipating future types of factories in order to ensure that their equipment will be an optimum match to the intended configuration. This also will ensure that the *concept of error-free manufacturing can be implemented with a high degree of spontaneity*. A continual tradeoff exists between flexibility and cost. In general, more flexible and "smarter" manufacturing equipment will cost more. Therefore, the objective in a particular setting will be to achieve just the required amount of flexibility, without any extra capability built into the equipment unit.

4.3 Sensors Tracking the Mean Time between Operator Interventions

The MTOI value should be matched to the factory configuration in use. In a highly manual operation, it may be acceptable to have an operator intervene frequently. On the other hand, if the objective is to achieve operator-independent manufacturing between manual setups, then the equipment must be designed so that the MTOI is longer than the planned duration between manual setups. The manufacturer of automated equipment with adequate sensors and control systems must try to assess the ways in which factories will be configured and produce equipment that can satisfy manufacturing needs without incurring any extra cost due to needed features.

4.4 Sensors Tracking the Mean Time of Intervention

Each time an intervention is required, it is desirable to compare the intervention interval with that for which the system was designed. If the intervention time becomes large with respect to the planned mean time between operator interventions, then the efficiency of the automated manufacturing equipment drops rapidly in terms of the fraction of time it is available to manufacture the desired product.

4.5 Sensors Tracking Yield

In a competitive environment, it is essential that all automated manufacturing equipment emphasize the production of quality product. If the automated manufacturing equipment produces a large quantity of product that must be either discarded or reworked, then the operation of the factory is strongly affected, and costs will increase rapidly. The objective, then, is to determine the product yields required for given configurations and to design automated manufacturing equipment containing sensors and control systems that can achieve these levels of yield. Achieving higher yield levels will, in general, require additional sensing and adaptability features for the equipment. These features will enable the equipment to adjust and monitor itself and, if it gets out of alignment, to discontinue operation.

4.6 Sensors Tracking the Mean Processing Time

If more product units can be completed in a given time, the cost of automated manufacturing equipment with sensors and control systems can be more widely amortized. As the mean processing time is reduced, the equipment can produce more product units in a given time, reducing the manufacturing cost per unit. Again, automated

manufacturing equipment containing sensory and control systems generally becomes more expensive as the processing time is reduced. Tradeoffs are generally necessary among improvements in the five parameters and the cost of equipment. If high-performance equipment is to be employed, the factory configuration must make effective use of the equipment's capabilities to justify its higher cost. On the other hand, if the factory configuration does not require the highest parameters, then it is far more cost-effective to choose equipment units that are less sophisticated but adequate for the purposes of the facility. This interplay between parameter values and equipment design and cost is an essential aspect of system design.

Table 4.1 illustrates the difference between available parameter values and optimum parameter values, where the subscripts for equipment E represent increasing levels of complexity. The table shows the type of data that can be collected to evaluate cost and benefits. These data have significant impact on system design and performance which, in turn, have a direct impact on product cost. Given the type of information in Table 4.1, system designers can evaluate the effects of utilizing various levels of sensors and control systems on new equipment and whether they improve performance enough to be worth the research and development and production investment.

Equipment	MTOI, min	MTI, min	Yield, %	Process, min	R&D Expense, Thousands of Dollars	Equipment Cost, Thousands of Dollars
Production Function A						
E_1	0.1	0.1	90	12		50
E_2	1.0	0.1	85	8		75
E_3	10	1.0	80	10		85
E_4	18	1.0	90	8	280	155
Production Function B						
E_1	1.0	0.1	95	10		150
E_2	10	0.5	90	2		300
Production Function C						
E_1	0.1	0.1	98	3		125
E_2	5.0	1.0	98	2		250
E_3	8.0	2.0	96	1		300
E_4	20	2.0	96	1	540	400

Table 4.1 Values of Available Parameters

One of the difficulties associated with manufacturing strategies in the United States is that many companies procure manufacturing equipment only from commercial vendors and do not consider modifying it to suit their own needs. Custom modification can produce pivotal manufacturing advantages, but also require the company to expand both its planning scope and product development skills. The type of analysis indicated in Table 4.1 may enable an organization to determine the value and return on investment of customizing manufacturing equipment to incorporate advanced sensors and control systems. Alternatively, enterprises with limited research and development resources may decide to contract for development of the optimum equipment in such a way that the sponsor retains proprietary rights for a period of time.

4.7 Network of Sensors Detecting Machinery Faults

A comprehensive detection system for automated manufacturing equipment must be seriously considered as part of the manufacturing strategy. A major component of any effort to develop an intelligent and flexible automatic manufacturing system is the concurrent development of automated diagnostic systems, with a network of sensors, to handle machinery maintenance and process control functions. This will undoubtedly lead to significant gains in productivity and product quality. Sensors and control systems are one of the enabling technologies for the "lights-out" factory of the future.

A flexible manufacturing system often contains a variety of manufacturing work cells. Each work cell in turn consists of various workstations. The flexible manufacturing cell may consist of a CNC lathe or mill whose capabilities are extended by a robotic handling device, thus creating a highly flexible machining cell whose functions are coordinated by its own computer. In most cases, the cell robot exchanges workpieces, tools (including chucks), and even its own gripping jaws in the cell (Fig. 4.1).

4.7.1 Diagnostic Systems

A diagnostic system generally relies on copious amounts of *a priori* and *a posteriori* information. *A priori* information is any previously established fact or relationship that the system can exploit in making a diagnosis. *A posteriori* information is the information concerning the problem at hand for which the diagnosis will be made. The first step in collecting data is to use sensors and transducers to convert physical states into electrical signals. After processing, a signal will be in an appropriate form for analysis (perhaps as a table of values, a time-domain waveform, or a frequency spectrum). Then, the analysis, including correlations with other data and trending, can proceed.



Figure 4.1 Flexible machining cell.

After the data have been distilled into information, the deductive process begins, leading finally to the fault diagnosis. Expert systems have been used effectively for diagnostic efforts, with the diagnostic system presenting either a single diagnosis or a set of possibilities with their respective probabilities, based on the *a priori* and *a posteriori* information.

4.7.2 Resonance and Vibration Analysis

Resonance and vibration analysis is a proven method for diagnosing deteriorating machine elements in steady-state process equipment such as turbomachinery and fans. The effectiveness of resonance and vibration analysis in diagnosing faults in machinery operating at variable speed is not proven, but additional study has indicated good potential for its application in robots. One difficulty with resonance and vibration analysis is the attenuation of the signal as it travels through a structure on the way to the sensors and transducers. Moreover, all motion of the machine contributes to the motion measured by the sensors and transducers, so sensors and transducers must be located as close as possible to the component of concern to maximize the signal-to-noise ratio.

4.7.3 Sensing Motor Current for Signature Analysis

Electric motors generate back electromotive force (emf) when subjected to mechanical load. This property makes a motor a transducer

for measuring load vibrations via current fluctuations. Motor current signature analysis uses many of the same techniques as vibration analysis for interpreting the signals. But motor current signature analysis is nonintrusive because motor current can be measured anywhere along the motor power cables, whereas a vibration sensor or transducer must be mounted close to the machine element of interest. The limited bandwidth of the signals associated with motor drive signature analysis, however, may restrict its applicability.

4.7.4 Acoustics

A qualified operator can tell from the machine-generated noise that a fault is developing. Consequently, It is natural to extend this concept to automatic diagnosis. The operator has access to subtle innate pattern recognition techniques, and thus is able to discern sounds from myriad background noises. Any diagnostic system based on sound would have to be able to identify damage-related sounds and separate the information from the ambient noise. Acoustic sensing (looking for sounds that indicate faults) is a nonlocal noncontact inspection method. Any acoustic technique is subject to outside disturbances, but is potentially a very powerful tool, provided that operating conditions are acoustically repeatable and that the diagnostic system can effectively recognize acoustic patterns.

4.7.5 Temperature

Using temperature as a measurement parameter is common, particularly for equipment running at high speed, where faults cause enough waste heat to raise temperature significantly. This method is generally best for indicating that a fault has occurred, rather than the precise nature of the fault.

4.7.6 Sensors for Diagnostic Systems

Assuming an automated diagnostic system is required, the necessary sensors are normally mounted permanently at their monitoring sites. This works well if data are required continuously, or if there are only a few monitoring locations. However, for those cases where many sites must be monitored and the data need not be continuously received during operation of the flexible manufacturing cell, it may be possible to use the same sensor or transducer, sequentially, in the many locations.

The robot is well-suited to gathering data at multiple points with a limited number of sensors and transducers. This would extend the mandate of the robot from simply moving workpieces and tools within the flexible manufacturing cell (for production) to include moving sensors (for diagnostic inspection).

Within the flexible manufacturing cell, a robot can make measurements at sites inside its work space by taking a sensor or transducer

from a tool magazine, delivering the sensor or transducer to a sensing location, detaching it during data collection, and then retrieving it before moving to the next sensing position.

Sensor mobility does, however, add some problems. First, the robot will not be able to reach all points within the flexible manufacturing cell because its work space is only a subspace of the volume taken up by the flexible manufacturing cell. The manipulator may be unable to assume an orientation desired for a measurement even inside its workspace. Also, the inspection procedure must limit the robot's influence on the measurement as much as possible. Finally, sensors require connectors on the robot end effectors for signals and power. The end effector would have to be able to accommodate all the types of sensors to be mounted on it.

4.7.7 Quantifying the Quality of a Workpiece

If workpiece quality can be quantified, then quality can become a process variable. Any system using product quality as a measure of its performance needs tight error checks so as not to discard product unnecessarily while the flexible manufacturing cell adjusts its operating parameters. Such a system would depend heavily, at first, on the continued supervision of an operator who remains in the loop to assess product quality. Since it is forbidden for the operator to influence the process while it is under automatic control, it is more realistic for the operator to look for damage to product after each stage of manufacture within the cell. In that way, the flexible manufacturing cell receives diagnostic information about product deficiencies close to the time that improper manufacture occurred.

In the future, these quality assessments will be handled by the flexible manufacturing cell itself, using sensors and diagnostic information for process control. Robots, too, will be used for maintenance and physical inspection as part of the regular operation of the flexible manufacturing cell. In the near term, the flexible manufacturing cell robot may be used as a sensor-transfer device, replacing inspectors who would otherwise apply sensors to collect data.

4.7.8 Evaluation of an Existing Flexible Manufacturing Cell Using a Sensing Network

A study was conducted at the Mi-TNO in the Netherlands of flexible manufacturing cells for low-volume orders (often called job production, ranging from 1 to 100 parts per order). The automated manufacturing equipment used in the study consisted of two free-standing flexible manufacturing cells. The first cell was a turning-machine cell; the second, a milling-machine cell. The turning cell contained two armed gantry robots for material handling. The study was mainly conducted to assess the diagnostics for flexible manufacturing systems (FMS). In considering the approach to setting up diagnostics for

an FMS, it was decided to divide development of the diagnostics program into three major blocks (Fig. 4.2):

- Analyzing the existing design
- Setting up diagnostics that are machine-based
- Choosing important points in the flexible manufacturing cells where critical failure can occur and where sensors are mounted
- Setting up a diagnostic decision system for the hardware system
- Establishing a workpiece-based diagnostic system that is actually part of quality control

The flexible manufacturing cells were divided into control volumes (as shown in Fig. 4.3 for the turning cell). For the turning cell,

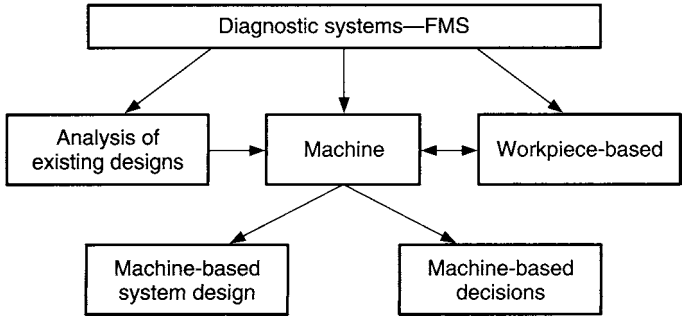
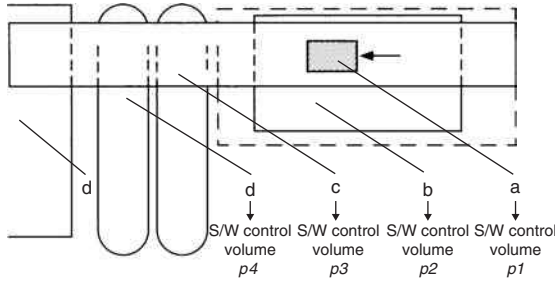


Figure 4.2 Diagnostics for an FMS.

Figure 4.3 Hardware control volumes in a manufacturing cell.



(a) Tool/workpiece interface
(b) Machining volume
(c) Robot volume
(d) Workpiece pallet volume
(e) Tool magazine volume

1. VAX, 8600
2. Microvax
3. PDP 11/231
4. Turning FMC controller—NC

for example, the hardware control volumes were denoted *a*, *b*, *c*, *d*, and *e*, and software control volumes *p1*, *p2*, *p3*, and *p4*. The failures within each of these control volumes were further categorized as:

- Fault detected by an existing sensor
- Fault that could have been detected if a sensor had been present
- Operator learning phase problem
- Failure due to manufacturer problem
- Software logic problem
- Repeat of a problem
- System down for an extended period

4.7.8.1 Software Problems

It is often assumed that disturbances in the cell software control system can be detected and evaluated relatively easily. Software diagnostics are common in most turnkey operations; however, it has been shown that software diagnostics are far from perfect. Indeed, software problems are of particular concern when a revised program is introduced into a cell that has been running smoothly (existing bugs having been ironed out). The availability of the cell plummets in these situations, with considerable loss of production capabilities and a concomitant higher cost. The frequency of software faults increases dramatically when revised packages are introduced to upgrade the system (Fig. 4.4). This is mainly due to human error on the part of either the vendor or the operator. Two possible approaches to software defect prevention and detection are:

- Investigate software methodologies and procedures and recommend alternative languages or more tools as defect prevention measurements. This is tedious and uses ambiguous results because such investigations are not based on data.
- Analyze the problems that result from the current design and develop a solution for each class of problem. This produces less ambiguous solutions and is typically used to solve only immediate problems, thereby producing only short-term solutions.

To identify the types of faults that occur in programs, it is necessary to know what caused the problem and what remedial actions were taken. Program faults can be subdivided into the categories shown next and restated in Fig. 4.5.
Matching faults:

- Wrong names of global variables or constants
- Wrong type of structure or module arguments

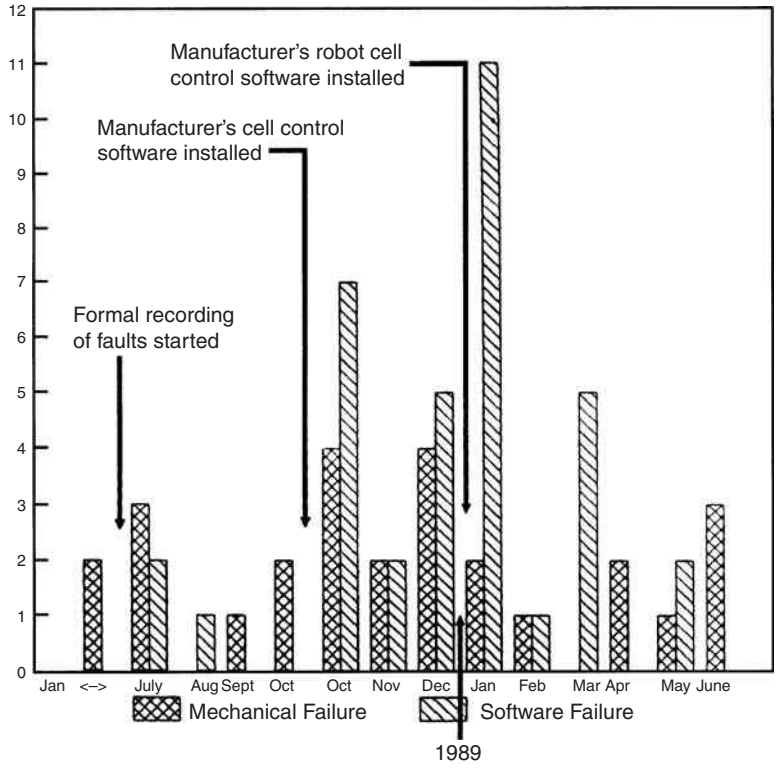


Figure 4.4 Frequency of software faults.

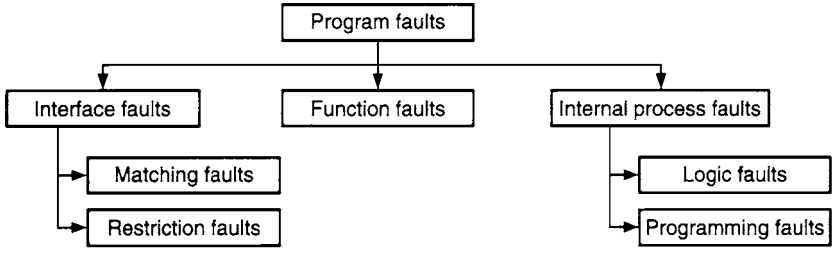
- Wrong number of hardware units
- Wrong procedures for writing data to hardware

Restriction faults:

- Omission of procedures to prevent invalid input or output of data
- Wrong limit value for validity check of arguments

Function faults:

- Omission of saving data to global variables
- Unnecessary calling modules
- Wrong limit value for judging whether or not hardware is set
- Reference to undefined local variables
- Omission of loop variable incrementation
- Logic expressions that are always true



Matching faults

- Wrong names of global variables or constants
- Wrong type of structure or module arguments
- Wrong number of hardware units
- Wrong procedures for writing data to hardware

Restriction faults

- Omission of procedures to prevent invalid input or output of data
- Wrong limit value for validity check of arguments

Function faults

- Omission of saving data to global variables
- Unnecessary calling modules
- Wrong limit value for judging whether or not hardware is set
- Reference to undefined local variables
- Omission of loop variable incrementation
- Logic expressions that are always true

Programming faults

- Comparison of local variables of different types
- Omission of comment marks

Figure 4.5 Program fault categories.

Programming faults:

- Comparison of local variables of different types
- Omission of comment marks

This categorization provides significant insight into the location of fault conditions, the reasons for their occurrence, and their severity. If the faults are in either a hardware (mechanical) or software category, then the frequency of failure by month can be summarized as indicated in Fig. 4.4. Two major and unexpected milestones in the program represented in Fig. 4.4 are the routine introduction of revised cell control software and revised robot control software. In both cases, a substantial increase occurred in the downtime of the flexible manufacturing cell. In an industrial environment, this would have been very costly.

In this study, it was found that interface faults (mismatched data transfer between modules and hardware) were the major cause of

downtime. Also, in the new machine software, it was found faults occurred because the software had not been properly matched to the number of tool positions physically present on the tool magazine. Once, such a fault actually caused a major collision within the machining volume.

4.7.8.2 Detecting Tool Failure

An important element in automated process control is real-time detection of cutting tool failure, including both wear and fracture mechanisms. The ability to detect such failures online allows remedial action to be undertaken in a timely fashion, thus ensuring consistently high product quality and preventing potential damage to the process machinery. The preliminary results from a study to investigate the possibility of using vibration signals generated during face milling to detect both progressive (wear) and catastrophic (breakage) tool failure are discussed next.

4.7.8.2.1 Experimental Technique The experimental studies were carried out using a 3-hp vertical milling machine. The cutting tool was a 381-mm-diameter face mill employing three Carboloy TPC-322E grade 370 tungsten carbide cutting inserts. The standard workpiece was a mild steel plate with a length of 305 mm, a height of 152 mm, and a width of 13 mm. While cutting, the mill traversed the length of the workpiece, performing an interrupted symmetric cut. The sensor sensed the vibration generated during the milling process on the workpiece clamp. The vibration signals were recorded for analysis. Inserts with various magnitudes of wear and fracture (ranging from 0.13 mm to 0.78 mm) were used in the experiments.

4.7.8.2.2 The Manufacturing Status of Parts Figure 4.6 shows typical acceleration versus time histories. Figure 4.7*a* is the acceleration for three sharp inserts. Note that the engagement of each insert in the workpiece is clearly evident and that all engagements share similar characteristics, although they are by no means identical.

Figure 4.7*b* shows the acceleration for the combination of two sharp inserts and one insert with a 0.39-mm fracture. The sharp inserts produce signals consistent with those shown in Fig. 4.6, while the fractured insert produces a significantly different output.

The reduced output level for the fractured insert is a result of the much smaller depth of cut associated with it. It would seem from the time-domain data that use of either an envelope detection or a threshold crossing scheme would provide the ability to automate the detection of tool fracture in a multi-insert milling operation.

Figure 4.7 shows typical frequency spectra for various tool conditions. It is immediately apparent that, in general, fracture phenomena are indicated by an increase in the level of spectra components within

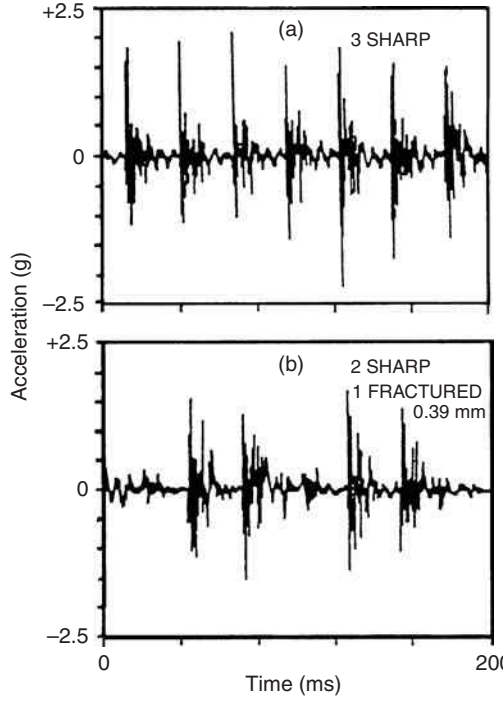


Figure 4.6 Typical acceleration level versus time plot.

the range of 10,000 to 17,000 Hz. A comparison of Fig. 4.7*a* and *b* indicates a notable increase in the spectra around 11 kHz when a single insert fracture of 0.39 mm is present. For two fractured inserts (Fig. 4.7*c*), the peak shifts to around 13.5 kHz. For three fractured inserts (Fig. 4.7*d*), both the 13.5-kHz peak and an additional peak at about 17 kHz are apparent.

Comparing Figs. 4.7*e* and *c*, it is seen that, in general, increasing the size of an insert fracture results in an increase in the spectral peak associated with the failure condition. Usually, the assemblies used to obtain the spectral data are not synchronously averaged. Therefore, it may be possible to improve the spectral data by utilizing a combination of synchronous averaging and delayed triggering to ensure that data representative of each insert in the cutter is obtained and processed.

In general, the acceleration–time histories for the worn inserts do not produce the noticeably different engagement signals evident in a case of fracture. However, by processing the data in a slightly different manner, it is possible to detect evidence of tool wear.

Figure 4.8 shows the amplitude probability density (APD) as a function of time for several tool conditions. The data are averaged for

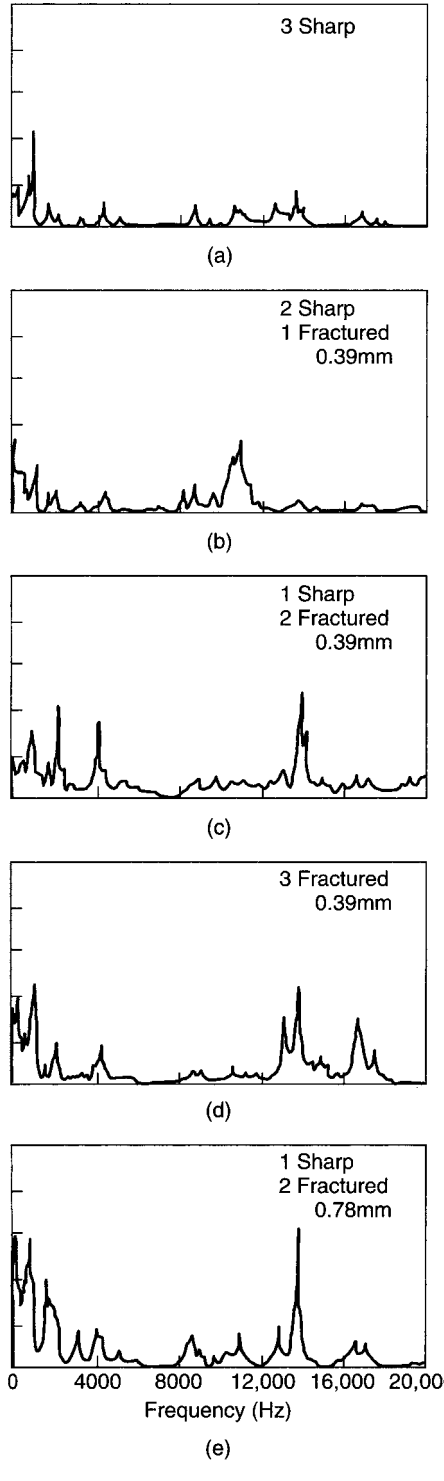


Figure 4.7
Acceleration level versus time for various inserts.

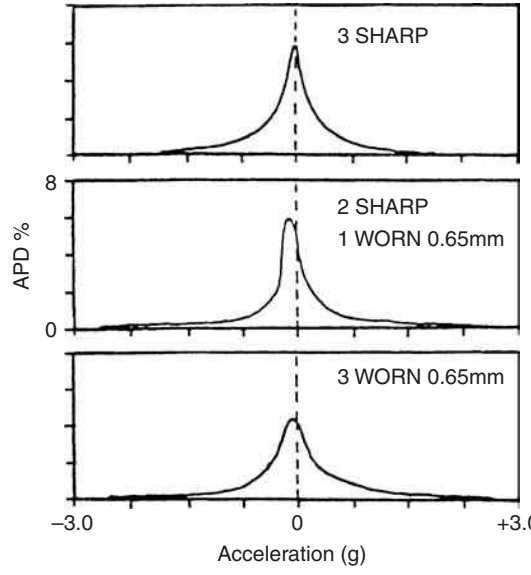


Figure 4.8 Amplitude probability density (APD) for several tool conditions.

eight assemblies. It thus seems possible that insert wear could be detected using such features as the location of the peak in the APD, the magnitude of peak, and the area under specific segments of the distribution.

As with fracture, the presence of insert wear resulted in a significant increase in the spectral components within the 10- to 13-kHz band. Although this would seem to indicate that the presence of flank wear could be detected by simple spectral analysis, it is not yet clear if this method would be sufficiently discriminating to permit reliable determination of the magnitude of flank wear.

4.8 Understanding Computer Communications and Sensors' Role

The evolution in computer software and hardware has had a major impact on the capability of computer-integrated manufacturing concepts. The development of smart manufacturing equipment and sensors and control systems, as well as methods of networking computers, has made it feasible to consider cost-effective computer applications that enhance manufacturing. In addition, this growth has changed approaches to the design of manufacturing facilities.

Messages are exchanged among computers according to various protocols. The open system interconnect (OSI) model developed by the International Standards Organization (ISO) provides such a

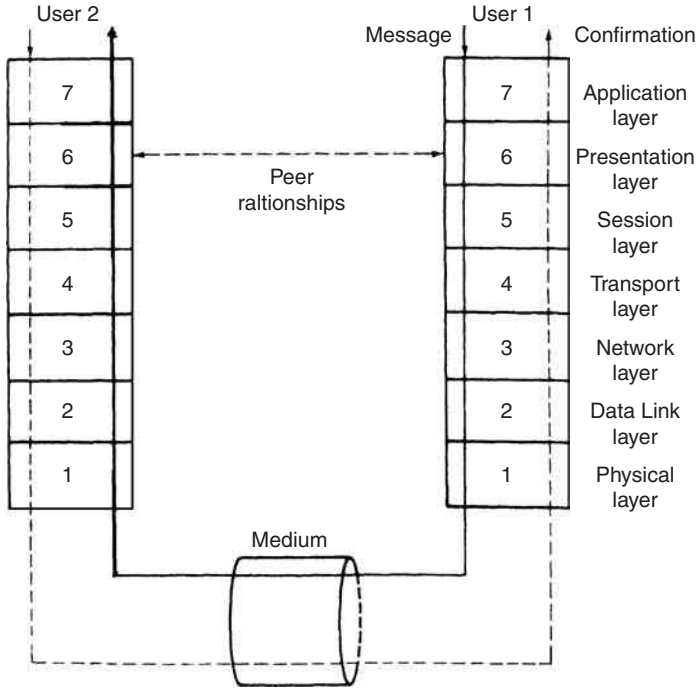


Figure 4.9 Message movement within a computer network.

framework. Figure 4.9 illustrates how two users might employ a computer network. As illustrated, the transfer of information takes place from User 1 to User 2. Each message passes through a series of layers that are associated with message processing:

Sequence of Layers—User 1	Sequence of Layers—User 2
(PhL)1 Physical layer	(PhL)1 Physical layer
(DL)2 Data link layer	(DL)2 Data link layer
(NL)3 Network layer	(NL)3 Network layer
(TL)4 Transport layer	(TL)4 Transport layer
(SL)5 Session layer	(SL)5 Session layer
(PL)6 Presentation layer	(PL)6 Presentation layer
(AL)7 Application layer	(AL)7 Application layer

A message is sent from User 1 to User 2, with message acknowledgment sent back from User 2 to User 1. The objective of this communication system is to transfer a message from User 1 to User 2 and to confirm message receipt. The message is developed at application

layer (AL)7, and passes from there to presentation layer (PL)6, from there to session layer (SL)5, and so forth until the message is actually transmitted over the communication path. The message arrives at physical layer (PhL)1 for User 2 and then proceeds from physical layer (PhL)1 to data link layer (DL)2, to network layer (NL)3, and so forth, until User 2 has received the message. In order for Users 1 and 2 to communicate with one another, every message must pass though all the layers.

The layered approach provides a structure for the messaging procedure. Each time a message moves from User 1 [application layer (AL)7 to physical layer (PhL)1], additional processing and addressing information is added to the beginning or end of the message. As the original message moves on, new information is added to ensure correct communication. Then, as the message moves to User 2 [from physical layer (PhL)1 to application layer (AL)7], this additional information is removed, as illustrated in Fig. 4.9.

The layers work together to achieve "peer" communication. Any information or operation instruction that is added at session layer (SL)5 for User 1 will be addressing session layer (SL)5 for User 2. The layers thus work as peers; each layer has a given operational or addressing task to make sure the message is correctly communicated from User 1 to User 2. Each layer associates only with the layers above and below itself. The layer receives messages from one direction, processes the messages, and then passes them on to the next layer.

4.8.1 Application Layer Communication

Communication begins when User 1 requests that a message be transferred from location 1 to location 2. In developing this message, it may be necessary to have application software that will provide supporting services to the users. This type of service is provided by application layer (AL)7. Common software application tools enable, for example, the transfer of files or the arranging of messages in standard formats.

4.8.2 Presentation Layer Communication

The initial message is passed down from the application layer (AL)7 to presentation layer (PL)6, where any necessary translation is performed to develop a common message syntax that User 2 will understand. If the two different users apply different computer or equipment languages, it will be necessary to define the differences in such a way that the users can communicate with one another. The basic message that began with User 1 is translated to a common syntax that will result in understanding by User 2. Additional information is added to the message at presentation layer (PL)6 to explain to User 2 the nature of the communication that is taking place. An extended message begins to form.

4.8.3 Session Layer Communication

The message now passes from presentation layer (PL)6 to session layer (SL)5. The objective of the session layer is to set up the ability for the two users to converse, instead of just passing unrelated messages back and forth. The session layer will remember there is an ongoing dialog and will provide the necessary linkages between the individual messages so that an extended transfer of information can take place.

4.8.4 Transport Layer Communication

The message then passes from session layer (SL)5 to transport layer (TL)4, which controls the individual messages as part of the communication sequence. The purpose of the transport layer is to make sure individual messages are transferred from User 1 to User 2 as part of the overall communication session that is defined by session layer (SL)5. Additional information is added to the message so that the transport of this particular portion of the communication exchange results.

4.8.5 Network Layer Communication

The message is then passed to network layer (NL)3, divided into packets, and guided to the correct destination. The network layer operates so that the message (and all accompanying information) is tracked and routed correctly so it ends up at User 2. This includes making sure addresses are correct and that any intermediate stops between User 1 and User 2 are completely defined.

4.8.6 Data Link Layer Communication by Fiber Optics or Coaxial Cable

The system now passes a message to data link layer (DL)2, which directs each frame of the message routing in transit. Each frame is loaded into the communication system in preparation for transmission. The message is prepared to leave the User 1 environment and move into the communication medium. The next step is from data link layer (DL)2 to physical layer (PhL)1. At this point, the frame is converted into a series of digital or analog electronic signals that can be placed on the communication medium itself—wire, fiber optics, coaxial cables, or other means—to achieve the transmission of the message from User 1 to User 2.

4.8.7 Physical Layer Communication

The electronic signal arrives at the correctly addressed location for User 2 and is received by physical layer (PhL)1. The physical layer

then converts the electronic signal back to the original frame that was placed on the medium. This extended message is passed up to the data link layer, which confirms that error-free communication has taken place and that the frame has been received at the correct location. Data link layer (DL)2 directs the frame routing in transit. When the full frame is received at data link layer (DL)2, the routing information is removed and the remaining information is transferred to network layer (NL)3. Network layer (NL)3 confirms the appropriate routing and assembles the packets. Afterward, the routing information is stripped from the message. The message is then passed to transport layer (TL)4, which controls the transport of the individual messages. Transport layer (TL)4 confirms that the correct connection has been achieved between User 1 and User 2 and receives the information necessary to achieve the appropriate connection for this particular exchange.

The remaining information is often passed to session layer (SL)5, which interprets whether the message is part of a continuing communication and, if so, identifies it as part of an ongoing conversation. The information is then passed to presentation layer (PL)6, which performs any necessary translation to make sure that User 2 can understand the message as it has been presented. The message is then passed to application layer (AL)7, which identifies the necessary support programs and software necessary for interpretation of the message by User 2. Finally, User 2 receives the message and understands its meaning.

4.8.8 Adding and Removing Information in Computer Networks Based on Open System Interconnect (OSI)

This step-by-step passing of the message from User 1 "down" to the actual communication medium and "up" to User 2 involves adding information to the original message prior to the transfer of information and removing this extra information on arrival (Fig. 4.10). At User 1, additional information is added step by step until the communication medium is reached, forming an extended message. When this information arrives at User 2, the additional information is removed step by step until User 2 receives the original message. As noted, a peer relationship exists between the various levels. Each level communicates only with the level above or below it, and the levels perform a matching service. Whatever information is added to a message by a given level at User 1 is removed from the message by the matching level associated with User 2. Communication is made possible by having User 1's physical layer (PhL)1 communicate with User 2's physical layer (PhL)1, data link layer (DL)2 communicate with data link (DL)2, network layer (NL)3 communicate with network layer (NL)3, and so forth to achieve an exchange between User 1 and User 2.

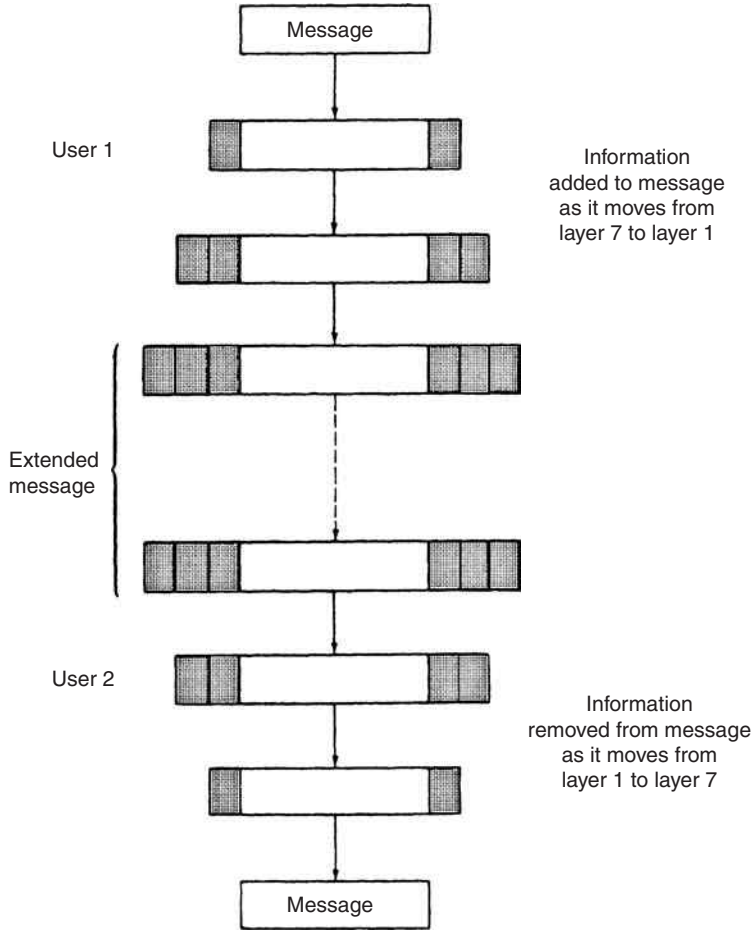


Figure 4.10 Open system interconnect (OSI) layered model.

The two users see only the message that is originated and delivered; they do not see all of the intermediate steps. This is analogous to the steps involved in making a telephone call or sending a letter, in which the two users know only that they have communicated with one another, but do not have any specific knowledge of the details involved in passing a message from one location to another. This orderly and structured approach to communications is useful because it separates the various tasks that must take place. It provides a means for assuring that the methods for processing and addressing messages are always the same at every node. Whether a message is being sent or is being received, a sequential processing activity always takes place.

4.9 Understanding Networks in Manufacturing

A seemingly simple problem in the development of computer networks is to establish the ability to interconnect *any* two computer system elements. This might involve a computer terminal, a modem, a barcode scanner, a printer, sensors, and other system elements that must exchange information in manufacturing. It might seem reasonable that such one-to-one interconnection would follow a well-defined strategy and make maximum use of standards. Unfortunately, for historical reasons and because of the wide diversity of the equipment units that are available today, this situation does not hold. In fact, achieving interconnection between typical system elements can be a disconcerting experience.

4.9.1 RS-232-Based Networks

One of the most common approaches used to interconnect computer system elements in manufacturing is associated with a strategy that was never really intended for this purpose. As noted by Campbell (1984), "In 1969, the EIA (Electronic Industries Association), Bell Laboratories, and manufacturers of communication equipment cooperatively formulated and issued 'EIA RS-232,' which almost immediately underwent minor revisions to become 'RS-232-C.' The RS-232 interface was developed to allow data equipment terminals to be connected to modems so that data could be transmitted over the telephone network. The entire purpose of this standard was to assure that the use of telephone lines for computer communications would be handled in a way acceptable to the telephone company.

Thus, in its general application today, RS-232 is not a standard. It is more a guideline for addressing some of the issues involved in interfacing equipment. Many issues must be resolved on an individual basis, which leads to the potential for difficulties. Essentially, a vendor's statement that a computer system element is RS-232-compatible provides a starting point to consider how the equipment unit might be interconnected. However, the detailed aspects of the interconnection require further understanding of the ways in which equipment units are intended to communicate. Campbell (1984) is a helpful introduction to applying RS-232 concepts.

In a sense, the history of the RS-232 interface illustrates the difficulties associated with creating a well-defined means for allowing the various elements of a computer network to interface. Past experience also indicates how difficult it is to develop standards that will apply in the future to all the difficult situations that will be encountered. As it has evolved, the RS-232 approach to the communications interface is an improvement over the "total anarchy" at position A in Fig. 4.11, but it still leads to a wide range of problems.

An entire computer network can be configured by using combinations of point-to-point RS-232 connections. In fact, a number of

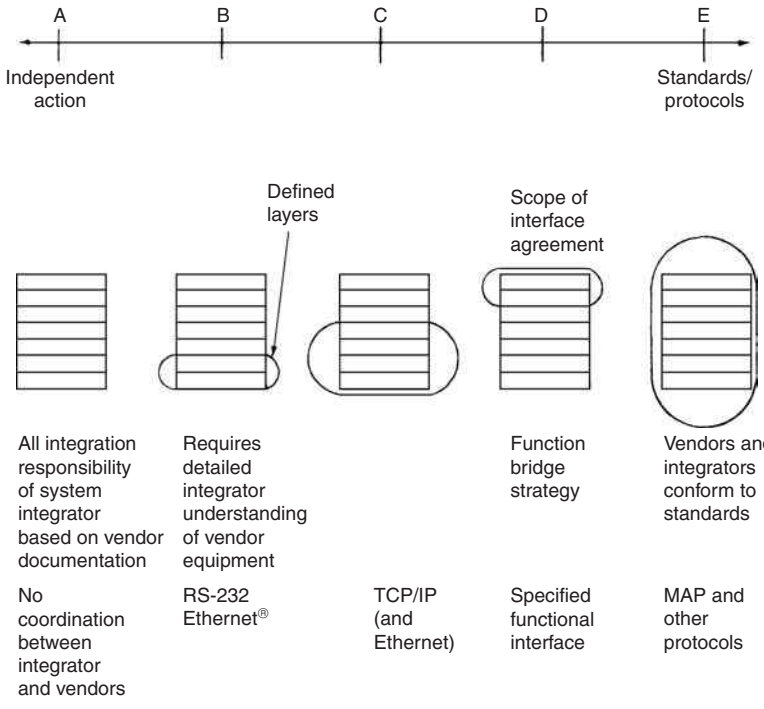


Figure 4.11 Levels of integration of computer networks.

networks of this type are in common use. Such networks require that multiple RS-232 interfaces be present on each equipment unit, as is often the case. Each particular interconnection must be customized for the two equipment units being considered. Thus, a system integrator not only must decide on the elements of the system and how they should perform in a functional sense, but also must develop a detailed understanding of the ways in which RS-232 concepts have been applied to the particular equipment units used for the network. The system integrator incurs a substantial expense in achieving the required interconnections.

It is essential to realize that the RS-232 pseudo-standard only addresses the ability to transfer serially information bit by bit from one system element to another. The higher level communications protocol in Fig. 4.11 is not considered. RS-232 provides the means for running wires or fiber-optic cables from one element to another in order to allow digital signals to be conveyed between system elements. The meaning associated with these bits of information is completely dependent on the hardware and software implemented in the system elements. RS-232 is a widely used approach for computer elements to transfer information. Using RS-232 is certainly much better

than using no guidelines at all. However, because RS-232 does not completely define all of the relationships that must exist in communication links, it falls far short of being a true standard or protocol.

4.9.2 Ethernet

As illustrated in Fig. 4.11, one approach to local area networks (LANs) is to define a protocol for the first two layers of a communication strategy and then allow individual users to define the upper layers. This approach has been widely applied using a method referred to as *Ethernet* [Metcalfe and Boggs (1976), Shock and Hupp (1980), Tanenbaum (1988)], which was created by Xerox Corp.

In every computer communication system, there must be a means of scheduling for each node to transmit onto the network and listen to receive messages. This may be done on a statistical basis. For example, when a unit needs to transmit over the network, it makes an effort to transmit. If another node tries to transmit at the same time, both nodes become aware of the conflict, wait for a random length of time, and try again. It might seem that this would be an inefficient means of controlling a network, since the various nodes are randomly trying to claim the network for their own use, and many collisions may occur. As it turns out, for lower communication volumes, this method works very well. As the number of nodes on the system and the number of messages being exchanged increases, however, the number of collisions between active nodes goes up and reduces the effectiveness of the system (Fig. 4.12).

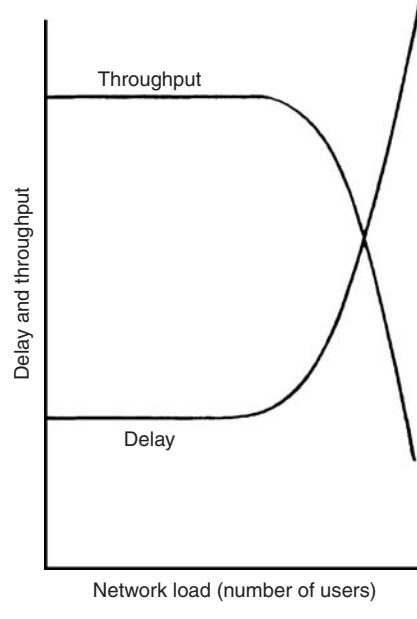


Figure 4.12 Delay and throughput versus network load (number of users).

This type of access control for a computer network is referred to as *carrier-sense multiple-access with collision detection* (CSMA/CD). Ethernet and similar solutions are widely applied to create CSMA/CD networks, particularly in settings in which a maximum waiting time for access to the network does not have to be guaranteed. This type of network is simple to install, and a wide range of hardware and software products are available for support. On the other hand, as indicated in Fig. 4.12, network performance can degrade significantly under high load; therefore, the utility of an Ethernet-oriented network will depend on the particular configuration and loads expected for the network.

4.9.3 Transmission Control Protocol (TCP)/Internet Protocol (IP)

TCP/IP applies to the transport and network layers indicated in Fig. 4.9. TCP/IP thus provides a means for addressing intermediate protocol levels, and in fact is often combined with Ethernet in a communication approach that defines both the lower and middle aspects of the system. TCP/IP functions by dividing any message provided to these middle layers into *packets* of 64 kbytes and then sending packets one at a time to the communication network. TCP/IP must also reassemble the packets in the correct order at the receiving user.

TCP/IP provides a common strategy to use for networking. It allows extension of the Ethernet lower layers to a midlayer protocol on which the final application and presentation layers may be constructed.

4.10 Manufacturing Automation Protocol

The manufacturing automation protocol (MAP) is one of the protocols developed for computer communication systems, and was developed specifically for use in a factory environment. General Motors Corp. has been the leading advocate of this particular protocol. When faced with a need for networking many types of equipment in its factory environment, General Motors decided that a new type of protocol was required. Beginning in 1980, General Motors began to develop a protocol that could accommodate the high data rate expected in its future factories and provide the necessary noise immunity expected for this environment. In addition, the intent was to work within a mature communications technology and to develop a protocol that could be used for all types of equipment in General Motors factories. MAP was developed to meet these needs. The General Motors effort has drawn on a combination of Institute of Electrical and Electronics Engineers (IEEE) and ISO standards, and is based on the open system interconnect (OSI) layered model, as illustrated in Fig. 4.10.

Several versions of MAP have been developed. One difficulty of several has been obtaining agreement among many different countries and vendor groups on a specific standard. Another problem is that the resulting standards are so broad they have become very complex, making it difficult to develop the hardware and software to implement the system and consequently driving up related costs. The early version of MAP addressed some of the OSI layers to a limited degree, and made provision for users to individualize the application layer for a particular use. The latest version of MAP makes an effort to define more completely all the application layer software support as well as the other layers. This has led to continuing disagreements and struggles to produce a protocol that can be adopted by every vendor group in every country to achieve MAP goals.

Because of its complexity, MAP compatibility among equipment units, or *interoperability*, has been a continuing difficulty. MAP has not been applied as rapidly as was initially hoped for by its proponents because of the complexity, costs, and disagreements on how it should be implemented. Assembling a complete set of documentation for MAP is a difficult activity that requires compiling a file of standards organization reports, a number of industry organization reports, and documentation from all the working committees associated with ISO.

4.10.1 Broadband System for MAP Protocol

The MAP protocol was developed with several alternatives for physical layer (PhL)1. It can be implemented through what is called a *broadband* system (Fig. 4.13), and so manufacturing units can talk to one another, transmitted messages are placed on the cable—a *head-end remodulator* then retransmits these messages and directs them to the receiving station. The broadband version of MAP has the highest capabilities because it allows several types of communications to take place on the same cabling at the same time. On the other hand, because of its greater flexibility, a broadband system is more complex and more expensive to install. It requires modems and MAP interface equipment for each item of equipment and a head-end remodulator to serve the entire network. The main cable used for broadband is unwieldy (approximately 25 mm in diameter) and is appropriate only for wiring very large factories. Multiple drop cables can be branched off the main cable for the different MAP nodes. Broadband communication can achieve a high data rate of 10 Mb/s and can split the frequency spectrum to allow several different communications to take place simultaneously. As indicated in Fig. 4.14, the three transmit frequencies and the three receive frequencies are separated from one another in the frequency domain. Three different channels can coexist on the MAP network. The headend remodulator transfers messages from the low frequencies to the high frequencies.

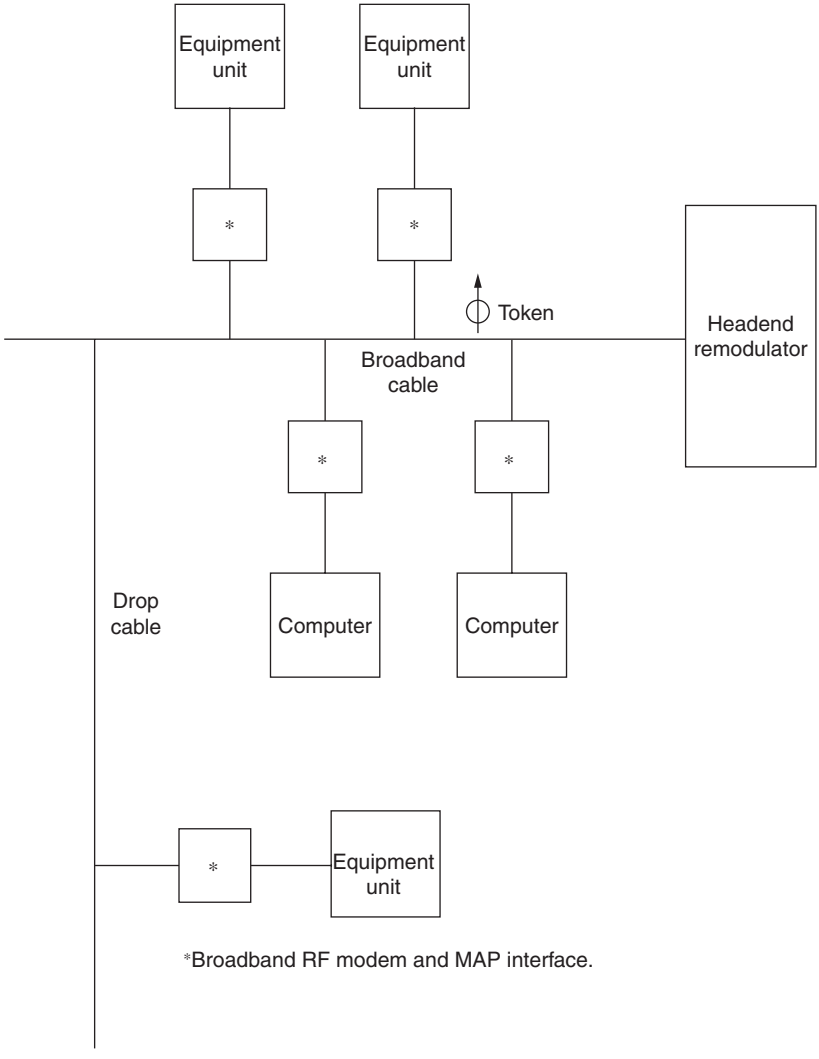


Figure 4.13 Broadband system.



Figure 4.14 Broadband frequencies.

4.10.2 Carrier-Band System for MAP Protocol

Another type of MAP network makes use of a carrier-band approach, which uses somewhat less expensive modems and interface units and does not require heavy duty cable (Fig. 4.15). For a small factory, a carrier-band version of MAP can be much more cost-effective. The carrier-band communication can also achieve a high data rate of 5 to 10 Mb/s, but only one channel can operate on the cable at a given time. A single channel is used for both transmission and reception.

4.10.3 Bridges MAP Protocol

It is possible to use devices called *bridges* to link a broadband factory-wide communication network to local carrier-band networks (Fig. 4.16). The bridge transforms the message format provided on one side to the message formats required on the other. In this sense, a bridge transforms one operating protocol to another.

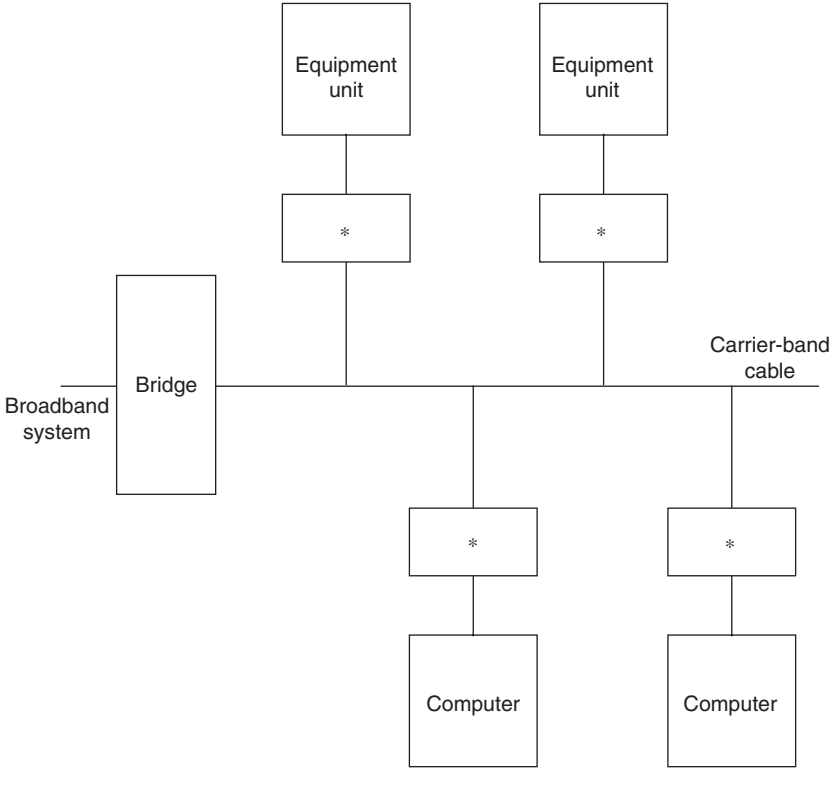


Figure 4.15 Carrier-band system for MAP protocol.

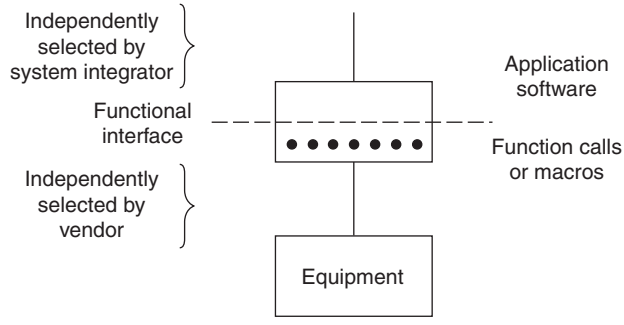


Figure 4.16 Bridges system for MAP protocol.

4.10.4 Token Systems for MAP Protocol

In developing MAP, General Motors was concerned with assuring that every MAP mode would be able to claim control of the network and communicate with other nodes within a certain maximum waiting time. Within this waiting time, every node would be able to initiate its required communications. To do this, MAP implements a *token passing* system (Fig. 4.17). The token in this case is merely a digital word that is recognized by the computer. The token is rotated from node address to node address; a node can claim control of the network and transmit a message only when it holds the token. The token prevents message collisions and also ensures that, for a given system configuration, the maximum waiting time is completely defined (so long as no failures occur). The token is passed around a logic ring defined by the sequence of node addresses, not necessarily by the physical relationships.

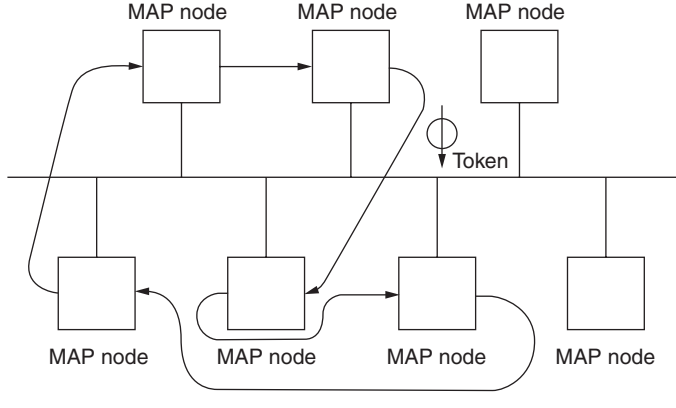


Figure 4.17 Token system for MAP protocol.

The token is a control word, and each MAP node can initiate communication only if it passes the token. It is interesting to note that MAP nodes that are not a part of the logic ring will never possess the token, but they may still respond to a token holder if a message is addressed to them. Token management is handled at the data link layer (DL)2 of the OSI model. This layer controls the features of the token application with respect to how long a token can be held, the sequence of addresses that is to take place, and the amount of time allowed for retrying communications before failure is assumed. If the logic ring is broken at some point—for example, if one equipment unit is no longer able to operate—the other nodes will wait a certain length of time and then will reform the token passing scheme. They will do this by using an algorithm, through which the token is awarded to the highest station address in contention. The rest of the stations on the ring are then determined by the highest-address successors. This process is repeated until the token ring is re-formed.

Physical layer (PhL)1 involves encoding and modulation of the message so that the digital data are transferred into analog and digital communication signals. Each MAP application requires a modem for this purpose. The modem takes the extended message that has been developed at higher layers and modifies it so it can be used to provide an electronic signal to the communications medium. The medium itself provides the means for transferring the signal from User 1 to User 2.

MAP continues to be an important protocol approach for sensors and control systems in computer-integrated manufacturing. For very large factories, the broadband option is available, and for smaller factories a carrier-band system is also available. A number of vendors now produce the hardware and software necessary to establish a MAP network. However, such networks typically are quite high in cost and, because of the complexity of the protocol, can also be difficult to develop and maintain. Thus, MAP is only one of several solutions available to planning and implementation teams.

4.11 Multiple-Ring Digital Communication Network—AbNET

An optical-fiber digital communication network has been proposed to support the data acquisition and control functions of electric power distribution networks. The optical-fiber links would follow the power distribution routes. Since the fiber can cross open power switches, the communication network would include multiple interconnected loops with occasional spurs (Fig. 4.18). At each intersection, a node is needed. The nodes of the communication network

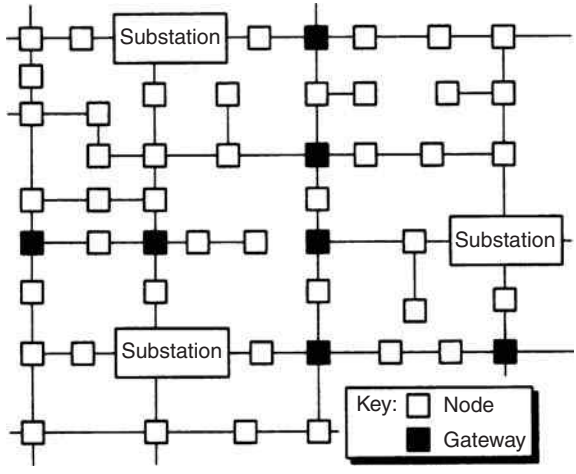


Figure 4.18 Multiple-ring digital communication network—AbNET.

would also include power distribution substations and power controlling units. In addition to serving data acquisition and control functions, each node would act as a repeater, passing on messages to the next node.

Network topology is arbitrary, governed by the power system. The token ring protocols used in single-ring digital communication networks are not adequate. The multiple-ring communication network would operate on the new AbNET protocol, which has already been developed, and would feature fiber optics for this more complicated network.

Initially, a message inserted anywhere in the network would pass from node to node throughout the network, eventually reaching all connected nodes. On the first reception of a message, each node would record an identifying number and transmit the message to the next node. On second reception of the message, each node would recognize the identifying number and refrain from retransmitting the message. This would prevent the endless repetition and recirculating of messages. This aspect of the protocol resembles the behavior of cells in the immune system, which learn to recognize invading organisms on first exposure and kill them with antibodies when they encounter the organisms again. For this reason, the protocol is called *AbNET* after the microbiologists' abbreviation *Ab* for *antibodies*. The *AbNET* protocols include features designed to maximize the efficiency and fault-tolerant nature of the approach. Multiple service territories can be accommodated, interconnected by *gateway* nodes (Fig. 4.18).

The AbNET protocol is expected to enable a network to operate as economically as a single ring that includes an *active monitor* node to

prevent the recirculation of messages. With AbNET, the performance of the proposed network would probably exceed that of a network that relies on a central unit to route messages. Communications would automatically be maintained in the remaining intact parts of the network even if fibers were broken.

For the power system application, the advantages of optical-fiber communications include electrical isolation and immunity to electrical noise. The AbNET protocols augment these advantages by allowing an economical system to be built with topology-independent and fault-tolerant features.

4.12 The Universal Memory Network

The universal memory network (UMN) is a modular digital data communication system that enables computers with differing bus architectures to share 32-bit-wide data between locations up to 3 km apart with less than 1 ms of latency (Fig. 4.19). This network makes it possible to design sophisticated real-time and near-real-time data processing systems without the data transfer bottlenecks that now exist when computers use the usual communication protocols. This enterprise network permits the transmission of the volume of data equivalent to an average encyclopedia each second (40 Mb/s). Examples of facilities that can benefit from the universal memory network include telemetry stations, real-time-monitoring through laser sensors, simulation facilities, power plants, and large laboratories (e.g., particle accelerators), or any facility that shares very large volumes of data. The main hub of the universal memory network uses a *reflection center*—a subsystem containing a central control processor (the reflection controller) and a data bus (the reflection bus) equipped with 16 dual memory parts. Various configurations of host computers, workstations, file servers, and small networks or subnetworks of computers can be interconnected by providing memory speed-bandwidth connectivity. The reflection center provides full duplex communication between the ports, thereby effectively combining all the memories in the network into dual-ported random-access memory. This dual-port characteristic eliminates the CPU overhead on each computer that is incurred with Ethernet.

The reflection bus carries write transfers only and operates at a sustained data rate of 40 Mb/s. This does not include address, error correction, and coordination information, which makes actual universal memory network bus traffic approach 100 Mb/s. The universal memory network can be implemented in copper cables for distances up to 15 m and in fiber optics for distances up to 3 km. A combination of both for media can be used in the same network. Multiple reflection centers can be interconnected to obtain configurations requiring more ports.

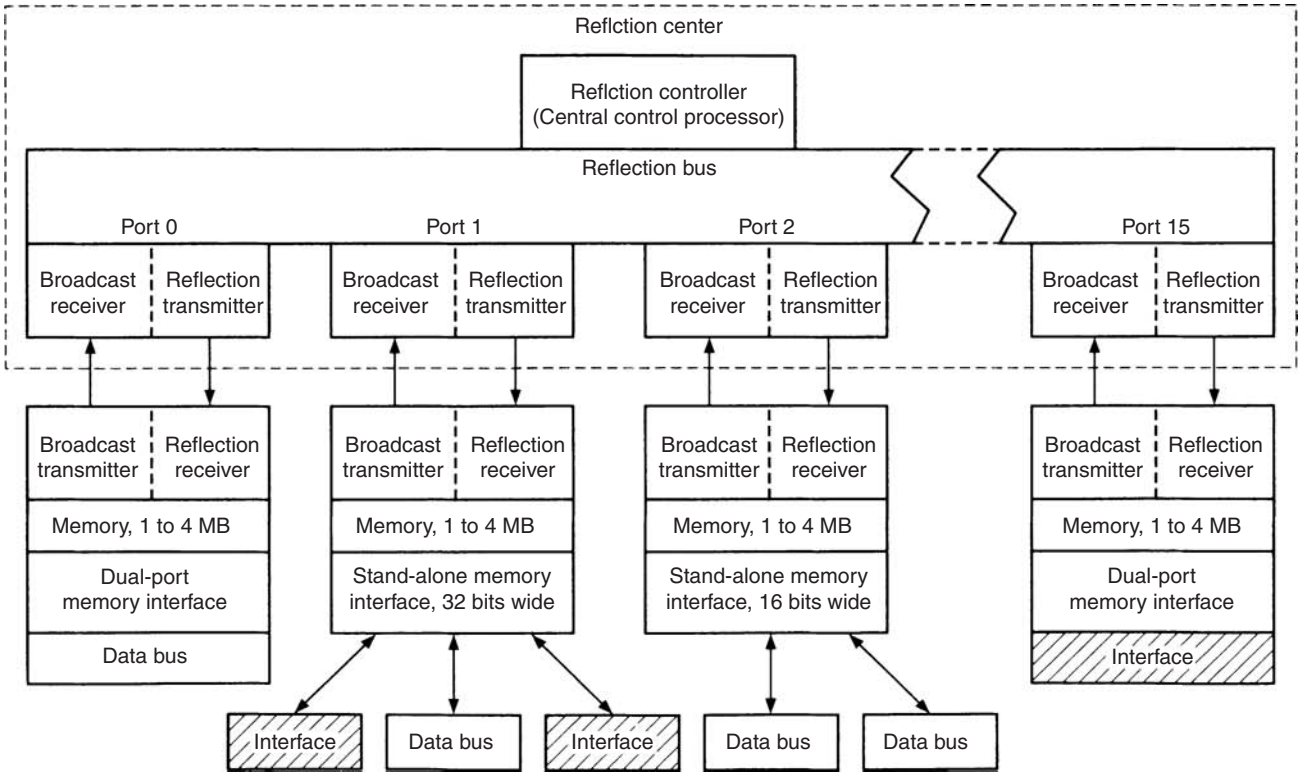


Figure 4.19 Universal memory network (UMN).

In addition to the reflection center of the main hub, the universal memory network includes smaller subsystems called *shared memory interfaces* (SMIs), which make it possible for computers based on different bus architectures (e.g., SELBus, DEC BI, Multi Bus, and VME, or other selected buses) to communicate via the reflection bus. Each host computer is attached to the reflection center by a bus-interface circuit card, which translates the read and write transfers of the host computer to and from the reflection-bus standard. This translation centers around the ordering of bits and conversation used by various vendor architectures to a common strategy required by the 100-ns cycle time of the reflection bus.

The standard memory interface enhances the modular nature of the network. It provides computer memory access to processors of lower cost and enables a large number of workstations to be supported from one reflection center. For example, one reflection center can support up to 12 SMI memory interfaces, each with the capacity to support between 8 and 16 workstations, depending on local hardware configurations. Multiple reflection centers can be interconnected to support even more workstations.

4.13 Satellite Sensor Network

The Satellite Sensor Data Service uses a spread spectrum technology, which is optimized to enable the reliable receipt of large numbers of simultaneous transmissions from multiple users in a distributed sensor network. The system consists of a field device containing a modem, the satellite relay system, network data processing equipment, a view data portal, and a data analysis tool.

The process unfolds as follows:

1. A one-way message containing the GPS location and sensor data is sent from a field device to the low-earth-orbit satellite network. The field device sends multiple messages to increase the reliability of message receipt to 99.99 percent or higher with correct installation in the coverage area.
2. The "bent-pipe" satellite relays messages to a ground station equipped with a Satellite Sensor Data Service decoder. Redundant messages are filtered and the customer only pays for the first message received. The message is decoded and forwarded to the Asset-View data portal.
3. Message data is made available to the end user via the Internet, voicemail, e-mail, or text.

Satelite Link of Global Network transmitting data of businesses and individuals, Fig. 4.20. The coverage area of the message data is illustrated in Fig. 4.21.

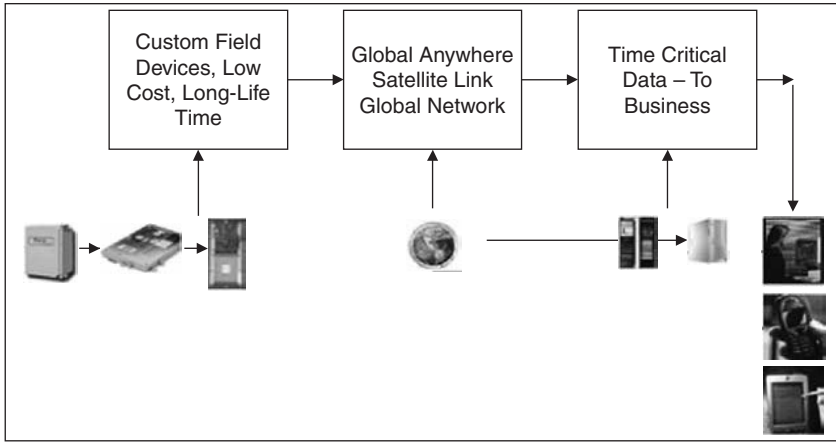


Figure 4.20 Global network data flow.

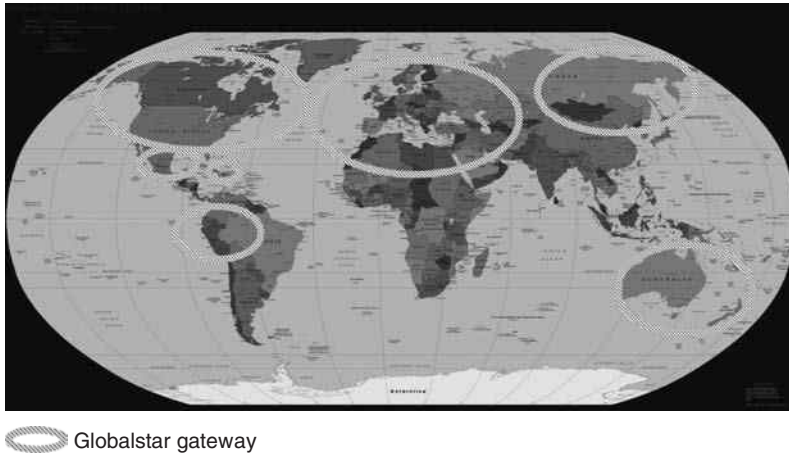


Figure 4.21 Map indicates areas of potential coverage only. Actual availability in any particular country may be subject to government verification.

Further Reading

Ames, J. G., "Which Network Is the Right One," *Manufacturing Engineering*, 56, May 1988.

Borggraaf, P., "Semiconductor Factory Automation: Current Theories," *Semiconductor International*, 88, October 1985.

Bux, W., "Local Area Subnetworks: A Performance Comparison," *IEEE Trans. Communications*, COM-29(10):1465 (1981).

Campbell, J., *The RS-232 Solution*, Sybex, Inc., Alameda, Ca., 1984.

Kaminski, A. M., Jr., "Protocols for Communicating in the Factory," *IEEE Spectrum*, 65, April 1986.

Kleinrock, L., and S. S. Lam, "Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation," *IEEE Trans. Communications*, COM-23(4):410 (1975).
Metcalfe, R. M., and D. R. Boggs, "Ethernet: Distributed Packet-Switching for Local Computer Networks," *Communications of the ACM*, 19:395 (1976).
Shock, J. F., and J. A. Hupp, "Measured Performance of an Ethernet Local Network," *Communications of the ACM*, 23:711 (1980).
Tanenbaum, A. S., *Computer Networks*, 2d ed., Prentice-Hall, Englewood Cliffs, N.J., 1988.
Talvage, J., and R. G. Hannam, *Flexible Manufacturing Systems in Practice: Application Design and Simulation*, Marcel Decker, New York, 1988.
"Support for The Manufacturing Floor," *Manufacturing Engineering*, 29–30, March 1989.
Vollmann, T. E., W. L. Berry, and D. C. Whyback, *Manufacturing Planning and Control Systems*, Richard D. Irwin, Homewood, Ill., 1984.
Voelcker, J., "Helping Computers Communicate," *IEEE Spectrum*, 61, March 1986.
Warndorf, P. R., and M. E. Merchant, "Development and Future Trends in Computer Integrated Manufacturing in the USA," *International Journal in Technology Management*, 1(1–2):162 (1986).
Wick, C., "Advances in Machining Centers," *Manufacturing Engineering*, 24, October 1987.
Wittry, E. J., *Managing Information Systems: An International Approach*, Society of Manufacturing Engineers, Dearborn, Mich., 1987.

This page intentionally left blank

CHAPTER 5

The Role of Sensors and Control Technology in Computer-Integrated Manufacturing

5.1 Introduction

According to various studies conducted in the United States, nearly 50 percent of the productivity increase during the period 1950–1990 was due to technological innovation. That is, the increase was due to the introduction of high-value-added products and more efficient manufacturing processes, which in turn have caused the United States to enjoy one of the highest living standards in the world. However, unless the United States continues to lead in technological innovation, the relative living standard of the country will decline over the long term.

This clearly means that the United States has to invest more in research and development, promote scientific education, and create incentives for technological innovation. In the R&D arena, the United States has been lagging behind other nations: about 1.9 percent of the U.S. gross national product (GNP) (versus about 2.6 percent of the GNP in Japan and West Germany) goes for R&D. The introduction of the computer-integrated manufacturing (CIM) strategy in U.S. industry has begun to provide a successful flow of communication,

which may well lead to a turnaround. *Sensors and control systems in manufacturing* are a powerful tool for implementing CIM. Current world business leaders view CIM as justifying automation to save our standard of living. The successful implementation of CIM depends largely on creative information gathering through sensors and control systems, with information flow as feedback response. Information gathering through sensors and control systems is *imbedded* in CIM, which provides manufacturers with the ability to react more quickly to market demands and achieve levels of productivity previously unattainable.

Effective implementation of sensors and control subsystems within the CIM manufacturing environment will enable the entire manufacturing enterprise to work together to achieve new business goals.

Database management systems will continue to manage large data volumes. Thus, efficient algorithms for accessing and manipulating large sets and sequences will be required to provide acceptable performance. The advent of object-oriented and extensible database systems will not solve this problem. On the contrary, modern data models exacerbate it: In order to manipulate large sets of complex objects as efficiently as today's database systems manipulate simple records, query processing algorithms and software will become more complex, and a solid understanding of algorithm and architectural issues is essential for the designer of database management software. This survey provides a foundation for the design and implementation of query execution facilities in new database management systems. It describes a wide array of practical query evaluation techniques for both relational and post-relational database systems, including iterative execution of complex query evaluation plans, the duality of sort- and hash-based set matching algorithms, types of parallel query execution and their implementation, and special operators for emerging database application domains.

5.2 The CIM Plan

This chapter will address implementation of a CIM plan through the technique of modeling.

A model can be defined as a tentative description of a system or theory, and accounts for many of the system's known properties. An *enterprise model* can be defined (in terms of its functions) as the function of each area, the performance of each area, and the performance of these areas interactively. The creation of a model requires an accurate description of the needs of an enterprise.

In any manufacturing enterprise, there is a unique set of business processes that are performed in order to design, produce, and market the enterprise's products. Regardless of how unique an enterprise or

its set of processes is, it shares with others the same set of high-level objectives. To attain the objectives, the following criteria must be met:

- Management of manufacturing finances and accounting
- Development of enterprise directives and financial plans
- Development and design of product and manufacturing processes utilizing adequate and economical sensors and control systems
- Management of manufacturing operations
- Management of external demands

5.2.1 The CIM Plan in Manufacturing

In manufacturing, CIM promotes customer satisfaction by allowing order entry from customers, faster response to customer inquiries and changes, via electronic sensors, and more accurate sales projections.

5.2.2 The CIM Plan in Engineering and Research

In engineering and research, CIM benefits include quicker design, development, prototyping, and testing; faster access to current and historical product information; and the paperless release of products, processes, and engineering changes to manufacturing.

5.2.3 The CIM Plan in Production Planning

In production planning, CIM offers more accurate and realistic production scheduling that requires less expediting, canceling, and rescheduling of production and purchase orders.

In plant operations, CIM helps control processes, optimize inventory, improve yields, manage changes to product and processes, and reduce scrap and rework. CIM also helps in utilizing people and equipment more effectively, eliminating production crises, and reducing lead time and product costs.

5.2.4 The CIM Plan in Physical Distribution

In physical distribution, where external demands are satisfied with products shipped to the customer, CIM helps in planning requirements; managing the flow of products; improving efficiency of shipping, vehicle, and service scheduling; allocating supplies to distribution centers; and expediting processing of returned goods.

5.2.5 The CIM Plan for Business Management

For business management activities such as managing manufacturing, finance, and accounting, and developing enterprise directives and financial plans, CIM offers better product cost tracking, more accuracy in financial projections, and improved cash flow.

5.2.6 The CIM Plan for the Enterprise

For the enterprise as a whole, these advantages add up to faster release of new products, shorter delivery times, optimized finished goods inventory, shorter production planning and development cycles, reduced production lead time, improved product quality, reliability and serviceability, increased responsiveness, and greater competitiveness. In effect, CIM replaces an enterprise's short-term technical improvements with a long-term strategic solution.

The advantages of CIM with sensors and control systems are not just limited to the four walls of an enterprise. It can also deliver real productivity gains in the outside world. For example, suppliers will be able to plan production, schedule deliveries, and track shipments more efficiently. Customers will benefit from shorter order-to-delivery times, on-time deliveries, and less expensive, higher-quality products.

5.3 The Manufacturing Enterprise Model

The integration and productivity gains made possible by CIM with sensors and control systems are the key to maintaining a competitive edge in today's manufacturing environments. The enterprise model defines an enterprise in terms of its functions. In a traditional enterprise that relies on a complex organization structure, operations and functional management are divided into separate departments, each with its own objectives, responsibilities, resources, and productivity tools.

Yet, for the enterprise to operate profitably, these departments must perform in concert. Sensors and control systems that improve one operation at the expense of another, and tie up the enterprise's resources, are counterproductive. New sensors and control systems in CIM can create a systematic network out of these insulated pockets of productivity. But to understand how, one must examine the elements of an enterprise model and see how its various functional areas work—independently and with each other.

Creating a model of the enterprise can help expose operations that are redundant, unnecessary, or even missing. It can also help determine which information is critical to a successful implementation once effective sensors and control systems are incorporated.

Obviously, this model is a general description. Many industry-unique variations to the model exist. Some enterprises may not require all of the functions described, while others may require more than those listed. Still other enterprises may use the same types of functions, but group them differently.

For example, in the aerospace industry, life-cycle maintenance of products is an essential requirement and may require extensions to

the model. In the process industry, real-time monitoring and control of the process must be included in the model.

Computer-integrated manufacturing harnesses sensors and control system technology to integrate these manufacturing and business objectives. When implemented properly, CIM can deliver increased productivity, cost-efficiency, and responsiveness throughout the enterprise. CIM can accomplish these objectives by addressing each of the major functional areas of the manufacturing enterprise:

- Marketing
- Engineering research and development of sensors in CIM strategy
- Production planning
- Plant operations incorporating sensors and shop floor control systems
- Physical distribution
- Business management

Integrating these functions and their resources requires the ability to share and exchange information about the many events that occur during the various phases of production; manufacturing systems must be able to communicate with the other information systems throughout the enterprise (Fig. 5.1). There must also be the means to capture data close to the source, then distribute the data at the division or corporate level, as well as to external suppliers, subcontractors, and even customers.

To meet these needs, the CIM environment requires a dynamic network of distributed functions. These functions may reside on independent system platforms and require data from various sources. Some may be general-purpose platforms, while others are tailored to specific environments. But the result is an environment that encompasses the total information requirements of the enterprise, from developing its business plans to shipping its products.

With this enterprise-wide purpose, CIM can deliver its benefits to all types of manufacturing operations, from plants that operate one shift per day to processes that must flow continuously, from unit fabrication and assembly to lots with by-products and coproducts. These benefits can also be realized in those enterprises where flexible manufacturing systems are being used to produce diversified products over shorter runs, as well as in those that use CIM with sensors and control systems to maintain an error-free environment.

By creating more efficient, more comprehensive management information systems through sensors and control systems, CIM supports management efforts to meet the challenges of competing effectively in today's world markets.

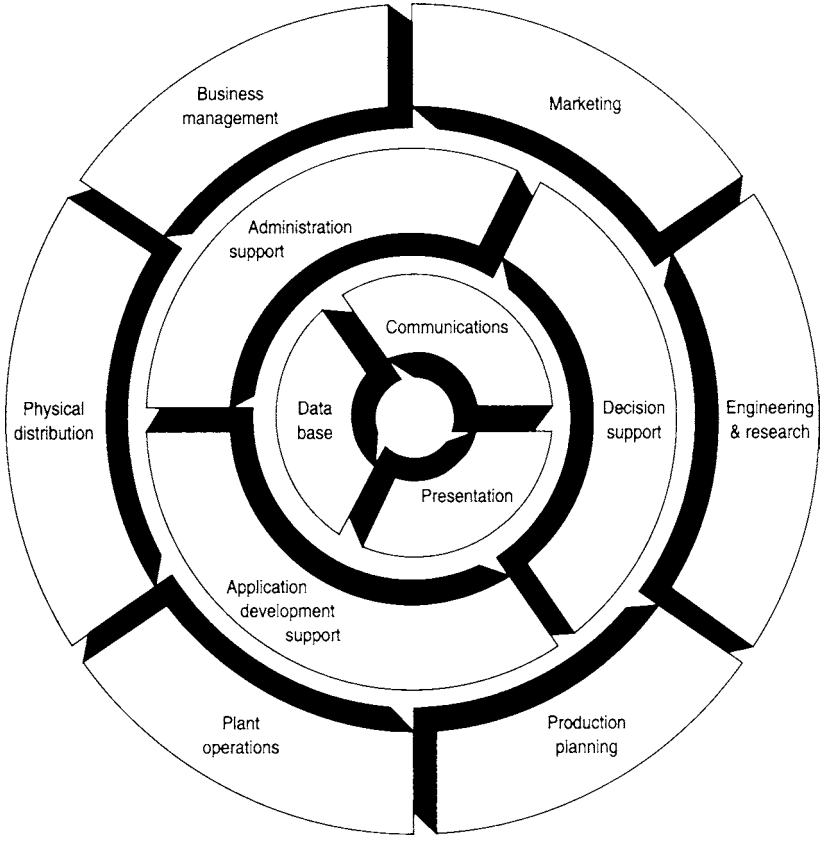


Figure 5.1 Major functional areas of manufacturing.

5.3.1 Marketing

Marketing acts as an enterprise's primary contact with its customer (Fig. 5.2). To help meet the key objectives of increasing product sales, a number of functions are performed within marketing. These include market research; forecasting demand and sales; analyzing sales; tracking performance of products, marketing segments, sales personnel, and advertising campaigns; developing and managing marketing channels; controlling profits and revenues; and managing sales personnel, sales plans, and promotion. Input comes from business management and customers. Output goes to customers, product development, customer order servicing, and master production planning.

The information handling requirements of marketing include monumental texts and graphics as well as queries and analysis of internal and external data. The internal data is gathered through a variety of software routines.

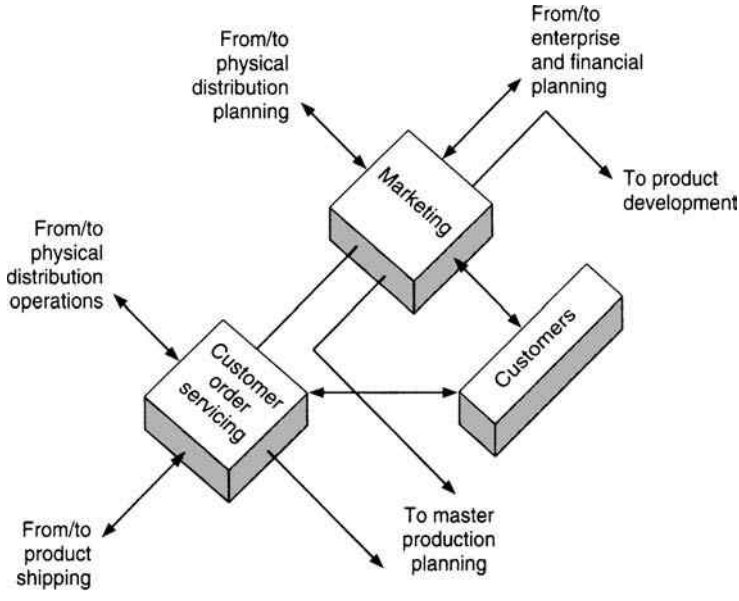


Figure 5.2 Marketing relationships.

Customer order servicing involves entering and tracking customer orders. This can be for standard products or custom-designed products. Other customer order servicing activities include providing product quotations, checking customer credit, pricing product, allocating order quantities, and selecting shipments from distribution centers.

Input to this area includes order and forecast data from marketing or directly from customers as well as available-to-promise data from production planning. Output can include allocation of all orders, quotations for custom products, communication with production engineering regarding custom products, order consideration, and shipping releases. Customer service will significantly improve through electronic data interchange (EDI).

5.3.2 Engineering and Research

The engineering and research areas of an enterprise can be broken down into separate activities (Fig. 5.3). Each of these has its own special needs, tools, and relationships to other areas.

The research activities include investigating and developing new materials, products, and process technology. Information processing needs include complex analyses, extensive texts, imaging, graphics, and videos.

Research input comes from such outside research sources as universities, trade journals, and laboratory reports. Then, research must

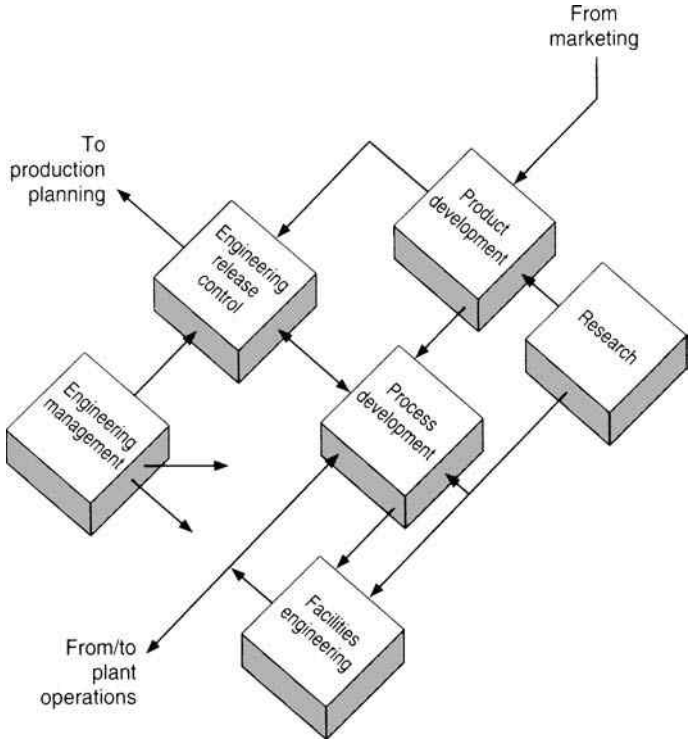


Figure 5.3 Engineering and research.

communicate its findings to three other functional areas—product development, process development, and facilities engineering.

5.3.2.1 Product Development

In today's increasingly competitive manufacturing markets, creation of new material and production technologies is essential for the successful development of products. The product development area uses these materials and production technologies to design, model, simulate, and analyze new products.

Product development activities include preparing product specifications and processing requirements, drawings, materials or parts lists, and bills of material for new products or engineering changes.

In this area, laboratory analysis tools, computer-aided design/computer-aided engineering (CAD/CAE) tools, and group technology (GT) applications are helping to reduce product development time, increase productivity, and improve product quality.

Product development comes from marketing, research, and plant operations. Its output—product specifications, manufacturing control requirements, drawings, text, and messages—is directed to process development and engineering release control.

5.3.2.2 Process Development

This functional area creates process control specifications, manufacturing routings, quality test and statistical quality control specifications, and numerical control (NC) programming. It also validates the manufacturability of product designs. Computer-aided process planning (CAPP) programs and group technology applications for routing similar parts have helped streamline these functions. Expert systems have also been used to supplement traditional product testing and defect-analysis processes. This area is also responsible for the application of new manufacturing technologies such as work cells, conveyer systems, and robotics. Sensors and control systems play a fundamental role within this work-cell structure.

Process development receives input from research and product development, as well as statistical process data from plant operations. Output includes providing routings, process control algorithms, and work-cell programming to plant operations by way of engineering release control.

5.3.2.3 Facilities Engineering

The chief responsibility of facilities engineering is the installation of plant automation incorporating new equipment with sensors and control systems, material flow, inventory staging space, and tools. Tools may include driverless material handling equipment, conveyers, and automated storage and retrieval systems. This area is also responsible for plant layout and the implementation of such plant services and utilities as electricity, piping, heat, refrigeration, and light.

Input to facilities engineering is from research and process development. Output, such as plant layouts, changes of schedule, and forecasts of new equipment availability, goes to plant operations.

5.3.2.4 Engineering Release Control

This function involves the coordination of the release of new products, processes, tools, and engineering changes to manufacturing. A major checkpoint in the product cycle is obtaining assurance that all necessary documentation is available, after which the information is released to manufacturing.

Input is received from product and process development activities. Specific output, including product and tool drawings, process and quality specifications, NC programs, and bills of material, is transferred to production planning and plant operations.

5.3.2.5 Engineering Management

Among the activities of engineering management are introducing engineering releases, controlling product definition data, estimating cost, entering and controlling process data, and defining production resources.

Input and output for engineering management include exchanging designs and descriptions with engineering design; reviewing product data, routings, and schedules with production planning; and accepting engineering change requests from plant operations.

5.3.3 Production Planning

Production planning can be viewed as five related functional areas (Fig. 5.4).

5.3.3.1 Master Production Planning

In master production planning, information is consolidated from customer order forecasts, distribution centers, and multiple plans in order to anticipate and satisfy demands for the enterprise's products. Output includes time-phased requirements sent to the material planning function, as well as the final assembly schedule.

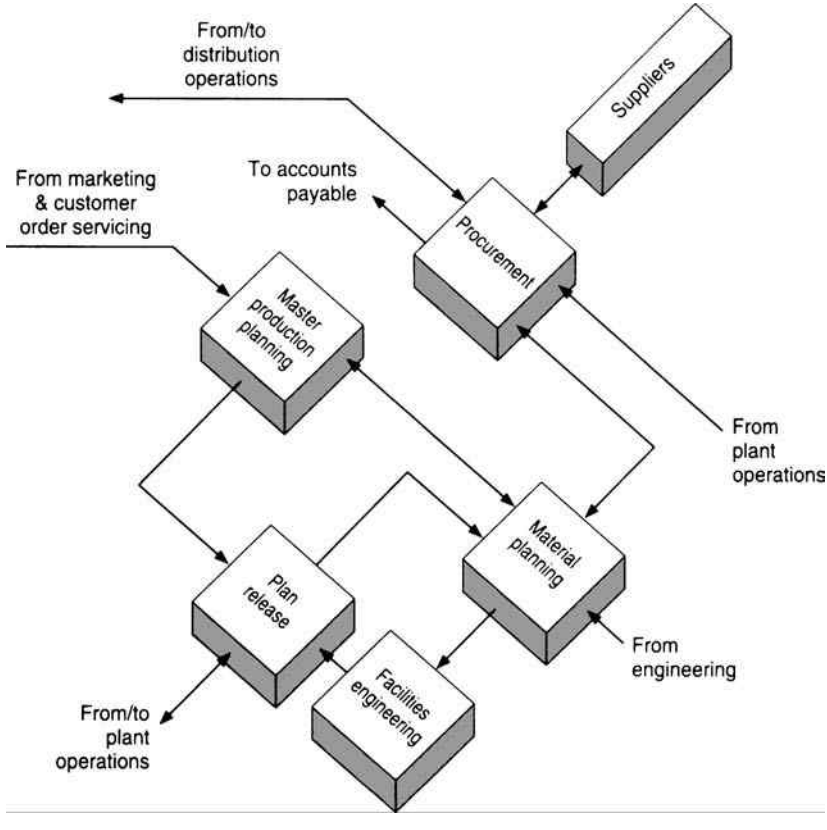


Figure 5.4 Production planning.

5.3.3.2 Material Planning and Resource Planning

Material planning and resource planning are two areas that require timely and accurate information—demand schedules, production commitment, inventory and work-in-progress status, scrap, actual versus planning receipts, shortages, and equipment breakdowns—in order to keep planning up to date with product demands.

Product and process definition data come from the engineering areas. Output is to plant operation and procurement and includes production schedules, order releases, and plans for manufactured and purchased items.

5.3.3.3 Procurement

Procurement involves selecting suppliers and handling purchase requirements and purchase orders for parts and materials. Among the input, is material requirements from material planning and just-in-time delivery requests from plant operations. Other input includes shipping notices, invoices, and freight bills.

Output to suppliers includes contracts, schedules, drawings, purchase orders, acknowledgments, requests for quotations, release of vendor payments, and part and process specifications. In order to streamline this output, as well as support just-in-time concepts, many enterprises rely on sensors for electronic data interchange with vendors.

5.3.3.4 Plant Release

The functions of this area can vary, depending on the type of manufacturing environment. In continuous-flow environments, for example, this area produces schedules, recipes to optimize use of capacity, specifications, and process routings.

For job-shop fabrication and assembly environments, this area prefers electronic or paperless-shop documents consisting of engineering change levels; part, assembly, setup, and test drawings and specifications; manufacturing routings; order and project control numbers; and bar codes, tags, or order-identification documents.

However, large-volume industries are evolving from typical job-shop operations to continuous-flow operations, even for fabrication and assembly-type processes.

Input—typically an exploded production plan detailing required manufacturing items—comes from material planning. Output—schedules, recipes, or shop packets—is used in plant operations for scheduling.

5.3.4 Plant Operations

Plant operations can be described in terms of nine functions (Fig. 5.5).

5.3.4.1 Production Management

The production management area provides dynamic scheduling functions for the plant floor by assigning priorities, personnel, and

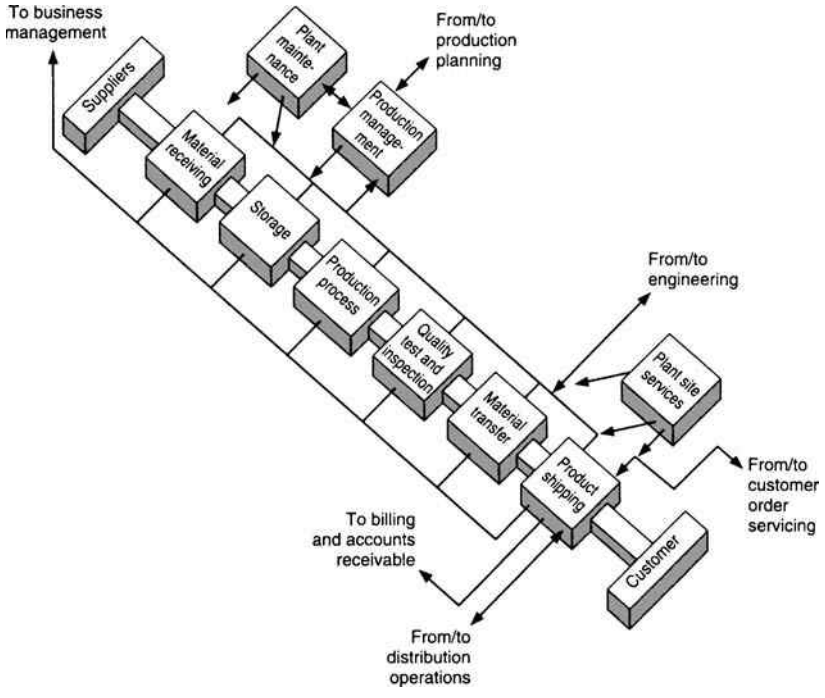


Figure 5.5 Plant operations.

machines. Other activities include sending material and tool requests for just-in-time delivery.

Input for production management includes new orders and schedules from production planning and real-time feedback from plant operation. Output includes dynamic schedules and priorities used to manage operations on the plant floor.

5.3.4.2 Material Receiving

The material receiving function includes accepting and tracking goods, materials, supplies, and equipment from outside suppliers or other locations within the enterprise. Input to this area includes receiving reports and purchase order notices.

Output includes identifying receipts with appropriate documentation, then routing materials to the proper destination. Data are also sent to accounting, procurement, and production management. Production management can also direct materials based on more immediate needs.

5.3.4.3 Storage

Storage represents inventory management—where materials are stored and are accessible to the proper production locations. The materials

include finished goods, raw materials, parts, supplies, work-in-progress, and tools, as well as nonproduction materials and equipment.

Storage functions include preparing item identification and storage tags, managing storage locations, processing pick requests, reporting picks and kit activities, and planning physical inventory cycles and reporting counts.

Storage input includes storage and picking requests from production management scheduling functions. Output includes receiving and disbursement reports for use in production management and accounting.

5.3.4.4 Production Process

Production process functions include managing the production process, processing materials, fabricating parts, grading or reworking components, assembling final products, and packaging for distribution.

One of today's trends in fabrication and assembly processes is toward continuous processing, such as continuous movement with minimal intermediate inventories. This can be described with such terms as continuous-flow manufacturing, flexible manufacturing cells, just-in-time logistics, and group technology. Unfortunately, in many instances, these automation efforts are autonomous, without regard to the other functions of the enterprise.

The information handling needs of the production process can include analog and digital data, text, graphics, geometries, applications programs—even images, video, and voice. Processing this information may require subsecond access and response time.

Input to this area includes shop documents, operator instructions, recipes, and schedules from production management as well as NC programs from process development. Output consists of material and tool requests, machine maintenance requests, material transfer requests, production, and interruption reports for production management, production and labor reports for cost accounting and payroll, and statistical process reports for production management and process development.

5.3.4.5 Quality Test and Inspection

Testing items and products to assure the conformity of specifications is the main activity in quality test and inspection. This includes analyzing and reporting results quickly by means of metrological sensors and control systems, in order to reduce scrap and rework costs.

Quality test and product specifications are input from engineering. Chief output includes purchased-item inspection results to procurement, manufactured-item inspection and product test results to production process and production management, quality test and inspection activity reports to cost accounting, and rejected part and product dispositions to material handling.

5.3.4.6 Material Transfer

Material transfer involves the movement of materials, tools, parts, and products among the functional areas of the plant. These activities may be manual. Or they may be semiautomated, using control panels, forklifts, trucks, and conveyers. Or they may be fully automated, relying on programmable logical controllers, distribution control systems, stacker cranes, programmed conveyers, automated guided vehicles, and pipelines.

Input in this area may be manual requests or those generated by the system. Output includes reporting completed moves to production management.

5.3.4.7 Product Shipping

The product shipping area supports the movement of products to customers, distributors, warehouses, and other plants. Among the activities are selecting shipment and routing needs, consolidating products for a customer or carrier order, preparing shipping lists and bills of lading, reporting shipments, and returning goods to vendors.

The primary input is from customer order servicing, and it includes the type of product and the method and date of shipment. Output includes reporting shipment dates to customer order servicing, billing, and accounts receivable.

5.3.4.8 Plant Maintenance

Plant maintenance includes those functions that ensure the availability of production equipment and facilities. Maintenance categories include routine, emergency, preventive, and inspection services. In addition, many of today's advanced users of systems are moving toward diagnostic tools based on expert systems, which reduce equipment downtime. Input (maintenance requests) can be initiated by plant personnel, a preventive maintenance and inspection system, or a process and equipment monitoring system. Output includes requests for purchase of maintenance items, schedules for maintenance for use in production management, requests for equipment from facilities engineering, and maintenance work order costs to cost accounting.

5.3.4.9 Plant Site Services

The final area of plant operation is plant site services. Input received and output provided cover such functions as energy supply and utilities management, security, environmental control, grounds maintenance, and computer and communications installations.

5.3.5 Physical Distribution

Physical distribution can be viewed as having two functional areas (Fig. 5.6).

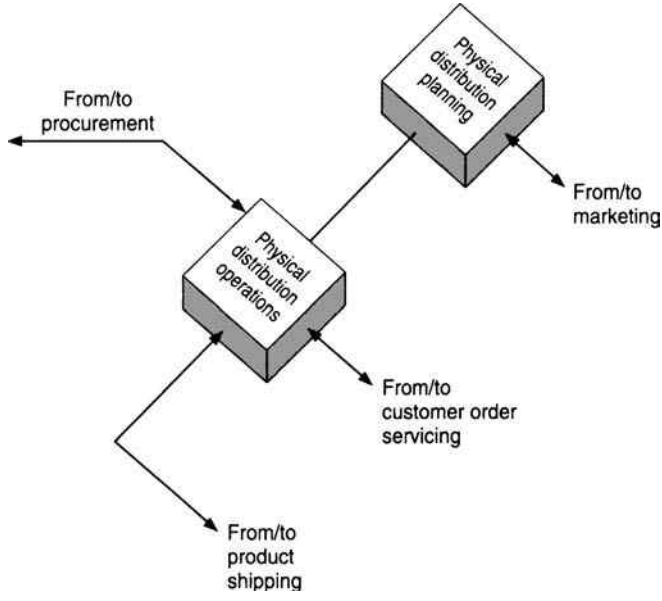


Figure 5.6 Physical distribution.

5.3.5.1 Physical Distribution Planning

This involves planning and control of the external flow of parts and products through warehouses, distribution centers, other manufacturing locations, and points of sale. These functions may also include allocating demand, planning finished goods inventory, and scheduling vehicles. Some industries require another major set of functions that relate to logistics reports. This includes spare parts, maintenance, training, and technical documentation.

Input and output are exchanged with marketing and physical distribution operations. The information handling requirements of physical distribution planning are usually medium to heavy, especially if multiple distribution centers exist.

5.3.5.2 Physical Distribution Operations

The physical distribution operations area includes receiving, storing, and shipping finished goods at the distribution center or warehouse. Receipts arrive from plants or other suppliers, and shipments are made to customers and dealers. Other functions can include scheduling and dispatching vehicles, processing returned goods, and servicing warranties and repairs.

Input is received from plant site product shipping, procurement, physical distribution planning, and customer order servicing. Output includes acknowledgments to plant site product shipping and procurement, as well as data for updates to physical distribution planning and customer order servicing.

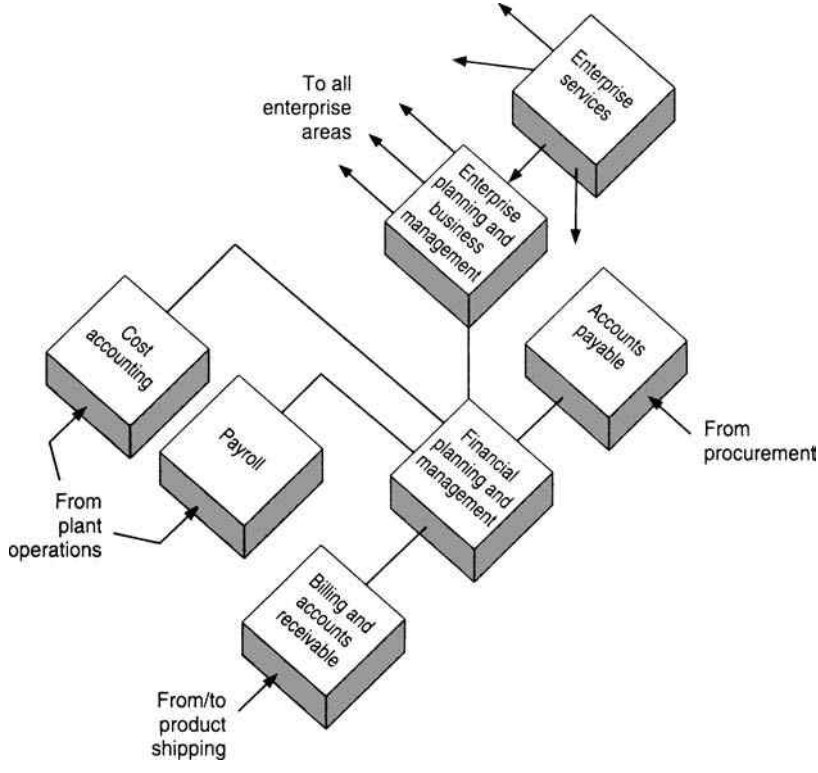


Figure 5.7 Business management.

5.3.6 Business Management

Within the enterprise model, a business management function may be composed of seven areas (Fig. 5.7).

5.3.6.1 Financial Planning and Management

In financial planning and management, financial resource plans are developed and enterprise goals are established. Among the functions are planning costs, budgets, revenues, expenses, cash flow, and investments.

Input includes financial goals and objectives established by management as well as summarized financial data received from all areas of the enterprise. Output includes financial reports to stockholders, departmental budgets, and general ledger accounting.

5.3.6.2 Accounts Payable

Accounts payable primarily involves paying vendors. Input includes vendor invoices and goods-received reports. The output includes discount calculations to vendors and payment checks. This last function

lends itself to the use of electronic data interchange to electronically transfer funds to vendors.

5.3.6.3 Billing and Accounts Receivable

The billing and accounts receivable area prepares invoices to customers and manages customer account collections. Input to this area consists of product shipping data and cash received. Output includes customer invoices, delinquent account reports, and credit ratings. Transferring funds electronically through EDI can simplify this area significantly.

5.3.6.4 Cost Accounting

Cost accounting activity supports product pricing and financial planning by establishing product costs. These costs can include those of materials, direct and indirect labor, fixed production elements (machinery and equipment), variable production elements (electricity, fuels, or chemicals), and overhead. Other functions can include establishing standard costs, reporting variances to standards, costing job orders, and determining accrued costs.

Cost accounting input is acquired primarily from plan operations. Output is sent to financial management and planning.

5.3.6.5 Payroll

The payroll area computes payments, taxes, and other deductions for employees. It also includes reporting and paying employee tax withholdings to government agencies. Input includes time and attendance information and production data from plant operations. Output includes payroll checks, labor distribution, and government reports and payments.

5.3.6.6 Enterprise Planning and Business Management

Enterprise planning and business management functions include establishing goals and strategies for marketing, finance, engineering and research, plant automation, sensors and control systems, and information systems. Input and output are exchanged through sensors and control systems with virtually every other area of the enterprise.

5.3.6.7 Enterprise Services

Enterprise services include such support services as office personnel, management information services, personnel resources, and public relations. These services require extensive administrative support tools, such as text processing, decision support, and graphic tools. But since input and output will be exchanged throughout the entire enterprise, it is imperative that these tools be integrated with the enterprise's other systems.

5.4 Design of CIM with Sensors and Control Systems

With the advent of low-priced computers and sensors and control systems, a number of technological developments have arisen related to manufacturing that can be used to make production more efficient and competitive. The primary purpose is to develop several computer-concepts as related to the overall manufacturing plan of CIM systems.

In order for the manufacturing enterprise to succeed in the future, it is imperative it adopt a manufacturing strategy that integrates its various functions. CIM systems have the potential to accomplish this task. The implementation of CIM with sensory and control systems on the shop floor represents a formidable, albeit obtainable, objective. To accomplish this goal, enterprises must have access to information on what is available in CIM. A secondary purpose of obtaining access to information is to provide a framework that can aid in the search for information. Once the information is obtained, it becomes necessary to look at the current system objectively and decide how to approach the problem of implementing CIM with sensors and control systems.

While many of the ideas associated with CIM are new and untried, progressive enterprises, with the realization that old methods are ineffective, are doing their part in implementing this new technology. Some of the concepts currently being implemented are flexible manufacturing systems, decision support systems (DSS), artificial intelligence (AI), just-in-time (JIT) inventory management, and group technology. While all of these concepts are intended to improve efficiency, each one alone can only accomplish so much. For example, an FMS may reduce work-in-process (WIP) inventory while little is accomplished in the area of decision support systems and artificial intelligence to relate all aspects of manufacturing management and technology to each other for FMS. The advent of inexpensive sensors and control systems enables the concept of CIM to be implemented with greater confidence.

Recent advances in computer technology and sensors in terms of speed, memory, and physical space have enabled small, powerful, personal computers to revolutionize the manufacturing sector and become an essential part of design, engineering, and manufacturing, through, for example, database management systems (DBMSs) and local area networks (LANs). The coordination of the various aspects of a manufacturing environment means that complex systems inherently interact with one another. Due to a lack of standards and poor communication between departments, many components and databases are currently incompatible.

Table 5.1 describes some benefits of CIM. The potential for CIM, according to Table 5.1, is overwhelming, but the main issue is how to

Application	Improvement with CIM, %
Manufacturing productivity	120
Product quality	140
Lead time (design to sale)	60
Lead time (order to shipment)	45
Increase in capital equipment utilization	340
Decrease in WIP inventory	75

Table 5.1 CIM Benefits

analyze and design CIM that incorporates sensors, control systems, and decision support so it is utilized effectively.

Manufacturing problems are inherently multiobjective. For example, improving quality usually increases cost and/or reduces productivity. Furthermore, one cannot maximize quality and productivity simultaneously; there is a tradeoff among these objectives. These conflicting objectives are treated differently by different levels and/or units of production and management. Obviously, without a clear understanding of objectives and their interdependence at different levels, one cannot successfully achieve CIM with sensors and control systems.

5.4.1 Components of CIM with Sensors and Control Systems

Decision making in designing CIM with effective sensors and control systems can be classified into three stages.

1. *Strategic level.* The strategic level concerns those decisions typically made by the chief executive officer (CEO) and the board of directors. Upper management decisions of this type are characterized by a relatively long planning horizon, lasting anywhere from one to ten years. Implementing CIM with sensors and control systems has to begin at this level. Even though small enterprises may not have as many resources at their disposal, they have the added advantage of fewer levels of management to work through while constructing CIM.
2. *Tactical level.* At the tactical level, decisions are made that specify how and when to perform particular manufacturing activities. The planning horizon for these decisions typically spans a period from 1 to 24 months. Activities at this level include such intermediate functions as purchasing and inventory control. They affect the amount of material on the shop floor but do not control the use of the material within the manufacturing process.

3. *Operational level.* Day-to-day tasks, such as scheduling, are performed at the operational level. The primary responsibility at this level is the effective utilization of the resources made available through the decisions made on the strategic and tactical levels. Because of the variability in demand or machine downtime, the planning horizon at this level must be relatively short, normally 1 to 15 days.

While each of these levels has certain responsibilities in a manufacturing plant, the objectives are often conflicting. This can be attributed to inherent differences between departments (e.g., sales and marketing may require a large variety of products to serve every customer's needs, while the production department finds its job easier if there is little product variation). One of the main causes of conflicting decisions is a lack of communication due to ineffective sensors and control systems between levels and departments. CIM with adequate sensors and control systems provides the ability to link together technological advances, eliminate much of the communication gap between levels, and bring all elements into a coherent production system.

5.4.2 CIM with Sensors and Control Systems at the Plant Level

Some of the important emerging concepts related to CIM with effective sensors and control systems are flexible manufacturing systems, material handling systems, automated storage and retrieval systems (AS/RS), computer-aided design (CAD), computer-aided engineering (CAE), computer-aided manufacturing (CAM), and microcomputers. These components of CIM can be classified into three major groups (Fig. 5.8).

5.4.2.1 Flexible Manufacturing Systems Incorporating Sensors and Control Systems

An FMS can link several elements on the shop floor through sensors in order to coordinate those elements. While CIM can be applied to any manufacturing industry, FMSs find their niche in the realm of discrete production systems such as job shops.

The most important FMS elements are numerical control machines and an automated material handling network to transport the product from raw material inventory, through the NC operations, and finally to the finished goods inventory.

Numerical control technology has made major advances with the advent of computer numerical control and direct numerical control (CNC/DNC). Microprocessors and sensors located on the machine itself can now provide the codes necessary for the parts to be operated on.

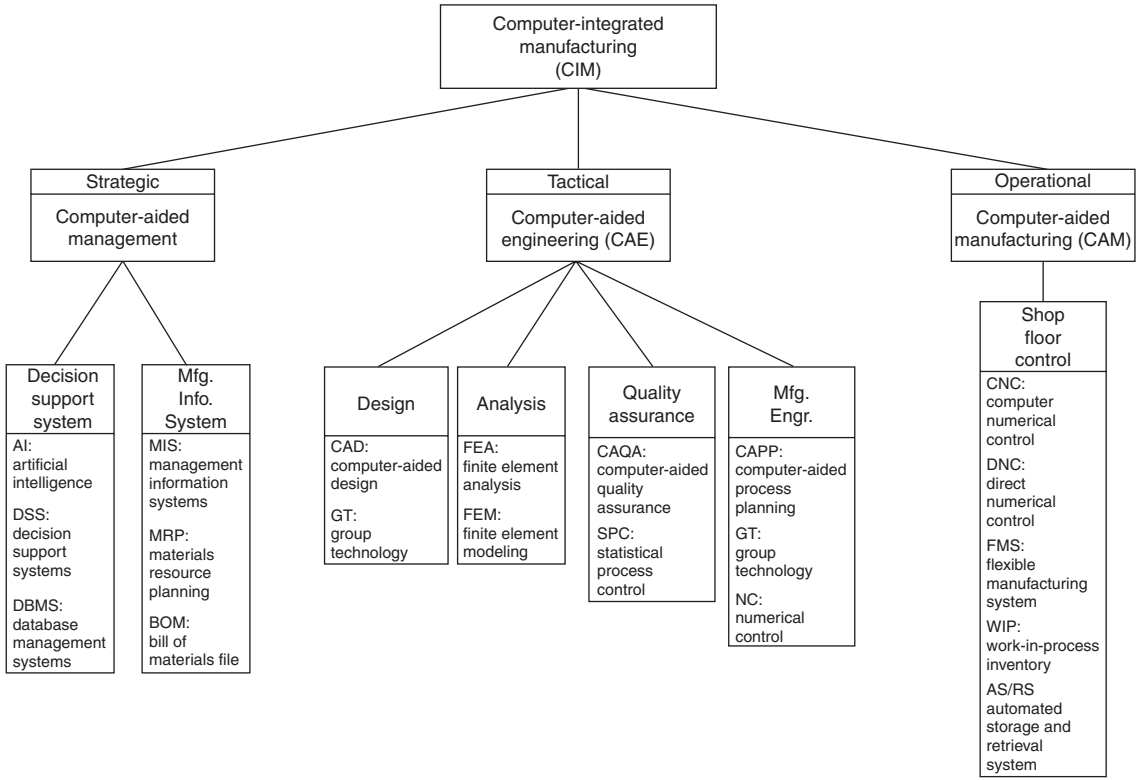


Figure 5.8 Components within CIM.

5.4.2.2 Material Handling

Material handling is the means of loading, unloading, and transporting workpieces among different machines and departments. It is accomplished in several ways:

- *Transfer lines* consist of fixed automation machinery such as conveyer belts. Their advantages are high speed and low cost. Their major disadvantage is their lack of flexibility. Dedicated transfer lines can handle only a limited number of parts and cannot be easily changed once in place, thus defeating the goals of an FMS.
- *Robots* provide another alternative for moving workpieces. Generally, robots can be made very flexible because of their programmability, but they are limited to their region of access.
- *Automated guided vehicles* can move workpieces a great distance, but they lack the speed found in both robot and transfer lines. Yet because of their ability to be programmed to different routes, they are more flexible than transfer lines.

5.4.2.3 Automated Storage and Retrieval Systems

By means of AGVs, raw materials can be taken from the loading dock and placed in a designated place in inventory. Using an AS/RS, inventory can be tracked throughout the manufacturing process and optimized for strategically locating items in storage. Because the process is computerized, data on what exactly is in inventory assist planners in determining order and production schedules.

Inventories consist of raw materials, work in process, and finished goods. Inventories should be controlled by keeping track of inventory locations and amounts. An AS/RS can accomplish this task.

5.4.2.4 Computer-Aided Engineering/Design/Manufacturing (CAE/CAD/CAM)

Computer-aided design helps the engineer in many ways during the design stage. Simply drawing the part on a computer increases the productivity of designers, but CAD is more than just automated drafting. It can facilitate group technology and the construction of a bill of materials (BOM) file.

Computer-aided engineering consists of the many computerized facets of engineering that go into a particular product. When a part has been designed, the CAE subgroup is responsible for generating the NC code that can be used by the NC machines on the floor.

By using GT, similar parts can be classified by similar attributes and placed in part families. By grouping parts this way, much redundancy in design and manufacturing is eliminated.

In *computer-aided manufacturing* (*CAM*), computers are used to plan and conduct production. They can correct and process data, control the manufacturing process, and provide information that can be used in decision making. CAM can involve distributed quality control and product testing and inspection, which are built into the manufacturing process to support the larger functional relationships.

5.4.2.5 Microcomputers for CIM

The integration of different elements of CIM and sensor systems can be accomplished only with the aid of a computer. Because of the magnitude of the CIM and sensor database, a mainframe (or a minicomputer for a small enterprise) is necessary for storage and retrieval of information. The power of microcomputers has increased so dramatically, however, that they are suitable for most other applications, such as:

- *Programmable logic controllers (PLCs).* On the factory floor, microcomputers are subject to harsh conditions. Dust, high temperature, and high humidity can quickly destroy an ordinary personal computer. Traditionally, the shop floor belongs to the programmable logic controller. The components can operate in extreme conditions (e.g., 55°C and 90 percent humidity). In addition, PLCs are geared to real-time control of factory operations. Through advances in microprocessor technology, functions once controlled by relays and mechanical switching mechanisms can now be performed by PLCs. The ladder diagrams used to notate logic circuits can now be programmed directly into the memory of the programmable controller. Because of microelectronic circuitry, PLCs can process control information quickly and shut down automatically in case of emergencies. Whereas PLCs have become commonplace on the floor of discrete product operations, process control computers have become indispensable in the control of process plants where conditions must be monitored constantly. They are also used in the control of office and factory environments. For example, a PLC can turn furnaces and air-conditioning units ON and OFF to provide suitable working conditions while optimizing energy use.
- *Industrial personal computers (PCs).* Until recently, PCs were commonly found in the protected environments of offices. Now, manufacturers have introduced rugged versions of popular personal computers. For example, IBM has developed the 5531 and 7531/32, industrialized versions of PC/XT and PC/AT, respectively, to withstand the environment of the factory floor. They have the advantage of being able to run any PC-DOS-based software, but are unable to perform real-time control. This problem has been remedied with the

advent of the industrial computer disk operating system (IC-DOS), a real-time operating system that is compatible with other IBM software. This allows a shop floor computer to provide real-time control while using software packages previously found only on office models.

- *Microsupercomputers.* The hardware of microsupercomputers has increased computing power significantly. Offering high performance, transportability, and low price, the new microsupercomputers compare favorably to mainframes for many applications.

5.5 Decision Support System for CIM with Sensors and Control Systems

With an increase in production volume and efficiency comes a need to have a more effective method of scheduling and controlling resources. Herein lies a connection between CAE and computer-aided management. The long-range plans of a company must include forecasts of what the demand will be for various products in the future. Through these forecasts, the enterprise determines what strategy it will take to ensure survival and growth.

For the enterprise to make intelligent decisions, reliable information must be available. In regard to the three levels of decision making, it is also important that the information be consistent throughout each level. The best way to assure this availability and consistency is to make the same database available to all individuals involved in the production process. Because of lack of good communication between levels, and sometimes the reluctance of upper-level managers to commit themselves to CIM, constructing a centralized database represents one of the most difficult problems in the implementation of CIM.

5.5.1 Computer-Integrated Manufacturing Database (CIM DB)

The creation of a CIM DB is at the heart of the effective functioning of CIM. Most manufacturers have separate databases set up for nearly every application. Since data from one segment of an enterprise may not be structured for access by other segments' software and hardware, a serious problem for meeting the CIM goal of having readily available data for all levels occurs. Another problem with multiple databases is in the redundancy of data. Both the strategic and tactical decision makers, for example, may need information concerning a bill of material file. Even with the assumption that the databases contain consistent data (i.e., the same information in each), maintaining them both represents inefficient use of computer time and storage

and labor. To install CIM, these databases must be consolidated. Unfortunately, bringing multiple databases into one CIM DB that remains available to everyone and consistent in all levels presents a significant obstacle because of the large investment needed in time and computer hardware and software.

5.5.2 Structure of Multiobjective Support Decision Systems

The success of CIM also depends largely on the ability to incorporate sensor technology with a database. The database is utilized in making decisions on all levels—decisions that are used to update to the database. Decision support systems can provide a framework for efficient database utilization by allowing storage and retrieval of information and problem solving through easy communications.

Decision-making problems in manufacturing can be grouped into two general classes:

- *Structured decisions* are those that are constrained by physical or practical limitations and can be made almost automatically with the correct input. An example is generating a group technology part code given the geometry of the part.
- *Unstructured decisions* typically are those that contain a large degree of uncertainty. Decisions considered by strategic planners are almost always unstructured. Deciding whether or not to expand a certain product line, for example, may be based on demand forecasting and on the expected growth of competitors. Due to the predictive nature of these decisions, they inherently contain more uncertainty than structured decisions. Long-range planning consists primarily of unstructured information.

Decision support systems mainly consist of three separate parts:

- *Language systems.* The function of a language system (LS) is to provide a means for the user to communicate with the DSS. Some considerations for the choice of a language are that the formulation should be easily understood, implementable, and modifiable. Moreover, processing the language should be possible on a separate level or on the problem processing system (PPS) level. An obvious choice for a language would be the spoken language of the user. This would require little or no training for the user to interact with a computer, but the complexity of sentences and the use of words that have multiple meanings present difficult problems that, when solved, would introduce unwanted inefficiency into the language system. An alternative would be to use a more formal language based on logic (e.g., PROLOG). The advantage here is

that a language can be used at all levels of the DSS. In the design and use of an LS for the user interface, one can consider objectives such as ease of communication, the level of complexity that can be presented by the LS, and the time needed for the user to learn it.

- *Knowledge systems.* The basic function of a knowledge system (KS) is the representation and organization of the "knowledge" in the system. Two possible approaches are storing in the information in database form or representing the data as a base for artificial intelligence using methods from, for example, predicate calculus. The objective of KS is to ease accessibility of data for the DSS. The KS should be able to organize and classify databases and problem domains according to objectives that are sensible and convenient for the user. Some of the objectives in the design of the KS are to reduce the amount of computer memory required, increase the speed with which the data can be retrieved or stored, and increase the number of classifications of data and problem domains possible.
- *Problem processing systems.* The problem processing system of a DSS provides an interface between the LS and the KS. The primary function is to receive the problem from the user via the LS and use the knowledge and data from the KS to determine a solution. Once a solution is found, the PPS sends it through the KS to be translated into a form the user can recognize. More importantly, in the model formulation, analysis, and solution procedure of PPS, the conflicting objectives of stated problems must be considered. The PPS should provide methodology that can optimize all conflicting objectives and generate a compromise solution acceptable to the user. Some of the objectives in the development of such multiobjective approaches are to reduce the amount of time that the user must spend to solve the problem, increase the degree of interactiveness (e.g., how many questions the user should answer), reduce the difficulty of questions posed to the user, and increase the robustness of the underlying assumptions and procedures.

5.6 Analysis and Design of CIM with Sensors and Control Systems

Many manufacturing systems are complex, and finding a place to begin a system description is often difficult. Breaking down each function of the system into its lowest possible level and specifying objectives for each level and their interactions will be an effective

step. The objectives and decision variables, as related to elements or units for all possible levels, are outlined in the following sections.

5.6.1 Structured Analysis and Design Technique (SADT)

The structured analysis and design technique is a structured methodology. It combines the graphical diagramming language of structured analysis (SA) with the formal thought discipline of a design technique (DT). The advantage of SADT is that it contains a formalized notation and procedure for defining system functions.

Psychological studies have shown that the human mind has difficulty grasping more than five to seven concepts at one time. Based on this observation, SADT follows the structured analysis maxim: "Everything worth saying must be expressed in six or fewer pieces." Limiting each part to six elements ensures that individual parts are not too difficult to understand. Even complex manufacturing systems can be subjected to top-down decomposition without becoming overwhelming.

The basic unit for top-down decomposition is the structural analysis box (Fig. 5.9). Each of the four sides represents a specific action for the SA box. The four actions implement:

- *Input*, measured in terms of different decision variables.
- *Control*, to represent constraints and limitations.
- *Output*, measured in the form of a set of objectives.
- *Mechanism for translation*, which performs the translations (for mapping) of input into output as constrained by the control action.

Since each box represents an idea in the system, each box contains a detailed diagram. A parent-child relationship exists between the

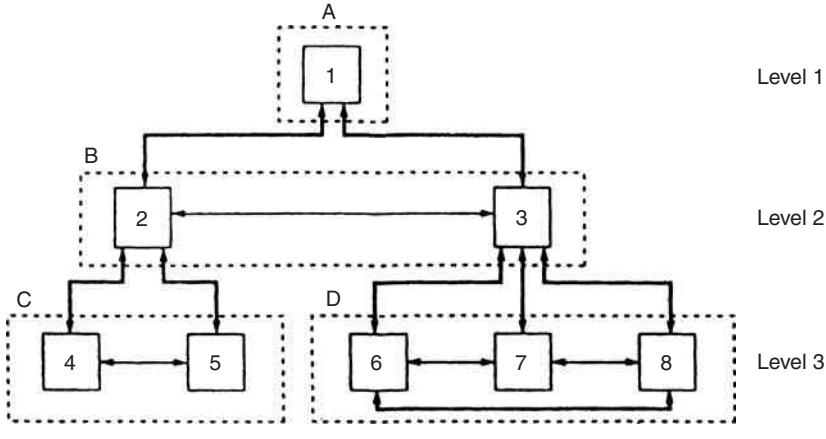


Figure 5.9 Structural analysis box.

box being detailed and those boxes under it. The same relationship holds for the parent diagram as the diagram directly preceding the child diagrams.

The interfaces linking boxes through their neighbors at the same level are the input, output, and control actions. In graphical terms, these interfaces are designated by arrows. Methodically proceeding through a given network, the entire system can be modeled in terms of boxes and arrows.

While SADT provides a realistic approach to modeling any system, it cannot provide the solution to any problem. The integrated computer-aided manufacturing definition method comes one step closer to the realization of a functional CIM system. It utilizes teamwork and demands that all correspondence and analysis be in written form so others can obtain a grasp of the situation and errors can be more readily located. Because the written word is required during the implementation phases, the documentation that usually is done at the end of most major projects can be nearly eliminated. Keeping accurate records also plays an important role in debugging the system in the future.

5.6.2 A Multiobjective Approach for Selection of Sensors in Manufacturing

Sensors are evaluated using six criteria:

- *Cost* is simply the price for the sensor and its integrated circuitry if it should be purchased.
- *Integrability* is the degree to which the sensor can be used in conjunction with the manufacturing system it serves. This can usually be measured in terms of compatibility with existing hardware control circuits and software.
- *Reliability* is the quality of the sensors as indicated by the *mean time between failures* (MTBF), and can be measured by performing a simple stress test on the sensor under severe limits of operation. If the sensor operates under a certain high temperature for a certain period of time, it will assure the user that the system will perform satisfactorily under normal operating conditions. It will also indicate that the electronic control circuits are reliable, according to the *burn-in philosophy*.
- *Maintenance* involves the total cost to update and maintain the sensor and how often the sensor needs to be serviced.
- *Expandability* is how readily the sensor can be modified or expanded as new needs arise because of a changing environment.
- *User friendliness* indicates the ease of using and understanding the unit. It may include the quality of documentation in terms of simplicity, completeness, and step-by-step descriptions of procedures.

5.7 Data Acquisition for Sensors and Control Systems in CIM Environments

The input signals generated by sensors can be fed into an interface board, called an *I/O board*. This board can be placed inside a PC-based system. As personal computers for CIM have become more affordable, and I/O boards have become increasingly reliable and readily available, PC-based CIM data acquisition has been widely implemented in laboratory automation, industrial monitoring and control, and automatic test and measurement.

To create a data acquisition system for sensors and control systems that really meets the engineering requirements, some knowledge of electrical and computer engineering is required. The following key areas are fundamental in understanding the concept of data acquisition for sensors and control systems:

- Real-world phenomena
- Sensors and actuators
- Signal conditioning
- Data acquisition for sensors and control hardware
- Computer systems
- Communication interfaces
- Software

5.7.1 Real-World Phenomena

Data acquisition and process control systems measure real-world phenomena, such as temperature, pressure, and flow rate. These phenomena are sensed by sensors, and are then converted into analog signals that are eventually sent to the computer as digital signals.

Some real-world events, such as contact monitoring and event counting, can be detected and transmitted as digital signals directly. The computer then records and analyzes this digital data to interpret real-world phenomena as useful information.

The real world can also be controlled by devices or equipment operated by analog or digital signals generated by the computer (Fig. 5.10).

5.7.2 Sensors and Actuators

A sensor converts a physical phenomenon—such as temperature, pressure, level, length, position, or presence or absence—into a voltage, current, frequency, pulses, and so on.

For temperature measurements, some of the most common sensors include thermocouples, thermistors, and resistance temperature detectors (RTDs). Other types of sensors include flow sensors, pressure sensors, strain gauges, load cells, and optical sensors.

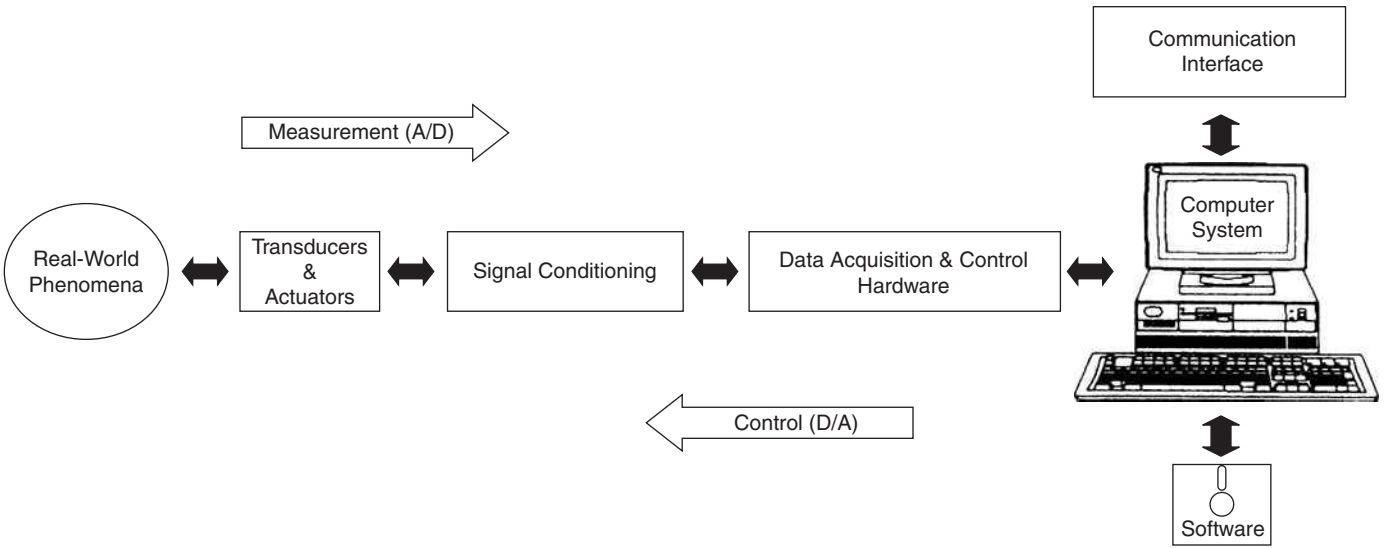


Figure 5.10 Integration of computer-controlled devices.

An actuator is a device that activates process control equipment by using pneumatic, hydraulic, electromechanical, or electronic signals. For example, a valve actuator is used to control fluid rate for opening and closing a valve.

5.7.3 Signal Conditioning

A signal conditioner is a circuit module specifically intended to provide signal scaling, amplification, linearization, cold junction compensation, filtering, attenuation, excitation, common mode rejection, and so on. Signal conditioning improves the quality of the sensor signals that will be converted into digital signals by the PC's data acquisition hardware.

One of the most common functions of signal conditioning is amplification. Amplifying a sensor signal provides an analog-to-digital (A/D) converter with a much stronger signal and thereby increases resolution. To acquire the highest resolution during A/D conversion, the amplified signal should be equal to approximately the maximum input range of the A/D converter.

5.7.4 Data Acquisition for Sensors and Control Hardware

In general, data acquisition for sensors and control hardware performs one or more of the following functions:

- Analog input
- Analog output
- Digital input
- Digital output
- Counter/timer

5.7.4.1 Analog Input

An analog-to-digital (D/A) converter produces digital output directly proportional to an analog signal input so it can be digitally read by the computer. This conversion is imperative for CIM (Fig. 5.11).

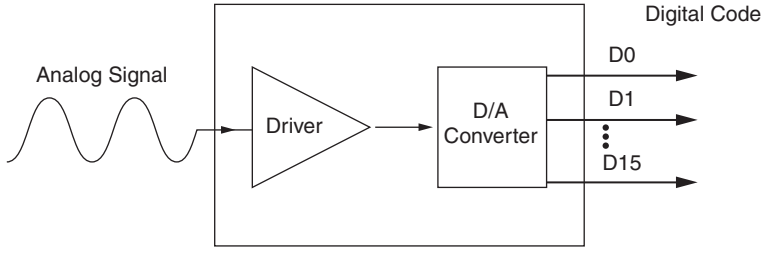


Figure 5.11 Analog-to-digital converter.

The most significant aspects of selecting A/D hardware are:

- Number of input channels
- Single-ended or differential input
- Sampling rate (in samples per second)
- Resolution (in bits)
- Input range (specified as full-scale volts)
- Noise and nonlinearity

5.7.4.2 Analog Output

A digital-to-analog (D/A) converter changes digital information into a corresponding analog voltage or current. This conversion allows the computer to control real-world events.

Analog output may directly control equipment in a process that is then measured as an analog input. It is possible to perform a closed loop or proportional integral-differential (PID) control with this function. Analog output can also generate waveforms in a function generator (Fig. 5.12).

5.7.4.3 Digital Input and Output

Digital input and output are useful in many applications, such as contact closure and switch status monitoring, industrial ON/OFF control, and digital communication (Fig. 5.13).

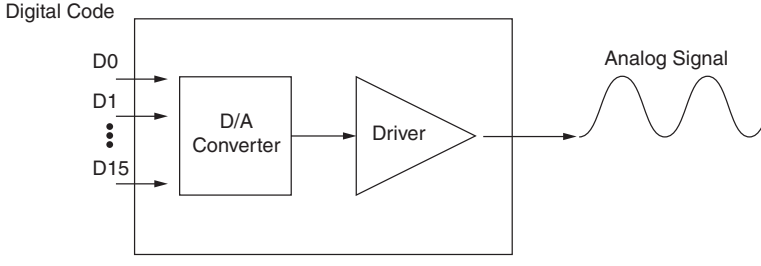


Figure 5.12 Analog output.

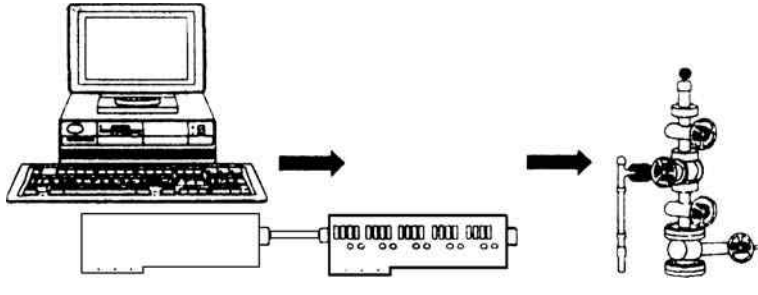


Figure 5.13 Application of digital input/output.

5.7.4.4 Counter/Timer

A counter/timer can be used to perform event counting, flow meter monitoring, frequency counting, pulse width and time period measurement, and so on.

Most data acquisition and control hardware is designed with the multiplicity of functions described earlier on a single card for maximum performance and flexibility. Multifunction data acquisition for high-performance hardware can be obtained through PC boards specially designed by various manufacturers for data acquisition systems.

5.7.5 Computer System

Today's rapidly growing PC market offers a great selection of PC hardware and software in a wide price range. Thus, a CIM strategy can be economically implemented.

5.7.5.1 Hardware Considerations

Different applications require different system performance levels. Currently, 286, 386, and 486 CPUs will allow a PC to run at benchmark speeds from 20 up to 150 MHz. Measurements and process control applications usually require 80286 systems. However, for applications that require high-speed real-time data analysis, an 80386 or 80486 system is much more suitable.

5.7.5.2 Industrial PCs

An *industrial PC* (*IPC*) is designed specifically to protect the system hardware in harsh operating environments. IPCs have rugged chassis that protect system hardware against excessive heat, dust, moisture, shock, and vibration. Some IPCs are even equipped with power supplies that can withstand temperatures from –20 to +85°C for added reliability in harsh environments.

5.7.5.3 Passive Backplane and CPU Card

More and more industrial data acquisition for sensors and control systems are using passive backplane and CPU card configurations. The advantages of these configurations are reduced mean time to repair (MTTR), ease of upgrading the system, and increased PC-bus expansion slot capacity.

A passive backplane allows the user to plug in and unplug a CPU card without the effort of removing an entire motherboard in case of damage or repair.

5.7.6 Communication Interfaces

The most common types of communication interfaces used in PC-based data acquisition for sensor and control system applications are RS-232, RS-422/485, and the IEEE-488 general-purpose interface bus (GPIB).

The RS-232 interface is the most widely used interface in data acquisition for sensors and control systems. However, it is not always suitable for distances longer than 50 m or for multidrop network interfaces. The RS-422 protocol has been designed for long distances (up to 1200 m) and high-speed (usually up to 56,000 bits/s) serial data communication. The RS-485 interface can support multidrop data communication networks.

5.7.7 Software

The driving force behind any data acquisition for sensors and control systems is its software control. Programming the data acquisition for sensors and control systems can be accomplished using one of three methods:

- *Hardware-level programming* is used to directly program the data acquisition hardware's data registers. In order to achieve this, the control code values must determine what will be written to the hardware's registers. This requires that the programmer use a language that can write or read data from the data acquisition hardware connected to the PC. Hardware-level programming is complex, and requires significant time—time that might be prohibitive to spend. This is the reason most manufacturers of data acquisition hardware supply their customers with either driver-level or package-level programs.
- *Driver-level programming* uses function calls with popular programming languages such as C, PASCAL, and BASIC, thereby simplifying data register programming.
- *Package-level programming* is the most convenient technique of programming the entire data acquisition system. It integrates data analysis, presentation, and instrument control capabilities into a single software package. These programs offer a multitude of features, such as pull-down menus and icons, data logging and analysis, and real-time graphic displays.

5.8 Developing CIM Strategy with Emphasis on Sensors' Role in Manufacturing

To develop a comprehensive CIM strategy incorporating sensors and control systems, an enterprise must begin with a solid foundation, such as a CIM architecture. A CIM architecture is an information system structure that enables the industrial enterprise to integrate information and business processes. It accomplishes this by (1) establishing the direction integration will take and (2) defining the interfaces between the users and the providers of this integration function.

Figure 5.14 shows how CIM architecture answers the enterprise's integration means. A CIM architecture provides a core of common

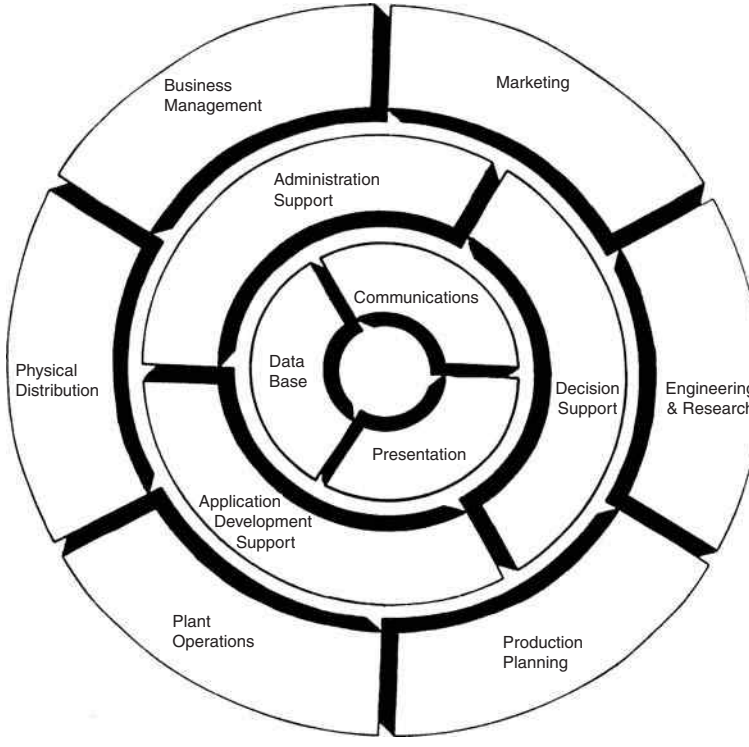


FIGURE 5.14 CIM architecture.

services. These services support every other area in the enterprise, from its common support function to its highly specialized business processes.

5.8.1 CIM and Building Blocks

The information environment of an industrial enterprise is subject to frequent changes in system configuration and technologies. A CIM architecture incorporating sensors and control systems can offer a flexible structure that enables it to react to the changes. This structure relies on a number of modular elements that allow systems to change easily so as to grow with the enterprise's needs.

Figure 5.17 shows a modular structure that gives CIM flexibility. It is based on three key building blocks:

- *Communications.* The communication and distribution of data
- *Data management.* The definition, storage, and use of data
- *Presentation.* The presentation of data to people and devices throughout the enterprise

Utilizing the building blocks, CIM can provide a base for integrating the enterprise's products, processes, and business data. It can define the structure of the hardware, software, and services required to support the enterprise's complex requirements. It can also translate information into a form that can be used by the enterprise's people, devices, and applications.

5.8.2 CIM Communications

Communications—the delivery of enterprise data to people, systems, and devices—is a critical aspect of CIM architecture. This is because today's industrial environment brings together a wide range of computer systems, data acquisition systems, technologies, system architectures, operating systems, and applications. This range makes it increasingly difficult for people and machines to communicate with each other, especially when they describe and format data differently.

Various enterprises, in particular IBM, have long recognized the need to communicate data across multiple environments. IBM's response was to develop *systems network architecture* (SNA) in the 1970s. SNA supports communication between different IBM systems, and over the years it has become the standard for host communications in many industrial companies.

However, the CIM environment with sensor communications must be even more integrated. It must expand beyond individual areas, throughout the entire enterprise, and beyond—to customers, vendors, and subcontractors.

Communications in the CIM environment involves a wide range of data transfer, from large batches of engineering or planning data to single-bit messages from a plant floor device. Many connectivity types and protocols must be supported so that the enterprise's people, systems, and devices can communicate. This is especially true in cases where response time is critical, such as during process alerts.

5.8.3 Plant Floor Communications

Plant floor communications can be the most challenging aspect of factory control. This is due to the wide range of manufacturing and computer equipment that have been used in production tasks over the decades.

IBM's solution for communicating across the systems is the IBM plant floor series, a set of software products. One of these products, *Distributed Automation Edition* (DAE), is a systems enabler designed to provide communication functions that can be utilized by plant floor applications. These functions include:

- Defining and managing networks
- Making logical device assignments

- Managing a program library for queuing and routing messages
- Establishing alert procedures
- Monitoring work-cell status

With these functions, Distributed Automation Edition can (1) help manufacturing engineers select or develop application programs to control work-cell operations and (2) provide communication capabilities between area- and plant-level systems (Fig. 5.15).

DAE supports several communication protocols to meet the needs of a variety of enterprises. For example, it supports SNA for connections to plant host systems and the IBM PC network, as well as

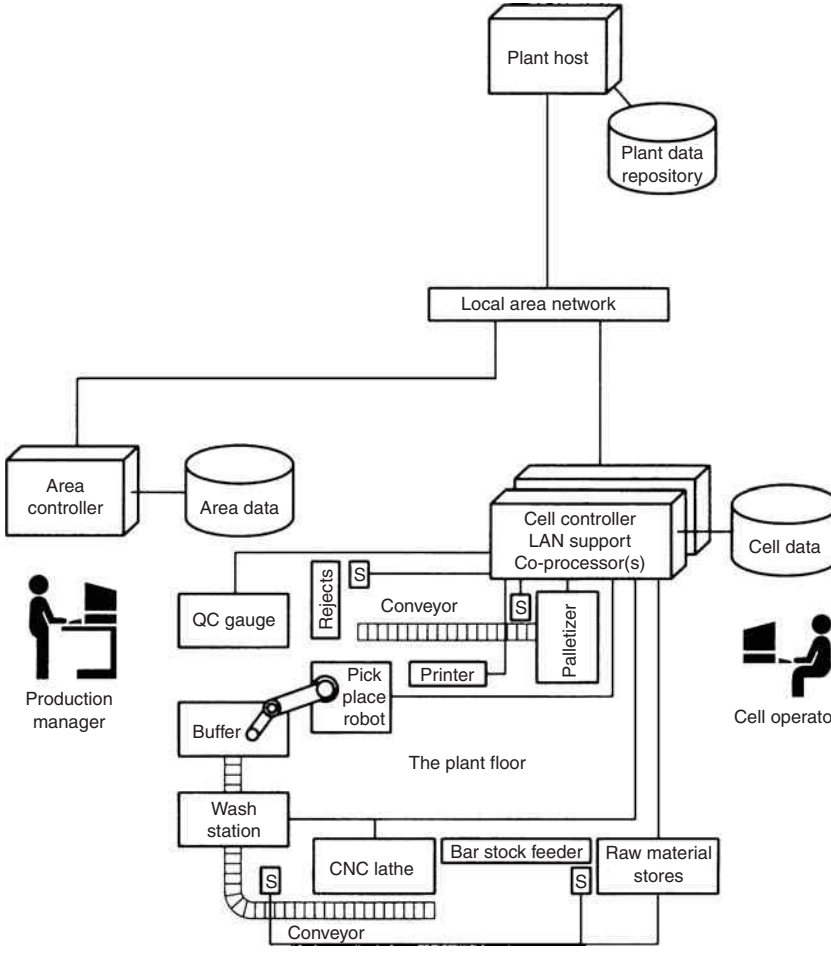


Figure 5.15 Distributed Automation Edition (DAE).

the IBM token-ring protocol and manufacturing automated protocol (MAP) for plant floor communications. MAP is the evolving plant floor communication industry standard, adopted by the International Standards Organization for communications among systems provided by different vendors.

5.8.4 Managing Data in the CIM Environment

The second building block of a CIM architecture incorporating sensors and control technology is data management. This includes how data are defined, how different data elements are related, where data are stored, and who has access to that data. Data management is particularly critical in today's industrial environment since many different databases, formats, and storage and access techniques are available.

Standards are evolving. For example, *Structured Query Language* (SQL) provides a medium for *relational database* applications and for users to access a database. Unfortunately, a significant amount of data exists today in other database technologies that are not accessible by current standards.

Data management defines and records the location of the data created and used by the enterprise's business functions. Data management also means enabling users to obtain the data needed without having to know where the data are located.

Relationships among several data elements must be known if data are to be shared by users and applications. In addition, other data attributes are important in sharing data. These include the type of data (text, graphic, image), their stat (working, review, completed), and their source (person, application, or machine).

In CIM with sensory architecture, data management can be accomplished through three individual storage functions: (1) the data repository, (2) the enterprise data storage, and (3) the local data files.

Some of the key data management functions—the repository, for example—are already being implemented by the *consolidated design file* (CDF) established through the IBM *Data Communication Service* (DCS).

The consolidated design file operates on a relational database and is built on SQL. One example of its use is as an engineering database to integrate CAD/CAM applications with the business needs of the engineering management function. This environment, IBM's DCS/CDF, provides the following repository functions:

- Transforming data to a user-selected format
- Storing CAD/CAM data
- Adding attributes to CAD/CAM data
- Enabling users to query data and attributes

DCS/CDF also provides communications functions to transfer data between the repository and CAD/CAM applications (Fig. 5.16).

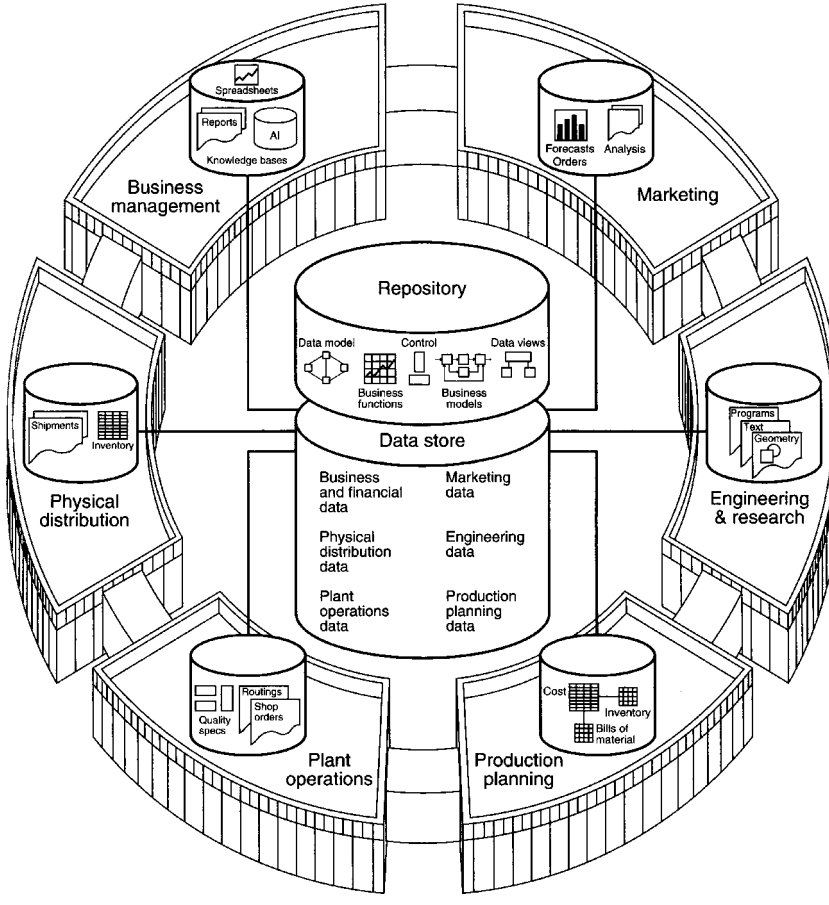


FIGURE 5.16 Data Communication Service/consolidated design file (DCS/CDF).

5.8.5 CIM Environment Presentation

Presentation in the CIM environment means providing data to, and accepting data from, people and devices. Obviously, this data must assume appropriate data definitions and screen formats to be usable.

Because today's industrial enterprise contains such a wide array of devices and information needs, it must have a consistent way to distribute and present information to people, terminals, workstations, machine tools, robots, sensors, bar-code readers, automated guided vehicles, and part storage and retrieval systems. The range of this information covers everything from simple messages between people to large data arrays for engineering design and applications (Fig. 5.17). It may originate from a CIM user in one functional area of the enterprise and be delivered to a CIM user or device in another area.

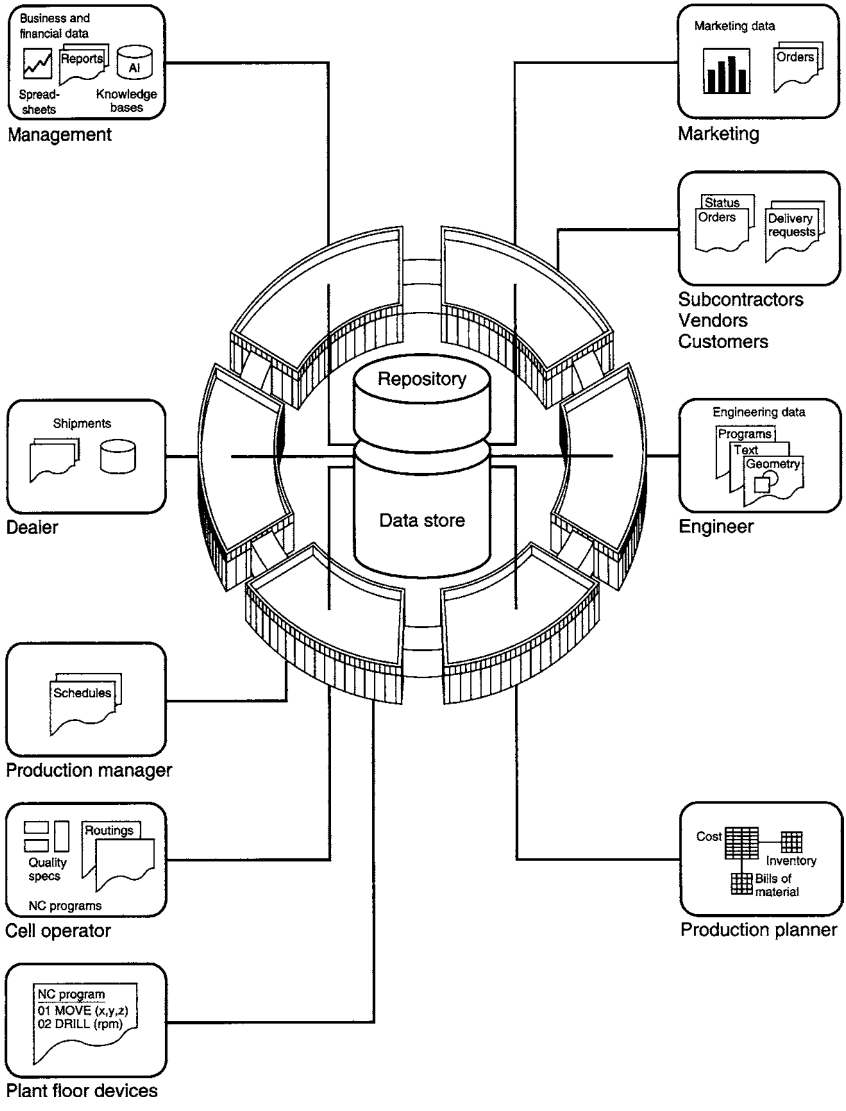


Figure 5.17 Presentation of data.

In today's environments, presentation occurs on displays that utilize various technologies. Some are nonprogrammable terminals, some are programmable workstations, and some are uniquely implemented for each application. As a result, the same information is often treated differently by individual applications.

For example, the same manufactured part may be referred to as a part number in a bill of material in production planning, as a drawing

in engineering's CAD application, and as a routing in a paperless shop order from plant operations.

As data are shared across the enterprise, they must be transformed into definitions and formats that support the need of individual users and applications. Applications must be able to access shared data, collect the required information, then format that information for delivery.

5.8.6 The Requirement for Integration

Communication, data management, and presentation each have their own set of technical requirements. In addition, before these three building blocks can be integrated, a CIM architecture must also address a number of enterprise-wide constraints. For example, a CIM architecture should be able to:

- Utilize standard platforms
- Integrate data
- Protect the installed base investment
- Work with heterogeneous systems
- Utilize industry-standard operator interfaces
- Reduce application support cost
- Provide a customizable solution
- Offer phased implementation
- Deliver selectable functions
- Improve the business process

5.8.6.1 Standard Platforms

Utilizing standard computing platforms is one step industrial enterprises can take toward integration.

Today, many products are available that utilize standard platforms. These include processors, operating systems, and enablers for communications, database management, and presentation. In addition, platforms such as IBM's *systems application architecture* (SAA) and *advance interactive executive* (AIX) help make application consistency a reality across strategic IBM and UNIX operating system environments.

SAA, for example, is a comprehensive IBM blueprint for consistency and compatibility of software products. SAA begins the process by establishing definitions for four key application aspects: common user access, common programming interface, common communication support, and common applications. Through these definitions, SAA will support the development of new applications across major operating environments.

AIX is IBM's version of the UNIX operating system and combines consistent user and application interfaces to aid in the development of an integrated application across UNIX environments. AIX consists of six related system enablers:

- Base systems
- Programming
- Interface
- User interface
- Communication support
- Distributed processing and applications

5.8.6.2 Data Integration

Integration requirements are often met by creating bridges between individual applications. Bridges usually copy a collection of data between two applications. A bridge between engineering and production planning allows these two functions to share a bill of material. Another bridge permits an engineering CAD/CAM application to download an NC program to a plant floor personal computer. Or a bridge between production planning and plant operations may be used to provide a copy of the production schedule to the plant floor system.

However, a problem with bridges is that changes made to the original set of data are not immediately incorporated into the copy of the data. This results in out-of-date information. Another problem is that bridges become difficult to maintain when more than two applications must work together.

As enterprises begin to integrate their operations, it will be imperative that the latest information is shared among multiple applications and across business functions. For example, engineering, marketing, cost accounting, production planning, and plant operations may all need access to inventory status information. At other times, the enterprise's various business functions may need information about product specifications, order status, operating cost, and more.

A CIM architecture must be able to simplify and accelerate this integration. It must provide the facilities to integrate data across the various applications of the business functions—facilities such as data query, data communication, controlled access and editing, and consistent data definitions.

5.8.6.3 Installed Base Investment

Today's industrial enterprises have made considerable investments in their installed bases, including systems, data, and even training. In the United States alone, manufacturers spend billions of dollars per annum on information systems hardware, software, and integration

services for production planning, engineering, and plant operations. CIM with sensory technology must help protect this investment by permitting the integration of existing systems, applications, and data.

5.8.6.4 Heterogeneous Systems

In today's heterogeneous environment, data are located on different systems and in different formats. Applications have different needs, which are answered by processors, communications, and displays utilizing different technologies and architectures.

In an enterprise model, production planning may automate its operations on a single mainframe using an interactive database. Engineering may store drawings in an engineering database, then design and analyze products on a network of graphics workstations. Plant operations and sensors and control systems may be automated with personal computers and specialized machine controllers connected by both standard and proprietary networks. The data needed to operate the enterprise are scattered across all these diverse systems.

The heterogeneous environment is also characterized by an installed system base provided by multiple computer system suppliers, software vendors, and systems integrators. A CIM architecture must allow the integration of these varied system solutions and operating platforms.

5.8.6.5 Industry Standards and Open Interfaces

As integration technologies mature, there will be the need to support an expanding set of industry standards. Today, these standards include communication protocols such as MAP, token ring, and Ethernet; data exchange formats like *initial graphics exchange specifications* (IGES) for engineering drawings; data access methods such as SQL; and programming interfaces like *programmer's hierarchical interactive graphics standard* (PHIGS). A CIM architecture must be able to accommodate these and other evolving standards. One framework for accomplishing this has already been established, the open systems architecture for CIM (CIM-OSA). CIM-OSA is being defined in the *Esprit* program by a consortium of European manufacturers, universities, and information system suppliers, including IBM. Data exchange formats are also being extended to accommodate product definition in the *product definition exchange specification* (PDES). In addition, a CIM architecture must be able to support well-established solutions, such as IBM's SNA, which have become de facto standards.

In this competitive marketplace, manufacturers must also be able to extend operations as needed and support these new technologies and standards as they become available. These needs may include adding storage to mainframe systems, replacing engineering workstations, installing new machine tools, upgrading operating systems,

and utilizing new software development tools. A CIM architecture with open interfaces will allow enterprises to extend the integration implementation over time to meet changing business needs.

5.8.6.6 Reduced Application Support Cost

A CIM architecture incorporating sensors must also yield application solutions at a lower cost than traditional stand-alone computing. This includes reducing the time and labor required to develop integrated applications and data. It also means reducing the time and effort required to keep applications up to speed with the changes in the enterprise's systems environment, technology, and integration needs.

5.8.6.7 Customizable Solutions

Every enterprise has its own business objectives, shared data, system resources and applications. For example, one enterprise may choose to address CIM requirements by reducing the cycle time of product development. It does this by focusing on the data shared between the product development and process development functions in accelerating the product release process.

Another enterprise's aim is to reduce the cycle time required to deliver an order to a customer. It addresses this by exchanging data between production planning and operations functions and automating the order servicing and production processes. It must also be able to customize operations to individual and changing needs over time.

5.8.6.8 Phased Implementation

Implementing enterprise-wide integration will take place in many small steps instead of through a single installation. This is because integration technology is still evolving, implementation priorities are different, installed bases mature at different rates, new users must be trained, and lessons will be learned in pilot installations.

As enterprises begin their implementation efforts in phases, they will be integrating past, present, and future systems and applications. A CIM architecture must be able to support the integration of this diverse installed base.

5.8.6.9 Selectable Functions

Most enterprises will want to weigh the benefits of integration against the impact this change will bring to each application and set of users. For example, an emphasis on product quality may require that production management gain greater insight into the quality of individual plant operations activities by implementing advanced sensors and control systems developed for particular applications. When an

existing shop floor control application adequately manages schedules and support shop personnel with operating instructions, some additional information, such as that on quality, may be added, but rewriting the existing application may not be justified.

However, the plant manager may plan to develop a new production monitoring application to operate at each workstation. This application will make use of various sensors, share data with shop floor control, and utilize software building blocks for communications, database management, and presentation.

As is evident, the existing application requires only a data sharing capability, while the new application can benefit from both data sharing and the architecture building blocks. A CIM architecture with selectable functions will provide more options that can support the variety of needs within an enterprise.

5.8.6.10 Improved Business Process

Obviously, an enterprise will not implement integration on the basis of its technical merits alone. A CIM architecture must provide the necessary business benefits to justify change and investment.

The integration of information systems must support interaction between business functions and the automation of business processes. This is a key function if corporate goals, such as improved responsiveness to customer demands and reduced operating cost, are to be met. A CIM architecture must provide the means by which an entire enterprise can reduce the cycle times of business processes required for order processing, custom offerings, and new products. It must also reduce the impact of changes in business objectives and those business processes.

Further Reading

Bucker, D. W., "10 Principles to JIT Advancement," *Manufacturing Systems*, 55, March 1988.

Campbell, J., *The RS-232 Solution*, Sybex, Inc., Alameda, Ca., 1984.

Clark, K. E., "Cell Control, The Missing Link to Factory Integration," International Industry Conference Proceedings, Toronto, 641–646, May 1989.

Datapro Research Corporation, "How U.S. Manufacturing Can Thrive," *Management and Planning Industry Briefs*, 39–51, March 1987.

Groover, M. P., and E. W. Zimmer, Jr., *CAD/CAM: Computer Aided Design and Manufacturing*, Prentice-Hall, Englewood Cliffs, N.J., 1984.

IBM Corp., *Introducing Advanced Manufacturing Applications*, IBM, Atlanta, Ga., 1985.

"APICS: Tweaks and Distributed Systems Are the Main Focus—MRP No Longer Looks to the Future for Finite Capacity Scheduling," *Managing Automation*, 31–36, January 1992.

Manufacturing Studies Board, National Research Council, *Toward a New Era in U.S. Manufacturing: The Need for a National Vision*, National Academy Press, Washington, D.C., 1986.

Orlicky, J., *Material Requirements Planning*, McGraw-Hill, New York, 1985.

Schonberger, R. J., "Frugal Manufacturing," *Harvard Business Review*, 95–100, September–October 1987.
Skinner, W., *Manufacturing: The Formidable Competitive Weapon*, Wiley, New York, 1985.
"Support for The Manufacturing Floor," *Manufacturing Engineering*, 29–30, March 1989.
Vollmann, T. E., W. L. Berry, and D. C. Whyback, *Manufacturing Planning and Control Systems*, Richard D. Irwin, Homewood, Ill., 1984.

CHAPTER 6

Advanced Sensor Technology in Precision Manufacturing Applications

6.1 Identification of Manufactured Components

In an automated manufacturing operation, one should be able to monitor the identification of moving parts. The most common means of automatic identification is *bar-code technology*, but other approaches offer advantages under certain conditions.

6.1.1 Bar-Code Identification Systems

The *universal product code* (UPC) used in retail stores is a standard 12-digit code. Five of the digits represent the manufacturer and five the item being scanned. The first digit identifies the type of number system being decoded (a standard supermarket item, for example) and the second is a parity digit to determine the correctness of the reading. The first six digits are represented by code in an alternating pattern of light and dark bars. Figure 6.1 shows two encodings of the binary string 100111000. In both cases, the minimum printed width is the same. The delta code requires nine such widths (the number of bits), while the width code requires 13 such widths (if a wide element is twice the width of the narrow element). Different bar widths allow for many character combinations. The remaining six digits are formed by dark alternating with light bars reversing the sequence of the first six digits. This allows backward scanning detection (Fig. 6.2).

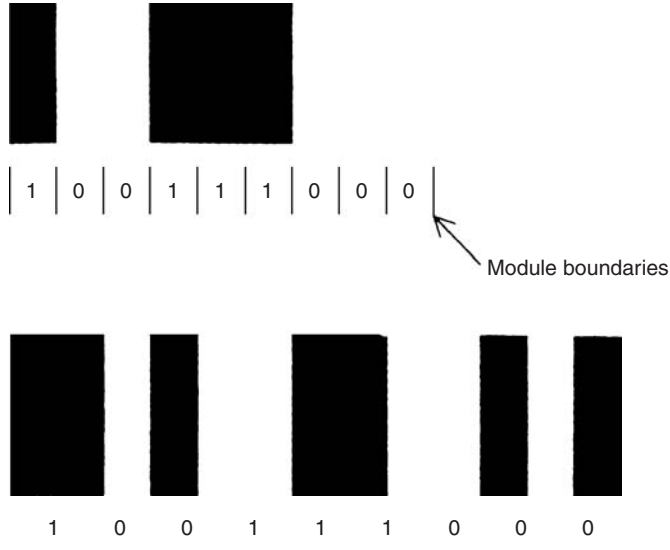


Figure 6.1 Encoding of the binary string 10011100 by delta code (top) and width code (bottom).

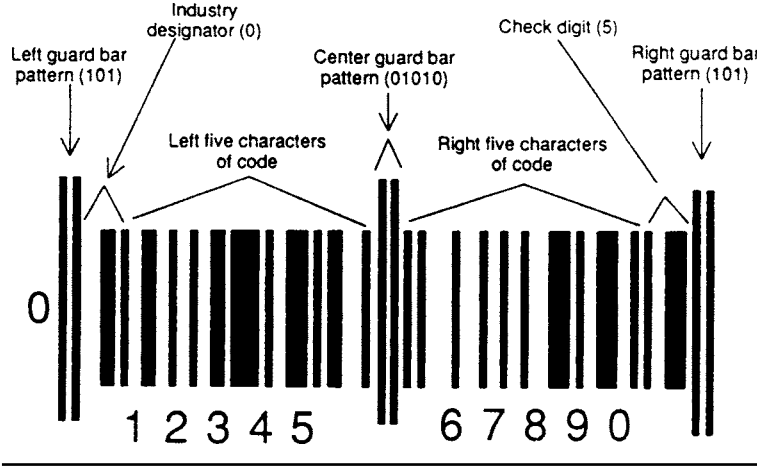


Figure 6.2 Specifications of the UPC symbol. The readable characters are normally printed in OCR-B font.

A bar-code reader can handle several different bar-code standards, decoding the stripes without knowing in advance the particular standard. The military standard, code 1189, specifies the type of coding to be used by the Department of Defense, which is a modification of code 39. Code 39 consists of 44 characters, including the letters A through Z. Because of its alphanumeric capabilities, code 39 is very effective for manufacturing applications. Code 39 is structured as

follows: three of nine bars (light and dark) form wide characters; the rest are narrow.

Bar-code labels are simple to produce. Code 39, for example, can be generated by a personal computer. Such labels are ideal for inventory identifications and other types of fixed-information gathering. Bar codes are not necessarily placed on labels. Tools, for example, have had the code etched on their surfaces to allow for tool tracking. Techniques have been developed for molding bar codes onto rubber tires. Holographic scanners allow reading around corners so that parts need not be oriented perpendicular to the reader as they feed down a processing line.

A difficulty with bar coding has been the fact that it cannot be read if the bars become obscured by dirt, grease, or other substances. Infrared scanners are used to read codes coated with black substances to prevent secrecy violations through reproduction of the codes. One way to generally offset the problem of a dirty environment is to use magnetic-stripe-encoded information.

6.1.2 Transponders

While bar-code labels and magnetic stripes are very effective on the shop floor, shop circumstances may require more information to be gathered about a product than can be realistically handled with encoded media. For instance, with automobiles being assembled to order in many plants, significant amounts of information are necessary to indicate the options for a particular assembly. Radio-frequency (RF) devices are used in many cases. An RF device, often called a transponder, is fixed to the chassis of a car during assembly. It contains a chip that can store a great amount of information. A radio signal at specific assembly stations causes the transponder to emit information that can be understood by a local receiver. The transponder can be coated with grease and still function. Its potential in any assembly operation is readily apparent. Several advanced transponders have read/write capability, thus supporting local decision making.

6.1.3 Electromagnetic Identification of Manufactured Components

Many other possible electronic schemes can be employed to identify manufactured parts in motion. Information can be coded on a magnetic stripe in much the same way that bars represent information on a bar-code label, since the light and dark bars are just a form of binary coding.

Operator identification data are often coded on magnetic stripes that are imprinted on the operators' badges. Magnetic stripe information can be fed into a computer. Such information might include the following: (1) the task is complete, (2) x number of units have been produced, (3) the unit part numbers, (4) the operator's identification

number, and so on. This same scanning station can also be set up using bar-code information; however, with magnetic striping, the information can be read even if the stripe becomes coated with dirt or grease. A disadvantage of magnetic striping is that the reader has to contact the stripe in order to recall the information.

6.1.4 Surface Acoustic Waves

A process similar to RF identification is *surface acoustic waves* (SAW). With this process, part identification is triggered by a radar-type signal that can be transmitted over greater distances than in RF systems.

6.1.5 Optical Character Recognition

Another form of automatic identification is *optical character recognition* (OCR). Alphanumeric characters form the information, which the OCR reader can "read." In mail processing centers, high-speed sorting by the U.S. Postal Service is accomplished using OCR. The potential application to manufacturing information determination is obvious.

Other means of part identification abound, such as vision systems and voice recognition systems. *Vision systems* utilize TV cameras to read alphanumeric data and transmit the information to a digital converter. OCR data can be read with such devices, as can conventionally typed characters. *Voice recognition systems* have potential where an individual's arms and hands are utilized in some function that is not conducive to reporting information. Such an application might be the inspection of parts by an operator who has to make physical measurements on the same parts.

In *laser scanning applications*, a laser beam scans and identifies objects at a constant speed. The object being scanned interrupts the beam for a time proportional to its diameter or thickness. Resolutions of less than 1 mm are possible.

In *linear array applications*, parallel light beams are emitted from one side of the object to be measured to a photooptical diode array on the opposite side. Diameters are measured by the number of array elements that are blocked. Resolutions of 5 mm or greater are possible.

In *TV camera applications*, a TV camera is used in the digitizing of the image of an object and the result is compared to the stored image. Dimensions can be measured, part orientation can be determined, and feature presence can be checked. Some exploratory work is being accomplished with cameras that can fit in a tool changer mechanism. The camera can be brought to the part like a tool and verify part characteristics.

6.2 Digital Encoder Sensors

Digital encoder sensors provide an output directly in a digital form and thus require only simple signal conditioning. They are also less

susceptible to electromagnetic interference, and are therefore useful for information processing and display in measurement and control systems. Their ability to rapidly scan a series of patterns provides additional manufacturing automation opportunities when light and dark patterns are placed in concentric rings in a disk. Figure 6.3 illustrates a portion of such a disk that can be rigidly attached to a shaft or an object and housed in an assembly containing optical sensors for each ring (Fig. 6.4). The assembly, called an *optical encoder*, automatically

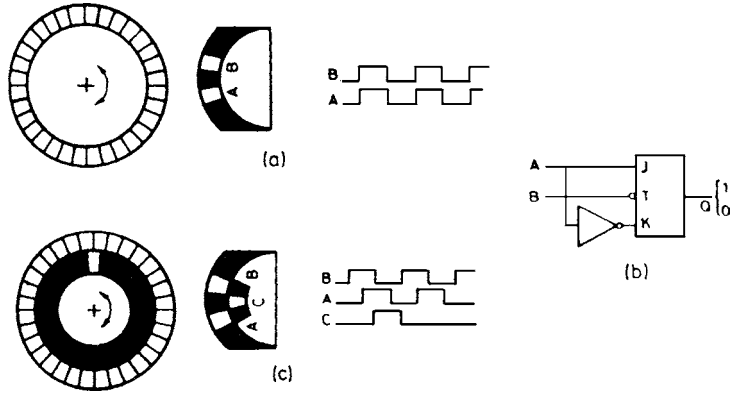


Figure 6.3 Detection of movement direction in directional encoders: (a) by means of two outputs with 90° phase shift; (b) output electronic circuit; (c) additional marker for absolute positioning.

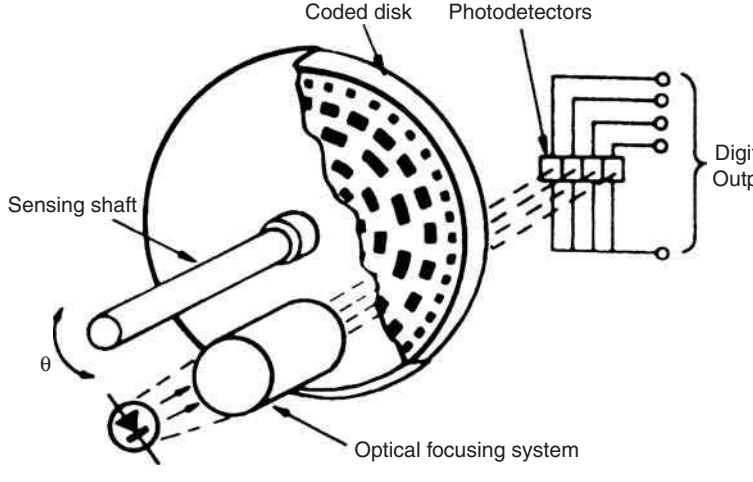


Figure 6.4 Principle of absolute position encoders for linear and rotary movements.

detects rotation of a shaft or an object. The shaft rotation information can be fed back into a computer or controller mechanism for controlling the velocity or position of the shaft. Such a device has application in robots and numerical control machine tools and for precision measurements of strip advancement to generate a closed-loop feedback actuation for displacement compensation.

The two classes of digital encoder sensors are:

- Encoder sensors yielding at the output a digital version of the applied analog input. This class of encoder sensors includes position encoders.
- Encoder sensors that rely on some physical oscillatory phenomenon that is transduced by a conventional modulating sensor. This class of sensors may require an electronic circuit acting as a digital counter in order to yield a desired digital output signal.

No sensors exist where the transduction process directly yields a digital output. The usual process is to convert an analog input quantity into a digital signal by means of a sensor without the requirement to convert an analog voltage into its digital equivalent.

6.2.1 Position Encoder Sensors in Manufacturing

Position encoder sensors can be categorized as linear and angular position encoder sensors. The optical encoder sensor can be either incremental or absolute. The incremental types transmit a series of voltages proportional to the angle of rotation of the shaft or object. The control computer must know the previous position of the shaft or object in order to calculate the new position. Absolute encoders transmit a pattern of voltages that describes the position of the shaft at any given time. The innermost ring reaches from dark to light every 180°, next ring every 90°, the next every 45°, and so on, depending on the number of rings on the disk. The resulting bit pattern output by the encoder reveals the exact angular position of the shaft or object. For an absolute optical encoder disk that has eight rings and eight LED sensors, and in turn provides 8-bit outputs, (10010110). Table 6.1 shows how the angular position of the shaft or object can be determined.

The incremental position encoder sensor suffers from three major weaknesses:

- The information about the position is lost whenever the electric supply fails or the system is disconnected, and when there are strong perturbations.
- The digital output, to be compatible with the input/output peripherals of a computer, requires an up/down counter.

Encoder Ring	Angular Displacement, Degrees	Observed Pattern	Computed Value, Degrees
1 (innermost)	180	1	180
2	90	0	
3	45	0	
4	22.5	1	22.5
5	11.25	0	
6	5.625	1	5.625
7	2.8125	1	2.8125
8	1.40625	0	210.94

TABLE 6.1 Absolute Optical Encoder

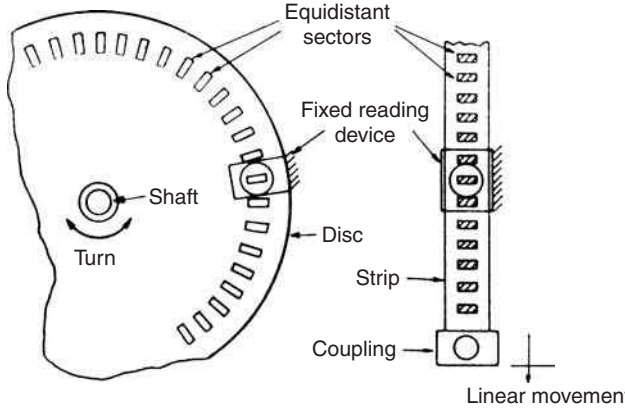


FIGURE 6.5 Principle of linear and rotary incremental position encoders.

- The incremental position encoder does not detect the movement direction unless elements are added to the system (Fig. 6.5).

Physical properties used to define the disk pattern can be magnetic, electrical, or optical. The basic output generated by the physical property is a pulse train. By differentiating the signal, an impulse is obtained for each rising or falling edge, increasing by two the number of counts obtained for a given displacement (Fig. 6.6).

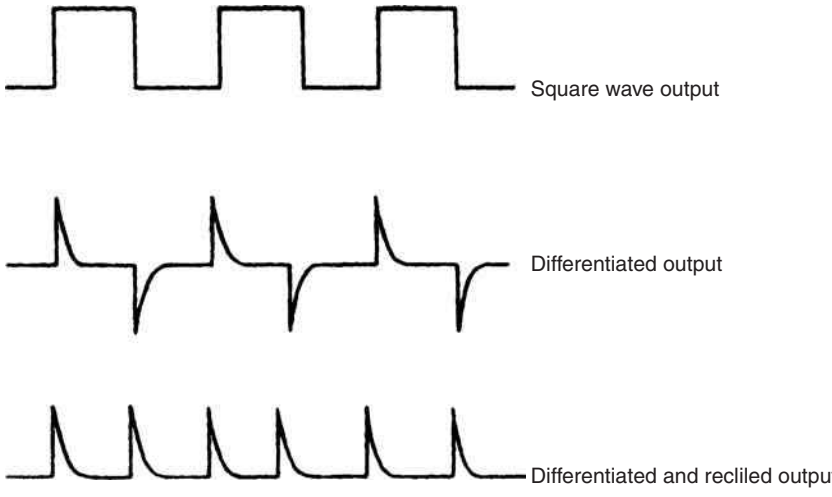


Figure 6.6 Improving output resolution of an incremental encoder by differentiation and rectification.

6.3 Fuzzy Logic for Optoelectronic Color Sensors in Manufacturing

Fuzzy logic will most likely be the wave of the future in practical and economical solutions to control problems in manufacturing. Fuzzy logic is simply a technique that mimics human reasoning. This technology is now being explored throughout various industries. Fuzzy logic color sensors can relay information to microprocessors to determine color variance within an acceptable range of colors. A conventional sensor could not perform this function because it could choose only a specific color and reject all other shades—it uses a very precise set of rules to eliminate environmental interference.

The research and development activities for fuzzy logic technology began in mid-1990. The research has lead to the creation of a fuzzy logic color sensor that can learn a desired color and compare it with observed colors. The sensor can distinguish between acceptable and unacceptable colors for objects on a conveyer belt. The development of new light source technology allows the color sensor to produce more accurate color measurement (Fig. 6.7). Also, the integration of the fuzzy logic sensor with a microprocessor enables the data to be collected and interpreted accurately.

6.3.1 Sensing Principles

The sensor is designed with a broad-spectrum solid-state light source utilizing a light-emitting diode cluster. The LED-based light source provides stable, long-lasting, high-speed target illumination capabilities.

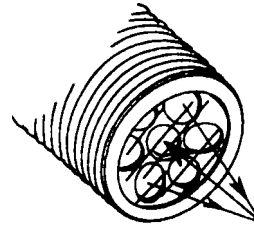


Figure 6.7 Integration of a fuzzy logic sensor in a sensing module.

The LED cluster is made up of three representative hues of *red*, *green*, and *blue*, which provide a triple bell-shaped spectral power distribution for the light source (Fig. 6.8). The light incident on the target is reflected with varying intensities, depending on the particular target color under analysis.

The reflected light is received by a semiconductor receiver in the center of the LED cluster. The amount of light reflected back onto the receiver is transduced to a voltage and converted to digital format immediately via an analog-to-digital converter. The internal processing of the converted red, green, and blue (RGB) values offers variable sample size and averaging to compensate for signal noise (Fig. 6.8).

Ambient light is sampled between every component pulse and immediately subtracted from the sampled signal so the effects of factory ambient light are suppressed. Thus, hooding the sensor is not totally necessary. In an area of a very bright or high-frequency lighting, it may be beneficial to provide some degree of hooding to at least limit the brightness. The sensor's electronics also employs temperature compensation circuitry to stabilize readings over temperature ranges.

6.3.2 Color Theory

Color science defines color in a space, with coordinates of *hue*, *saturation*, and *intensity* (HSI). These three general components uniquely

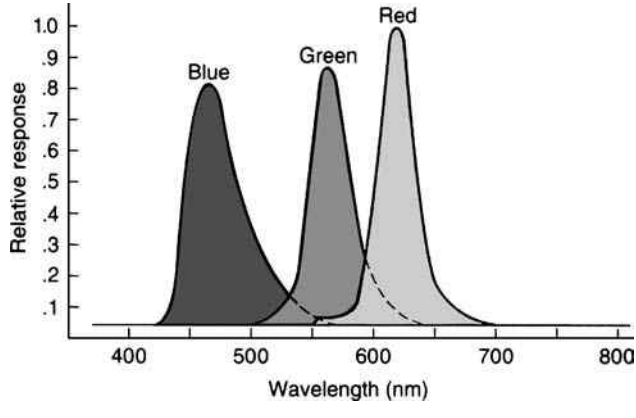


Figure 6.8 Spectra of red, blue, and green light.

define any color within HSI color space. Hue is related to the reflected wavelength of a color when a white light is shined on it. Intensity (lightness) measures the degree of whiteness, or gray scale, of a given color. Saturation is a measure of the vividness of a given hue. The term *chromaticity* primarily includes elements of the hue and saturation components. Researchers depict color in space using hue as the angle of a vector, saturation as the length of it, and intensity as a plus or minus height from a center point (Fig. 6.9).

The concepts of hue, saturation, and intensity can be further clarified by a simplified pictorial presentation. Consider Fig. 6.10, where a color is depicted at a molecular level. Color is created when light interacts with pigment molecules. Color is generated by the way pigment

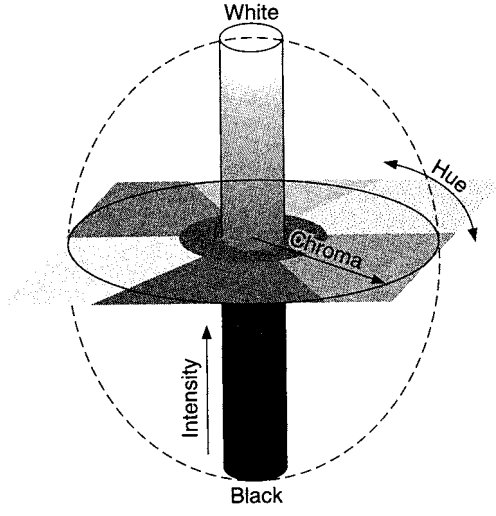


Figure 6.9 Coordinates of hue, saturation, and intensity of color in space.

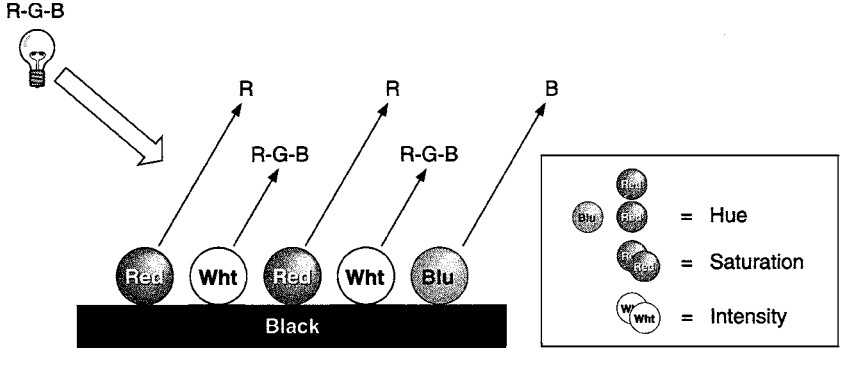


Figure 6.10 Model of color interpretation.

molecules return (bend) incoming light. For example, a red pigment causes a measurable hue component of the color. The relative density of the pigment molecules leads to the formation of the saturation component. Some molecules return almost all wavelengths, and appear white as a result, leading to the intensity (lightness) component.

6.3.3 Units of Color Measurement

If color description depends on measuring the interaction of a target color with a given white light source, it is clear that in order to have the system of measurement standardized, both the light source and means of detection must be well-defined. One very popular set of standardization rules has been set up by the Commission International de l'Eclairage (CIE), a color standardization organization. From color theory, it is known that the response to a color stimulus (its determination), depends on the spectral power distribution of the light source (illuminant), times the spectral reflectance of the target (color) surface, times the spectral response of the detector (observer) (Fig. 6.11).

With this principle in mind, the CIE presented a detailed description of the standard light source and a standard observer (photodetector). The result of the study was the popular CIE diagram, which creates a two-dimensional mapping of a color space (Fig. 6.12).

Further manipulation of the CIE observations has lead to another color coordinate system, the so-called L.a.b numbers for describing a color. The L.a.b. numbering system is fairly prevalent in industrial applications. The machines that measure color according to this theory are referred to as *color spectrophotometers* or *color meters*. These machines are typically expensive, bulky, and not well-suited for distributed online color sensing.

The fuzzy color sensor does not offer CIE-based color measurement; however, it is a very high-resolution color comparator. The sensor learns a color with its own standard light source (trio-stimulus LEDs) and its own observer (semiconductor photoreceiver). It thereby sets up its own unique color space with the three dimensional coordinates being the red, blue, and green readings (Fig. 6.12).

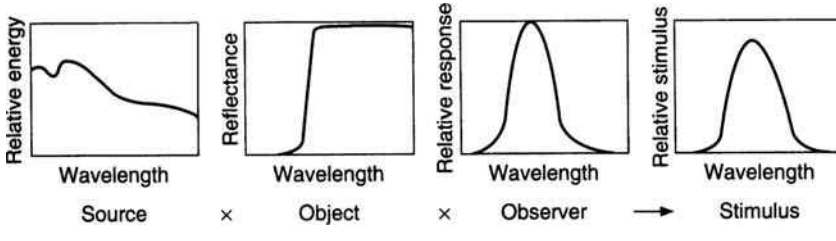


Figure 6.11 Stimulus response to a color (detector/determination) = illuminant × target × observer.

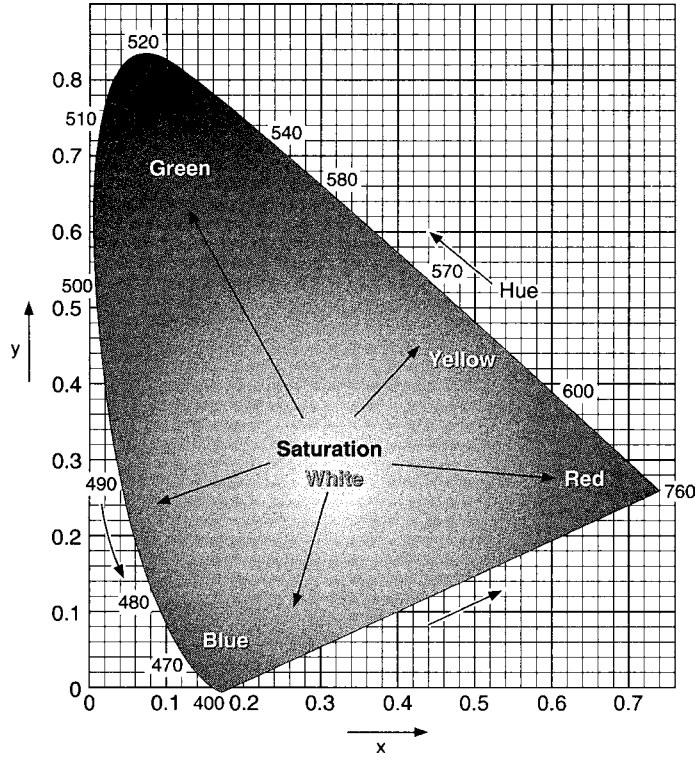


Figure 6.12 Two-dimensional mapping of color in a space.

Theoretically, there would be 256^3, or 16,777,216, unique positions in its color space for defining color (Fig. 6.13). In reality, the actual number of colors a sensor can reliably distinguish is much less because of optical noise and practical limitations of the design.

The device compares the RGB colors it observes to the internally learned standard. Each time a standard color is relearned, it essentially recalibrates the sensor.

6.3.4 Color Comparators and True Color Measuring Instruments

If the learned color is initially defined by a color spectrometer, the fuzzy logic color sensor (comparator) can be installed in conjunction with it. The color sensor can learn the same area that the spectrometer has read, thereby equating its learned standard to the absolute color. By using the color sensor in this manner, a temporary correlation to an absolute color standard can be established (Fig. 6.14). This permits

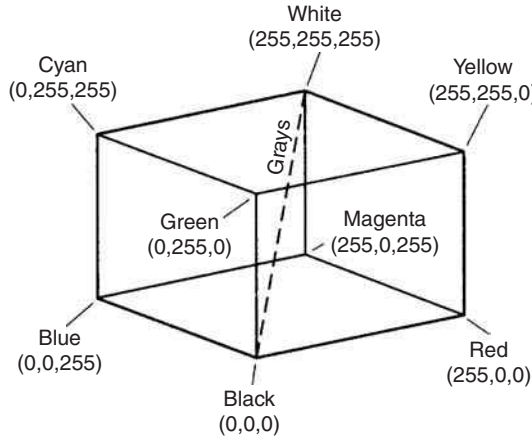


Figure 6.13 Three-dimensional coordinates of color in a space.

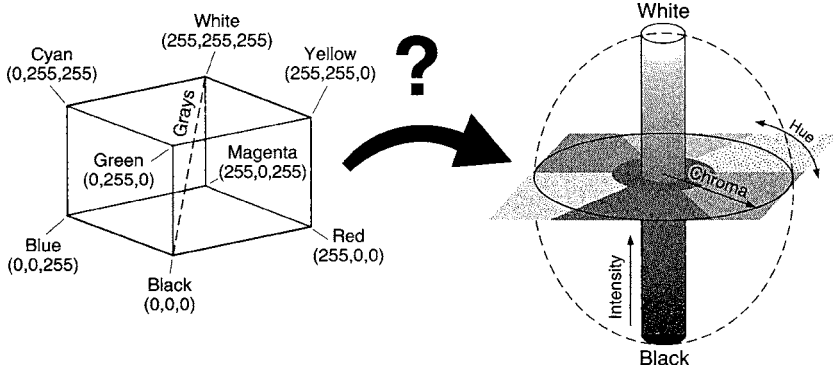


Figure 6.14 Correlation to an absolute color standard.

the relative drift from the standard to be monitored. The advantage is that more color sensing can be economically distributed across the target area. When a significant relative deviation is detected, an alert signal can flag the absolute color sensor to take a reading in the suspect area. If enough storage space is available in a central processing computer, lookup tables can be constructed to relate the serially communicated color sensor readings to a standard color coordinate system like the CIE system.

6.3.5 Color Sensor Algorithms

There are two internal software algorithms, or sets of rules, for analyzing fuzzy logic color sensor data:

- *The absolute algorithm.* Compares color on the basis of absolute voltages.
- *The relative algorithm.* Compares color on the basis of relative percentages of each RGB component voltage.

The choice of algorithm depends on sensing distance variation and the type of color distinction one needs. If the outputs vary excessively with distance when the absolute algorithm is used, a relative (ratio) algorithm must be considered. While a relative algorithm does not retain the lightness information, it greatly reduces unwanted distance-related variations. The relative algorithm shows changes in chromaticity (hue and chroma) that exist in most color differences. If the sensing distances can be held constant, the absolute algorithm works well at detecting subtle changes in density (shades) of a single color.

6.4 Optimum Detectors in Light Sensors

Deciding on the type of detectors to use for light sensing applications can be overwhelming, given the kinds available, which include photodiodes, phototransistors, photodarlingtons, photomultiplier tubes, photo-resistors, integrated circuits, various hybrids, and even thermopiles. A vital insight on selecting the appropriate approach for ultraviolet, visible, and near-infrared light sensing applications is imperative.

The light source spectral characteristics require optical power mating of various electronic packages with the following constraints:

- Image size
- Signal-to-noise ratio
- Frequency bandwidth
- Cost

Most corporations do not have resident experts in the fields of modern optics—especially in the area of light detection. As a result, when a new product is being developed, the light sensing design project is usually assigned to a mechanical or electrical engineer. To meet project schedules (for these optics for non-experts), individuals must acquire the necessary knowledge on the various light sensing methods available. A simple guideline briefly describing a variety of light sensing technologies and options is presented next.

6.4.1 Available Light Sensing Options

Light sensing applications vary widely from specialized scientific instrumentation that needs to detect individual light particles

(photons) to systems that control high-speed welding and cutting lasers that produce kilowatts of optical power.

Sensors exist for almost any application imaginable: from a photomultiplier tube, which gives a large voltage pulse for every photon it detects, to cooled thermopiles that absorb kilowatts of power providing a thermocouple voltage proportional to the optical power absorbed. The following describes the most popular light sensing technologies. Their characteristics are summarized in Table 6.2.

Photomultiplier tubes are special vacuum tubes that have a light sensing surface, the photocathode, which absorbs incoming light photons and emits secondary electrons. These secondary electrons are accelerated and multiplied within the photomultiplier tube by dynode plates. Each time an electron strikes a dynode, it has gained enough momentum to create a larger number of secondary electrons. This multiplication process continues for each dynode within the tube. Tubes with 10 to 12 dynodes can easily generate multiplications of more than a million, resulting in sufficient current to develop hundreds of millivolts across an output 50-ohm load resistor for a single incident photon.

Photomultiplier tubes provide the ultimate in detection sensitivity. They can sense the smallest amount of optical energy there is—an individual photon. When cooled, it can be essentially noise-free, with at most one false photon pulse in a one-second time period.

However, these light sensor detectors have a few drawbacks:

- They are mechanically fragile.
- They need an extremely stable high-voltage power supply.
- They are relatively expensive.
- Shapes and sizes are very limited.
- They are susceptible to external magnetic fields.
- The available photocathodes are limited to sensing ultraviolet to near-infrared wavelengths between 190 to 900 nm, and can be extended to 1100 nm.

Photomultiplier tubes are generally used to detect the lowest light levels where the application demands their superior sensitivity, such as in high-speed spectroscopy applications.

6.5 Photodiodes

Photodiodes are light-sensitive semiconductor devices that are manufactured in essentially the same way as semiconductor diodes used in conventional electronic circuits. The primary differences are that photodiode chips are larger and are packaged to allow light onto the sensitive area of the diode.

Electrical Characteristics	Photo-multiplier	Tubes Photo-diodes	Photo-transistors	CdS Photocells	Other Photo-conductors	Integrated Circuits	Hybrids	Sensor Electronic Assembly
Available Wavelengths (μm)	0.2–0.9	0.2–2.0	0.4–1.1	0.4–0.7	2–15	0.2–1.1	0.2–15.0	0.2–15.0
Performance- to-cost ratio	Fair	Good	Excellent	Excellent	Fair	Fair	Fair	Good
Sensitivity	Excellent	Very Good	Very Good	Very Good	Very Good	Very Good	Very Good	Very Good
Linearity	Good	Excellent	Good	Good	Good	Good	Good	Good
Ambient Noise Performance	Excellent	Fair	Very Good	Very Good	Very Good	Very Good	Excellent	Excellent
Dynamic Range	Very Good	Excellent	Very Good	Good	Good	Very Good	Very Good	Very Good
Stability	Very Good	Very Good	Good	Poor	Fair	Very Good	Very Good	Very Good
Other Characteristics								
Reproducibility	Fair	Excellent	Fair	Poor	Fair	Very Good	Very Good	Very Good
Cost	High	Low	Very Low	Very Low	High	Medium	High	Medium
Ruggedness	Poor	Excellent	Excellent	Excellent	Good	Excellent	Very Good	Excellent
Physical Size	Large	Small	Small	Small	Small	Small	Medium	Medium
Ease of Customization	Poor	Easy	Fair	Fair	Poor	Poor	Poor	Fair
Cost of Customization	Very High	Low	Medium	Low	High	Very High	High	Medium
Lead time for Customization (in weeks)	40	12	14	12	20	40	30	16

Table 6.2 Comparison of Light Sensor Characteristics in Photomultiplier Tubes

Photodiodes offer many conveniences and advantages that make them very practical for a wide range of applications:

- They can easily measure from picowatts to milliwatts of optical power.
- They come in standard packages, or the package can be tooled to the application.
- Depending on the semiconductor material used, they can detect wavelengths from 190 to over 2,000 nm.
- They are small and lightweight.
- Almost any photosensitive shape can be fabricated for small capital investment as low as $3,000 for tooling.
- They have very reproducible sensitivity.
- Very large areas can be fabricated: > 10 cm^2.
- They can be very responsive, with rise-times as fast as 10 picoseconds.

If noise presents an obstacle when measuring a few picowatts of light with a standard photodiode, consider the advantages of an avalanche photodiode (APD) which offers a current gain internal to the photodiode structure of up to 100 times.

Photodiodes generally require a pre-amplifier to give signal gain for applications to detect picowatts of light power. But for high optical power of <10 microwatt levels, a simple load resistor configuration can give adequate performance and TTL-compatible voltage swings.

Silicon based photodiodes cover the wide range of wavelengths from 190 to 1100 nm. The lower limit is set by absorption of ultraviolet light in air. Germanium (Ge) photodiodes overlap the silicon response spectrum and are usable to about 1600 nm. Semiconductors that are compounds of gallium, arsenide, indium, antimonide, and phosphorous can be specially fabricated to cover small sections of the 190-to-2000-nm spectral range. For instance, the fiber optics industry uses indium-gallium-arsenide "InGaAs" detectors for the 800 to 1800 nm range. More exotic and expensive photodiodes can sense energy much further out in the IR spectrum.

Photodiodes are widely used in high-tech societies, in applications ranging from sensors for door openings, assembly line controls, load levelers in luxury cars, to personal blood sugar meters for diabetics, sun-tan exposure meters, smoke detectors, x-ray baggage inspection systems, and even cranial pressure sensors for head injury patients.

6.6 Phototransistors and Photodarlingtons

Phototransistors are transistors designed to capture light and are assembled in a transparent package. They are often more convenient

than photodiodes because they have built in gain: the absorbed light creates a current in the base region of the phototransistor, resulting in current gains from 100 to several thousands. Photodarlingtons have two stages of gain, with net gains that can be greater than 100,000.

The built in gain allows the phototransistor to be coupled with a load resistor to accommodate TTL level voltages for a wide range of light levels. Because of their ease of use, low cost, and TTL-compatible signal levels, phototransistors have become popular for applications where there is more than a few hundred nanowatts of available optical power.

These devices however, do have some drawbacks compared to photodiodes. The frequency bandwidth and linearity are relatively limited and spectral response is restricted to between 350 and 1100 nm. In addition, there are very large variations in sensitivity between individual devices and few standard package options.

6.7 Photoconductive Sensors

A photoconductive sensor is a thick film semiconductor material whose electrical resistance decreases with increasing incident light. These rugged assemblies that can withstand hundreds of volts are typically smaller than 0.25 inches in diameter.

Photoconductive sensors based on cadmium sulfide (CdS) have sensitivity curves that closely match the sensitivity of the human eye. Accordingly, they are useful in applications involving human light perception such as headlight dimmers and intensity adjustments on information displays. These sensors can be designed for measuring microwatts to milliwatts of optical power and are very inexpensive at high volume (less than $0.10 each). These characteristics make CdS photoconductors the sensor of choice in applications such as street light control and in the toy industry where economy is a major consideration.

There are, however, considerations that limit the use of CdS photoconductors in more sophisticated applications requiring sensitivities over a wide spectral range, small variations between individual parts, or no history-dependent response. The resistance of these sensors depends on the thick-film microstructure, so the resistance specification has a wide tolerance—a max/min ratio of 3 is not uncommon. The resistance also has long-term memory that depends, at any given time, on the amount of light actually incident on the sensor plus the sensor light history for the past several days.

Photoconductors made from materials other than CdS such as lead telluride and mercury cadmium telluride are also available. These materials have spectral sensitivities that cover the range that photodiodes cannot: above 2 μm out to 15 μm. This longer wavelength sensitivity is very important for infrared imaging cameras and for long wave instrumentation such as that used to monitor carbon

dioxide laser emission and atmospheric physics. These sensors tend to be more expensive than both silicon photodiodes and CdS photoconductors.

6.7.1 Integrated Circuits

Incorporating additional electronics directly onto a semiconductor sensor chip makes it possible to add additional functions to the sensor. An optical IC is an integrated circuit comprising photodiode and electronic-signal-processing-circuits. Such additional functions as current-to-voltage conversion and reference-level sensing (a Schmitt trigger, for example) can be incorporated. Other optical ICs can provide signals highly immune to noise, such as a current-to-frequency conversion.

The principal advantages of an optical IC are ease of use, small size, and immunity to electronic noise compared to a photodiode with separate electronics. Typically, these devices are much more expensive and offer a very limited active light-sensing area. Custom tooling for specific applications is also expensive.

6.7.2 Hybrids

The electronic functions of an optical IC can also be provided by a hybrid circuit that has an unpackaged IC components "die" attached to a substrate that also contains a photodiode.

This type of sensor combines the ease of use and immunity to electrical noise of an optical IC with increased design flexibility and lower tooling costs. In addition, the sensitivity can easily be increased with a larger photodiode active area without the added cost of a separate detector. The primary disadvantages of a hybrid sensor are its cost and reliability. Cost can be several times higher than the electronic-assembly option discussed next, and reliability testing is difficult to amortize, so either limited reliability screening is implemented, or the cost per unit becomes high.

6.8 Sensor Electronic Assemblies

Combining any of the sensors listed earlier with printed-circuit-based electronic signal processing creates sensor assemblies or black boxes. The user defines specifications for light input and the desired output response; the vendor builds and tests the systems to ensure that the specifications are met. An assembly can also include optical components such as lenses and special wavelength filters. The user just bolts the assembly in place and connects it to the high-level electronics; there are no concerns about mismatch between the purchased sensor and front-end amplifiers or diagnostic electronics. The system is relatively immune from noise and is highly reliable because of the mature manufacturing technologies used.

6.9 Other Sensors

Many other types of sensors exist. These include avalanche photodiodes, APD, bolometers, self-scanned arrays, and photon drag detectors. A sensor manufacturer can provide information about these devices and can discuss the physics and advantages of each detector technology.

6.10 Selecting a Sensor

Reviewing a few key design aspects generally provides enough information for making an optimum choice of detector for a given application (Fig. 6.15).

6.11 Infrared Light Domain

For sensing wavelengths below 1100 nm, photoconductive cells or a silicon-based detector should be appropriate. At wavelengths above 1100 nm, the costs and technology options are not straightforward, and a detector vendor consulted at the beginning of a design program will provide the most effective guidance.

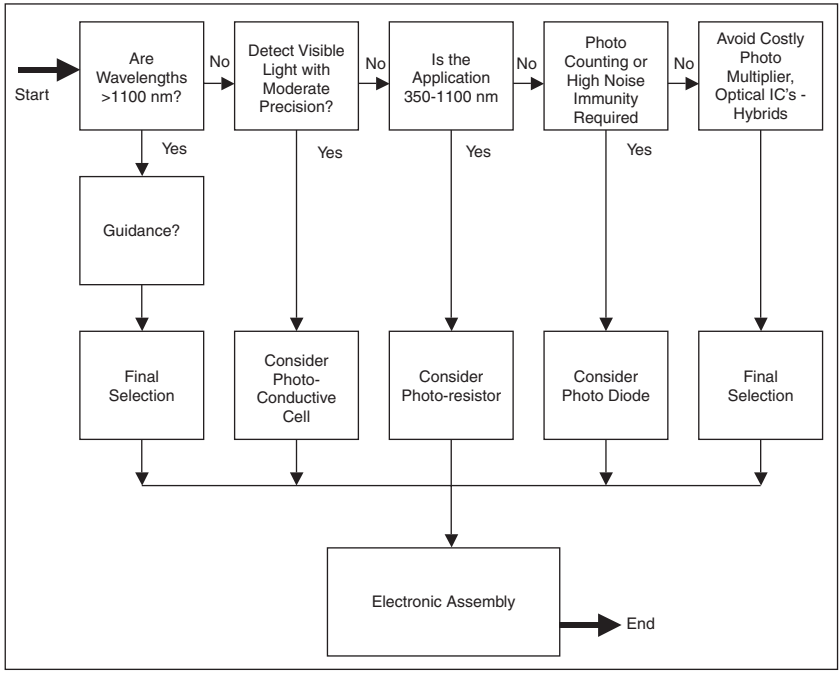


Figure 6.15 Selecting appropriate detectors for specific sensing applications.

6.12 Visible Light Region

Applications at visible wavelengths with at least microwatts of optical power, in which the sensor is simply required to detect if light is present, can use one of the least expensive and most rugged detectors available: the photoconductive cell. Many standard devices are available as well as custom-designed options, including complete electronic assemblies.

6.13 Nanowatts Power

Silicon phototransistors and photodarlingtons—the second-lowest-cost sensors—are also convenient and should be considered as the next possible alternative. Only nanowatts of optical power within a 5-mm diameter spot are required at wavelengths between 350 and 1100 nm. The application must also tolerate some unit-to-unit sensitivity variation.

6.14 Considering High-Performance Photodiodes

For UV to near IR-wavelengths, photodiodes offer the best overall performance. They are only slightly more expensive than phototransistors, but their spectral range is broader and they have lower noise, more uniform sensitivity and reproducibility, a larger dynamic range, better linearity, and more package options. Also, photodiodes can routinely detect picowatts of optical power. Thus, if phototransistors or photoconductive cells are not appropriate for an application, more often than not a photodiode will afford the best alternative.

6.15 Hybrids of Photomultipliers Options

At least 90 percent of detector applications should be satisfied by one of the sensors discussed earlier in this chapter. However, when light levels are extremely low, or ambient electronic noise levels are high, or space requirements are particularly limited, other alternatives, such as optical ICs, hybrids, or photomultipliers should be seriously investigated.

6.16 Design Considerations in Fuzzy Logic Color Sensors

The design of fuzzy logic color sensors aims to achieve maximum color sensing ability, while maintaining the expected simplicity of operation and durability of typical discrete industrial sensors. Several other key goals must be considered in system design such as high speed, small size, configurability, high repeatability, and long light source life.

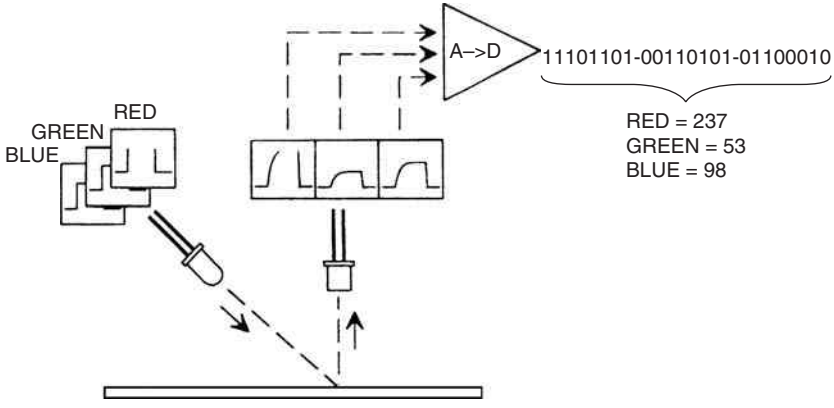


Figure 6.16 Conversion of red, green, and blue LED outputs from analog to digital.

The choice of a solid-state light source satisfies the majority of the criteria for a good industrialized color sensor design. The fuzzy logic color sensor utilizes two sets of three different LED photodiodes as its illumination source. The three LED colors—red, green, and blue—were chosen essentially for their coverage of the visible light spectrum (Fig. 6.16).

The light from each LED is sequentially pulsed onto the target, and its reflected energy is collected by a silicon photoreceiver chip in the LED cluster. Ambient light compensation circuitry is continually refreshed between each LED pulse, so the reported signals are almost entirely due to the LED light pulses. The LED sources offer a very fast (microsecond response) and stable (low spectral drift, steady power) source of a given wavelength band, without resorting to filters. Emerging blue LEDs, in combination with the more common red and green LEDs, have made it possible to use three solid-state spectra to define a hue. The choice of the specific LED bandwidth (or spectral distribution) is made so as to obtain the best color distinction through broader coverage of the illumination spectrum.

6.16.1 Fuzzy Logic Controller Flowcharts

The entire sensor operation is illustrated in Fig. 6.17. An internal microcontroller governs the device operations. It directs signals in and out of the sensor head, to maintain local and remote communications, and provides color discrimination algorithms to produce the appropriate signal output at the control pin. As the device proceeds out of reset, it checks the locally (or remotely) set configuration dipswitches, which in turn define the path, or operating menu, to proceed through. There is permanent storage of learned or remotely loaded values of RGB readings, tolerance, number of readings to

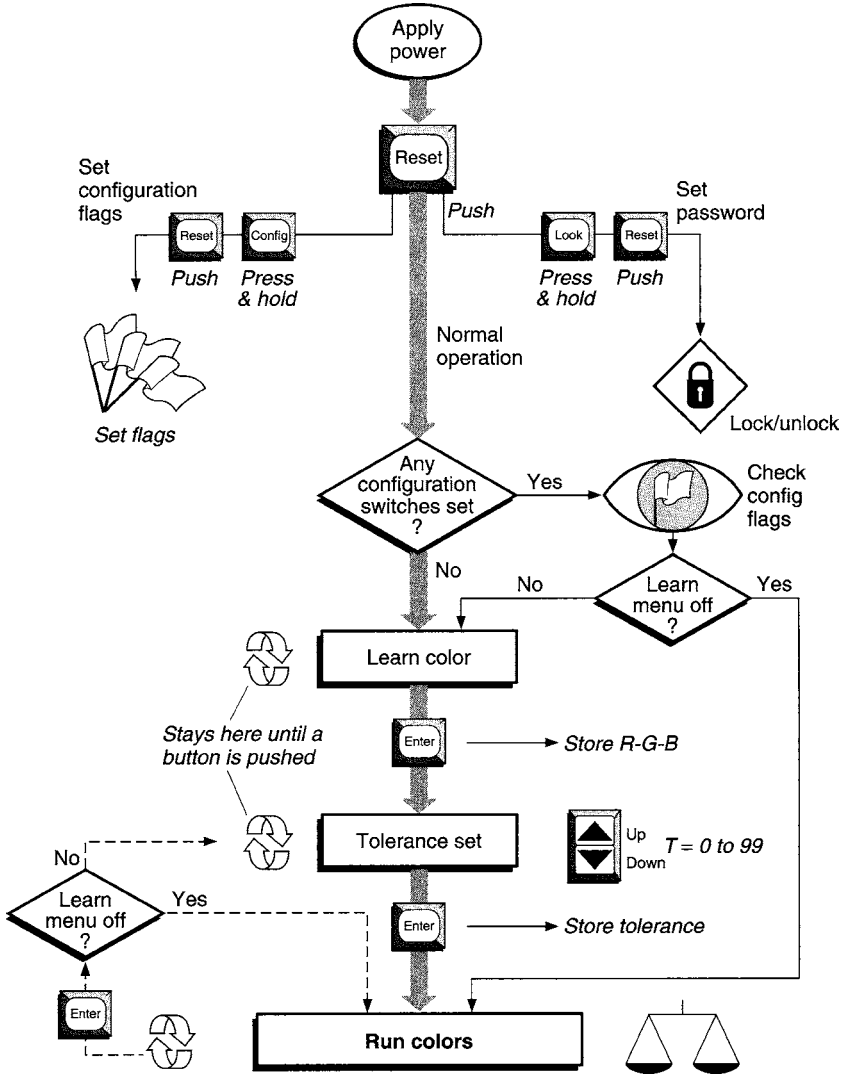


Figure 6.17 A fuzzy-logic controller flowchart.

average, and the white-card calibration value, so these settings can be available at reset or at power-up.

One or more of three optional menus can be selected before entering into the main operation of learning and running colors. The three alternative menus are (1) white-card gain set, (2) number of reads to average set, and (3) the observed stored reading menus.

If none of the alternative menus is activated, the sensor will proceed directly to the primary modes, which are learn, tolerance set,

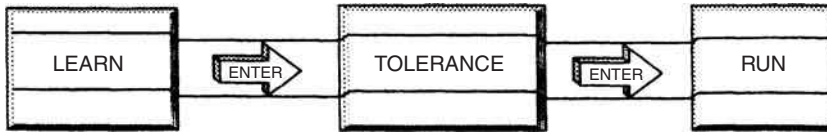


Figure 6.18 Simple path of operation.

and run modes (Fig. 6.18). By pressing and holding the appropriate buttons while pushing the reset button, two other programming menus can be entered. These are the set configuration flags (dip-switch) menu and the set password menu.

6.17 Sensors Detecting Faults in Dynamic Machine Parts (Bearings)

A system consisting of analog and digital signal processing equipment, computers, and computer programs would detect faults in ball bearings in turbomachines and predict the remaining operating time until failure. The system would operate in real time, extracting the diagnostic and prognostic information from vibrations sensed by accelerometers, strain gauges, and acoustical sensors, and from the speed of the machine as measured by a tachometer.

The vibrations that one seeks to identify are those caused by impact that occurs when pits in balls make contact with races and pits in races make contact with balls. These vibrations have patterns that are unique to bearings and repeat at known rates related to ball-rotation, ball-pass, and cage-rotation frequencies. These vibrations have a wide spectrum that extends up to hundreds of kilohertz, where the noise component is relatively low.

The system in Fig. 6.19 would accept input from one of two sensors. Each input signal would be amplified, bandpass-filtered, and digitized. The digitized signal would be processed in two channels: one to compute the keratosis of the distribution of the amplitudes, the other to calculate the frequency content of the envelope of the signal. The *keratosis* is the fourth statistical moment and is known, from theory and experiment, to be indicative of vibrations caused by impact on faults. The keratosis would be calculated as a moving average for each consecutive digitized sample of the signal by using a number of samples specified by the technician. The trend of a keratosis moving average would be computed several times per second, and the changes in the keratosis value deemed to be statistically significant would be reported.

In the other signal processing channel, the amplitude envelope of the filtered digitized signal would be calculated by squaring the signal. Optionally, the high-frequency sample data would be shifted

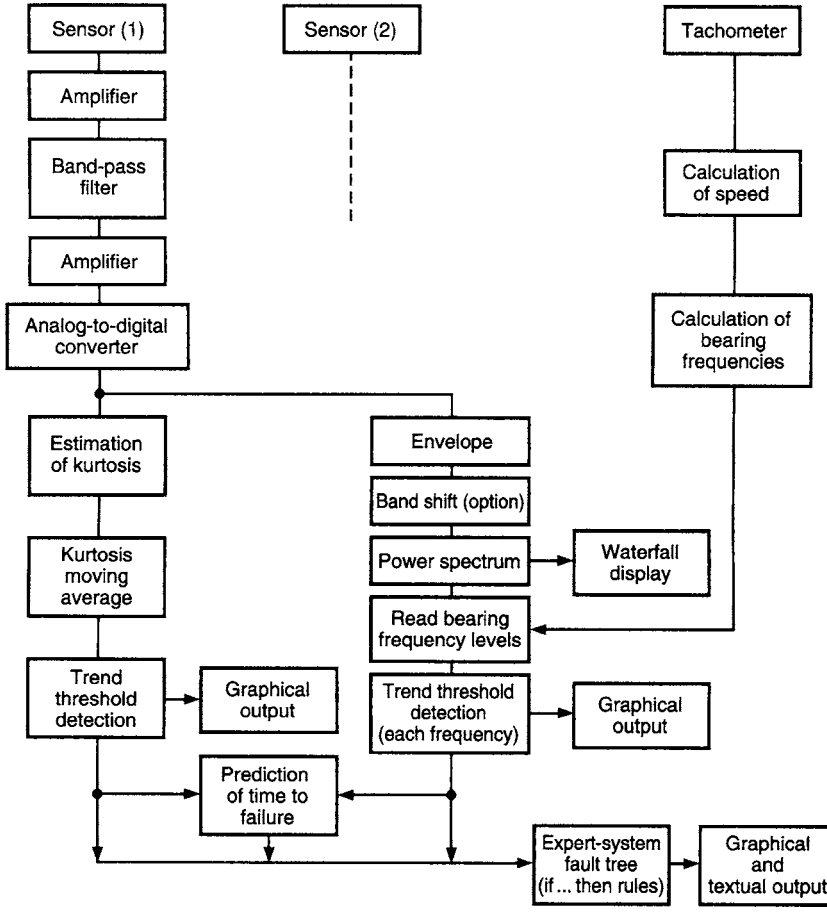


Figure 6.19 Data flow for an automatic bearing fault detection system.

to a lower frequency band to simplify processing by use of a Fourier transformation. This transformation would then be applied to compute the power spectrum.

The output of the tachometer would be processed in parallel with the spectral calculations so that the frequency bins of the power spectrum could be normalized on the basis of the speed of rotation of the machine. The power spectrum would be averaged with a selected number of previous spectra and presented graphically as a "waterfall" display; this is similar to a sonar display with which technicians can detect a discrete frequency before an automatic system can.

The bearing frequencies would be calculated from the measured speed and the known parameters of the bearings, with allowances for slip. The power spectrum levels would be read for each bearing frequency; a moving average of the amplitude at each bearing

frequency and harmonic would be maintained, and trends representing statistically significant increases would be identified by threshold detection and indicated graphically.

By using algorithms based partly on analyses of data from prior tests, the results of both keratosis and power spectrum calculations would be processed onto predictions of the remaining operating time until failure. All the results would then be processed by an expert system. The final output would be a graphical display and text that would describe the condition of the bearings.

6.18 Sensors for Vibration Measurement of a Structure

An advanced sensor was developed to gauge structure excitations and measurements that yield data for design of robust stabilizing control systems (Fig. 6.20).

An automated method for characterizing the dynamic properties of a large flexible structure estimates model parameters that can be used by a robust control system to stabilize the structure and minimize undesired motions. Although it was developed for the control of large flexible structures in outer space, the method is also applicable to terrestrial structures in which vibrations are important—especially aircraft, buildings, bridges, cranes, and drill rigs.

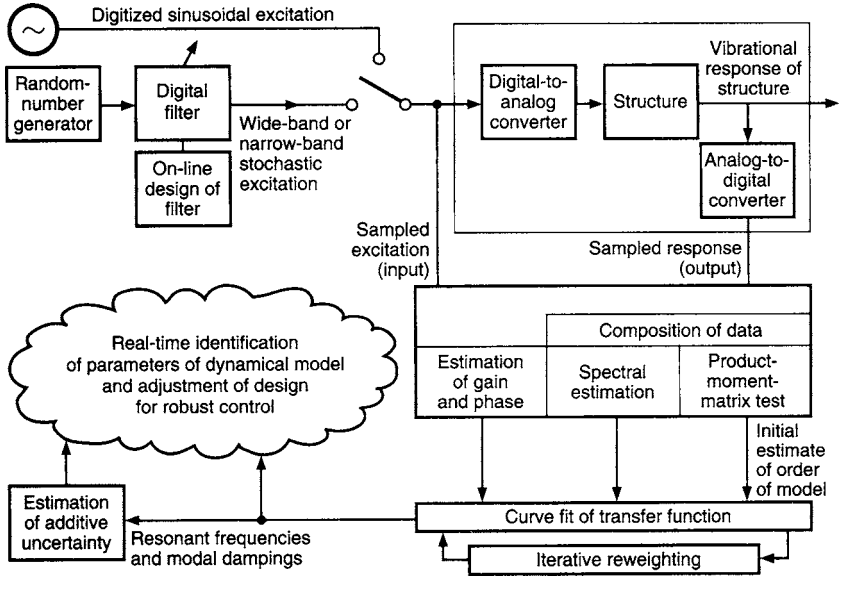


Figure 6.20 Automated characterization of vibrations of a structure.

The method was developed for use under the following practical constraints:

- The structure cannot be characterized in advance with enough accuracy for purposes of control.
- The dynamics of the structure can change in service.
- The numbers, types, placements, and frequency responses of sensors that measure the motions and actuators that control them are limited.
- Time available during service for characterization of the dynamics is limited.
- The dynamics are dominated by a resonant mode at low frequency.
- In-service measurements of the dynamics are supervised by a digital computer and are taken at a low rate of sampling, consistent with the low characteristic frequencies of the control system.
- The system must operate under little or no human supervision.

The method is based on extracting the desired model and control-design data from the response of the structure to known vibrational excitations (Fig. 6.20). Initially, wideband stochastic excitations are used to obtain the general characteristics of the structure. Narrow-band stochastic and piece-wise-constant (consistent with sample-and-hold discretizations) approximations to sinusoidal excitations are used to investigate specific frequency bands in more detail.

The relationships between the responses and excitations are first computed nonparametrically—by spectral estimation in the case of stochastic excitations and by estimation of gains and phases in the case of approximately sinusoidal excitations. In anticipation of the parametric curve fitting to follow, the order of a mathematical model of the dynamics of the structure is estimated by use of a *product moment matrix* (PMM). Next, the parameters of this model are identified by a least-squares fit of transfer-function coefficients to the nonparametric data. The fit is performed by an iterative reweighting technique to remove high-frequency emphasis and assure minimum-variance estimation of the transfer-function coefficient. The order of the model starts at the PMM estimate and is determined more precisely thereafter by successively adjusting a number of modes in the fit at each iteration until an adequately small output-error profile is observed.

In the analysis of the output error, the additive uncertainty is estimated to characterize the quality of the parametric estimate of the transfer function and for later use in the analysis and design of

robust control. It can be shown that if the additive uncertainty is smaller than a certain calculable quantity, then a conceptual control system could stabilize the model structure and could also stabilize the real structure. This criterion can be incorporated into an iterative design procedure. In this procedure, each controller in a sequence of controllers for the model structure would be designed to perform better than the previous one did, until the condition for robust capability was violated. Once the violation occurred, one could accept the penultimate design (if its performances were satisfactory) or continue the design process by increasing a robustness weighting (if available). In principle, convergence of this iterative process guarantees a control design that provides high performance for the model structure while guaranteeing robustness of stability to all perturbations of the structure within the additive uncertainty.

6.19 Optoelectronic Sensor Tracking Targets on a Structure

The location and exact position of a target can be accurately sensed through optoelectronic sensors for tracking a retroreflective target on a structure. An optoelectronic system simultaneously measures the positions of as many as 50 retroreflective targets within 35° of view with an accuracy of 0.1 mm. The system repeats the measurement 10 times per second. The system provides an unambiguous indication of the distance to each target that is not more than 75 m away from its sensor module. The system is called *spatial high-accuracy position-encoding sensor* (SHAPES).

SHAPES fills current needs in the areas of system identification and control of large flexible structures, such as large space- and ground-based antennas and elements of earth-orbiting observational platforms. It is also well-suited to applications in rendezvous and docking systems. Ground-based applications include boresight determination and precise pointing of 70-m deep-space-network antennas.

SHAPES illuminates the retroreflective targets by means of a set of lasers in its sensor module. In a typical application (Fig. 6.21) a laser diode illuminates each target with 30-ps pulses at a repetition rate of 100 MHz. Light reflected from the target is focused by a lens and passed through a beam splitter to form images on a charge-coupled device (CCD) and on the photocathode of a streak tube. The angular position of the target is determined simply from the position of its reflection on the charge-coupled device.

The measurement of the distance to the target is based on the round-trip time of the optical pulses. The round-trip distance can be measured in terms of the difference between the phase of the train of return pulses incident on the photocathode and the phase of a reference sine wave that drives the deflection plate of the streak tube.

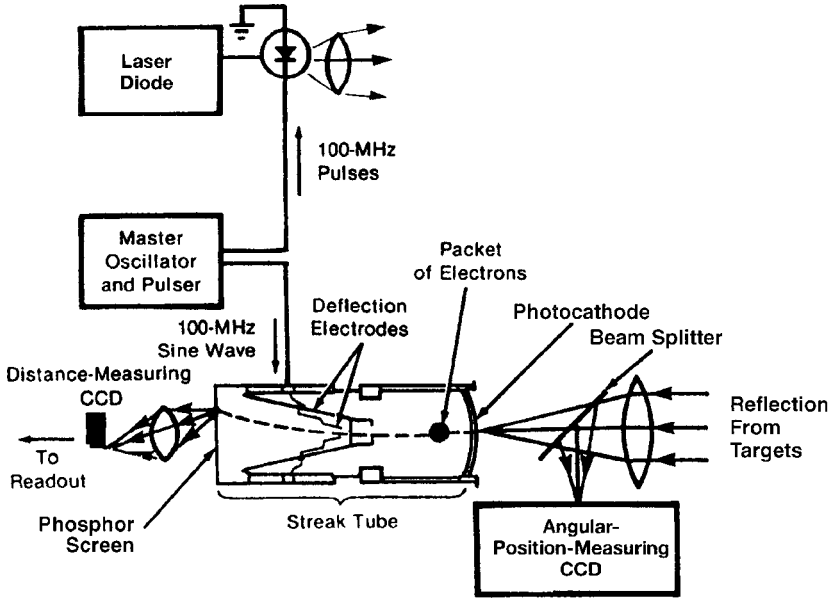


Figure 6.21 Beam splitter diverts reflections from a continuous-wave laser into a CCD camera for measurement of angles of reflection.

This difference, in turn, manifests itself as a displacement between the swept and unswept positions, at the output end of the streak tube, of the spot of light that represents the reflection from the target. The output of the streak tube is focused on a CCD for measurement and processing of the position of this spot. Three microprocessors control the operation of SHAPES and convert the raw data required from the angular-position and distance-measuring CCDs into the position of the target in three dimensions.

6.20 Optoelectronic Feedback Signals for Servomotors through Fiber Optics

In what is believed to be among its first uses to close a digital motor-control loop, fiber-optics transmission provides immunity to noise and rapid transmission of data. An optoelectronic system affects closed-loop control of the shaft angle of four servomotors and could be expanded to control as many as 16. The system includes a full-duplex fiber-optic link (Fig. 6.22) that carries feedforward and feedback digital signals over a distance of many meters, between commercial

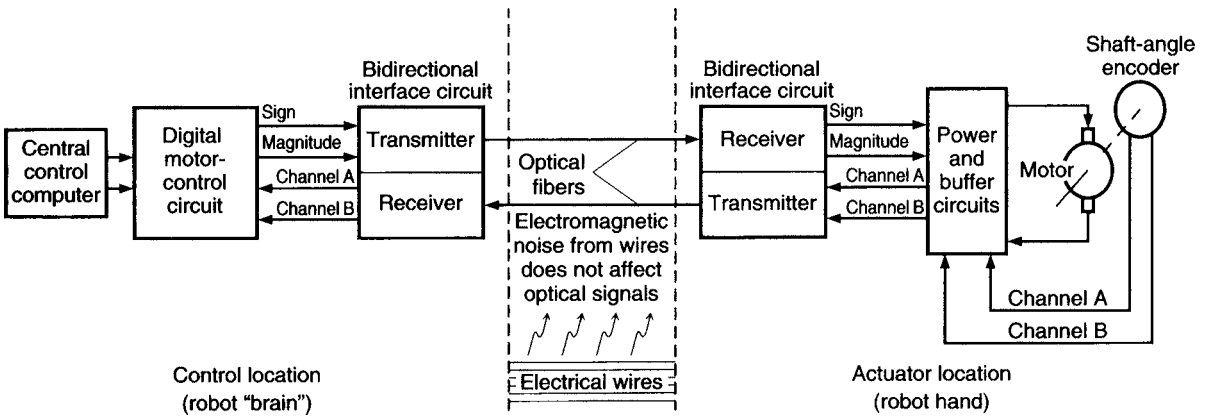


FIGURE 6.22 Full-duplex fiber-optic transmission link.

digital motor-control circuits that execute a PID control algorithm with programmable gain (one such control circuit dedicated to each servomotor) and modules that contain the motor-power switching circuits, digital-to analog buffer circuits for the feedforward control signals, and analog-to-digital buffer circuits for the feedback signals from the shaft-angle encoders (one such module located near, and dedicated to, each servomotor).

Besides being immune to noise, optical fibers are compact and flexible. These features are particularly advantageous in robots, which must often function in electromagnetically noisy environments and in which it would otherwise be necessary to use many stiff bulky wires (which could interfere with movement) to accommodate the required data rates.

Figure 6.22 shows schematically the fiber-optic link and major subsystems of the control loop of one servomotor. Each digital motor-control circuit is connected to a central control computer, which programs the controller gains and provides the high-level position commands. The other inputs to the motor-control circuit include the sign of the commanded motor current and pulse-width modulation representing the magnitude of the command motor current.

The fiber-optic link includes two optical fibers—one for feedforward, one for feedback. The ends of the fibers are connected to identical bidirectional interface circuit boards, each containing a transmitter and a receiver. The fiber-optic link has a throughput rate of 175 MHz; at this high rate, it functions as though it were a 32-bit parallel link (8 bits for each motor control loop), even though the data are multiplexed into a serial bit stream for transmission. In the receiver, the bit stream is decoded to reconstruct the 8-bit pattern and a programmable logic sequencer expands the 8-bit pattern to 32 bits and checks for errors by using synchronizing bits.

6.21 Acoustooptical/Electronic Sensors for Synthetic-Aperture Radar Utilizing Vision Technology

An acoustooptical sensor operates in conjunction with analog and digital electronic circuits to process frequency-modulated *synthetic-aperture radar* (SAR) return signals in real time. The acoustooptical SAR processor will provide real-time SAR imagery aboard moving aircraft or space SAR platforms. The acoustooptical SAR processor has the potential to replace the present all-electronic SAR processors that are currently so large and heavy and consume so much power that they are restricted to use on the ground in the postprocessing of the SAR in-flight data recorder.

The acoustooptical SAR processor uses the range delay to resolve the range coordinates of a target. The history of the phase of the train of radar pulses as the radar platform flies past a target is used to

obtain the azimuth (cross-range) coordinate by processing it coherently over several returns. The range-compression signal processing involves integration in space, while the azimuth-compression signal processing involves integration in time.

Figure 6.23 shows the optical and electronic subsystems that perform the space and time integrations. The radar return signal is heterodyned to the middle frequency of an acoustooptical sensor and added electronically to a reference sinusoid to capture the history of the phase of the return signal interferometrically for compression in azimuth. The resulting signal is applied to the acoustooptical sensor via a piezoelectric transducer. The acoustooptical sensor thus becomes a cell that encodes the evolving SAR return.

Meanwhile, pulses of light a few tens of nanoseconds long are generated by a laser diode in synch with the transmitted pulses and

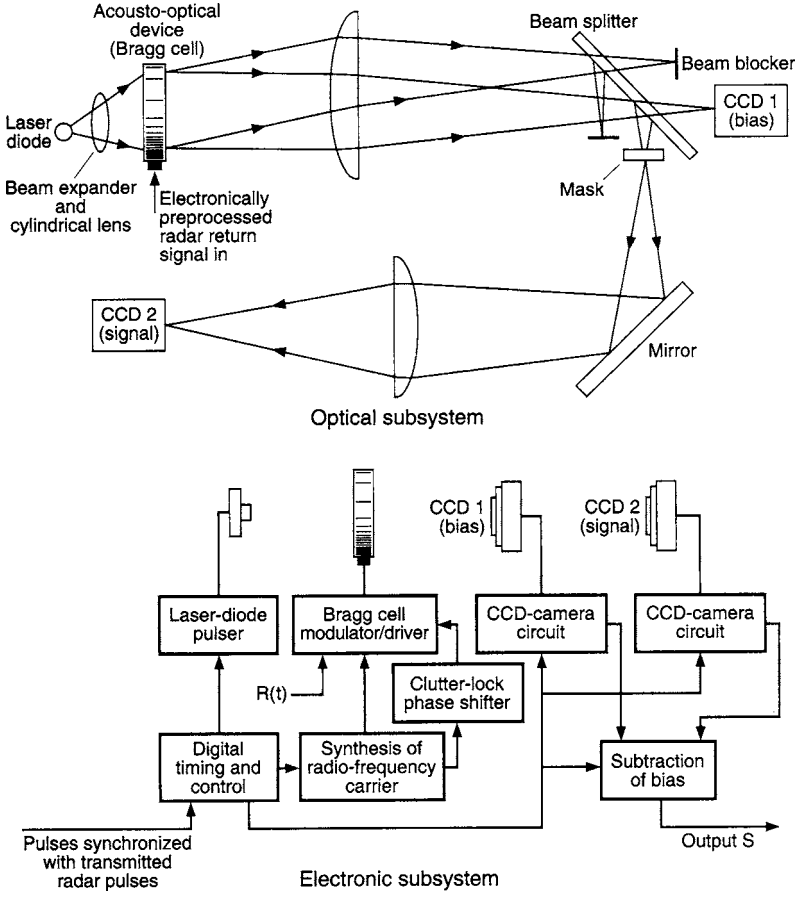


Figure 6.23 Acoustooptical synthetic-aperture radar.

are used to sample and process the return signal. Lenses shape that laser light into a plane wave incident upon the acoustooptical sensor. The integration in space is effected at the moment of sampling by the focusing action. The position of the focal point in the cell depends on the range delay of the corresponding target, and light is brought to focus on two CCD imaging arrays at positions that depend on the range.

The sinusoidal reference signal component of the cell interacts with laser radiation to generate a plane wave of light that interferes with the light focused by the cell. This produces interference fringes that encode the phase information in the range-compressed optical signal. These fringes are correlated with a mask that has a predetermined spatial distribution of density and that is placed in front of, or on, one of the CCD arrays. This CCD array is operated in a delay-and-integrate mode to obtain the desired correlation and integration in time for the azimuth compression. The output image is continuously taken from the bottom picture element of the CCD array.

Two CCDs are used to alleviate a large undesired bias of the image that occurs at the output as a result of optical processing. CCD_1 is used to compute this bias, which is then subtracted from the image of CCD_2 to obtain a better image.

6.22 The Use of Optoelectronic/Vision Associative Memory for High-Precision Image Display and Measurement

Storing an image of an object often requires large memory capacity and a high-speed interactive controller. Figure 6.24 shows schematically an optoelectronic associative memory that responds to an input

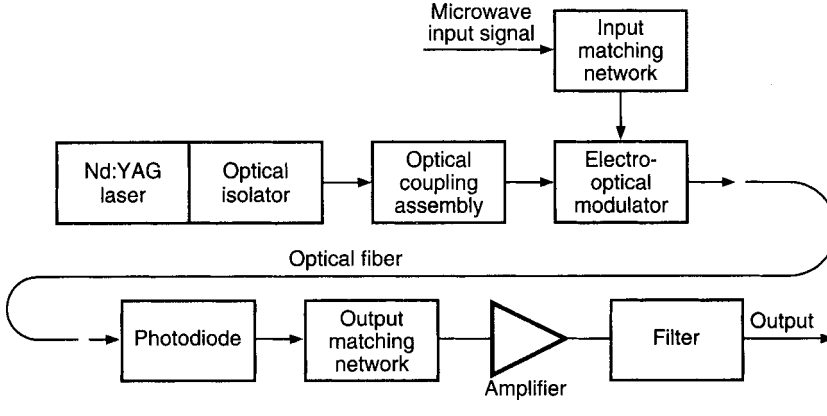


Figure 6.24 A developmental optoelectronic associative memory.

image by displaying one of *M* remembered images. The decision about which if any of the remembered images to display is made by an optoelectronic analog computation of an inner product-like measure of resemblance between the input image and each of the remembered images. Unlike associative memories implemented as all-electronic neural networks, this memory does not rely on the precomputation and storage of an outer-product synapse matrix. Instead, the optoelectronic equivalent of this matrix is realized by storing remembered images in two separate spatial light modulators placed in tandem. This scheme reduces the required size of the memory by an order of magnitude.

A partial input image is binarized and displayed on a liquid-crystal light valve spatial modulator which reprocesses the image in real time by operating in an edge-enhancement mode. This preprocessing increases the orthogonality (with respect to the inner product) between the input image and each of the remembered images, thereby increasing the ability of the memory to discriminate among different images.

The light from the input image is passed through a polarizing beam splitter, a lens, a binary diffraction grating, and another lens, to focus an array of *M* replicas of the input image on one face of a liquid-crystal-television spatial light modulator that is displaying the *M* remembered images. The position of each replica of the input image coincides with that of one of the remembered images. Light from the array of pairs of overlapping input and remembered images is focused by a corresponding array of lenslets onto a corresponding array of photodetectors. The intensity of light falling on each photodetector is proportional to the inner product between the input image and the corresponding remembered image.

The outputs of the photodetectors are processed through operational amplifiers that respond nonlinearly to inner product level (in effect executing analog threshold functions). The outputs of the amplifiers drive point sources of white light, and an array of lenslets concentrates the light from each source onto the spot occupied by one of *M* remembered images displayed on another liquid-crystal-television spatial light modulator. The light that passes through this array is reflected by a pivoted ray of mirrors through a lens, which focuses the output image onto a CCD television camera. The output image consists of superpositioned remembered images, the brightest of which are those that represent the greatest inner products (the greatest resemblance to the input image). The television camera feeds the output image to a control computer, which performs a threshold computation, then feeds the images through a cathode-ray tube back to the input liquid-crystal light valve. This completes the associative recall loop. The loop operates iteratively until one (if any) of the remembered images is the sole output image.

6.23 Sensors for Hand-Eye Coordination of Microrobotic Motion Utilizing Vision Technology

The micro motion of a robotic manipulator can be controlled with the help of dual feedback by a new method that reduces position errors by an order of magnitude. The errors—typically of the order of centimeters—are differences between real positions on the one hand and measured and computed positions on the other; these errors arise from several sources in the robotic actuators and sensors and in the kinematic model used in control computations. In comparison with current manufacturing methods of controlling the motion of a robot with visual feedback (the robotic equivalent of hand-eye coordination), the novel method requires neither calibration over the entire work space nor the use of an absolute reference coordinate frame for computing transformations between field of view and robot joint coordinates.

The robotic vision subsystem includes five cameras: three stationary ones that provide wide-angle views of the workspace and two mounted on the wrist of an auxiliary robot arm to provide stereoscopic close-up views of the workspace near the manipulator (Fig. 6.25).

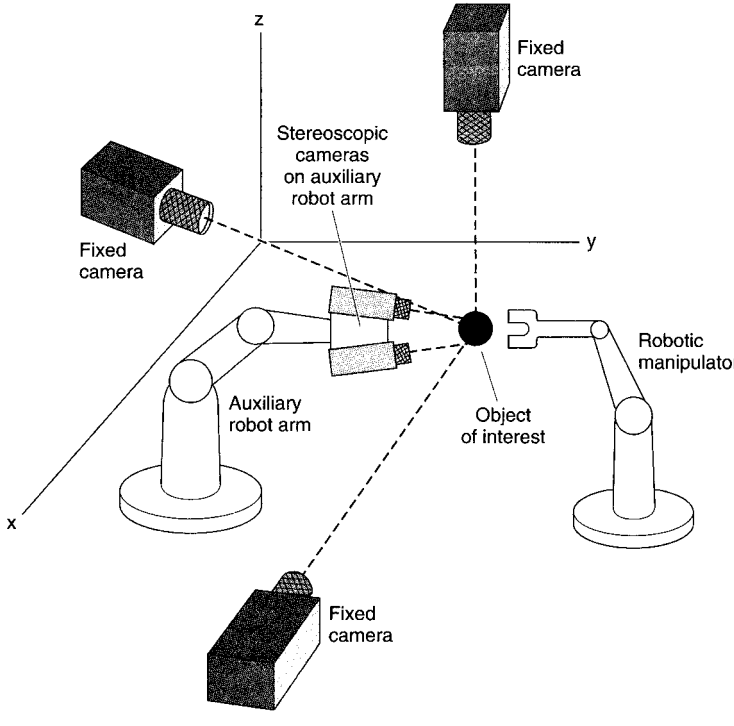


Figure 6.25 Stereoscopic cameras on an auxiliary robot arm.

The vision subsystem is assumed to be able to recognize the objects to be avoided and manipulated and to generate data on the coordinates of the objects from sent positions in the field-of-view reference frame.

The new method can be implemented in two steps:

1. The close-up stereoscopic cameras are set initially to view a small region that contains an object of interest. The end effector is commanded to move to a nominal position near the object and within the field of view. Typically, the manipulator stops at a slightly different position, which is measured by the cameras. Then, the measured error in position is used to compute a small corrective motion. This procedure is designed to exploit the fact that small errors in relative position can be measured accurately and small relative motions can be commanded accurately.
2. The approximate direct mapping between the visual coordinates and the manipulator joint-angle coordinates can be designed without intermediate transformation to and from absolute coordinates. This is, in effect, a calibration, but it requires fewer points than does a conventional calibration in an absolute reference frame over the entire work space. The calibration is performed by measuring the position of a target (in field-of-view coordinates) when the target is held rigidly by the manipulator at various commanded positions (in manipulator joint-angle coordinates) and when the cameras are placed at various commanded positions. Interpolations and extrapolations to positions near the calibration points are thereafter performed by use of the nonlinear kinematic transformations.

6.24 Force and Optical Sensors Controlling Robotic Grippers for Agriculture and Manufacturing Applications

A robotic gripper operates in several modes to locate, measure, recognize (in a primitive sense), and manipulate objects in an assembly subsystem of a robotic cell that is intended to handle geranium cuttings in a commercial greenhouse. The basic concept and design of the gripper could be modified for handling other objects—for example, rods or nuts—including sorting the objects according to size. The concept is also applicable to real-time measurement of the size of an expanding or contracting part gripped by a constant force and to measurement of the size of a compliant part as a function of the applied gripping force.

The gripper is mounted on an industrial robot. The robot positions the gripper at a fixed distance above the cutting to be processed.

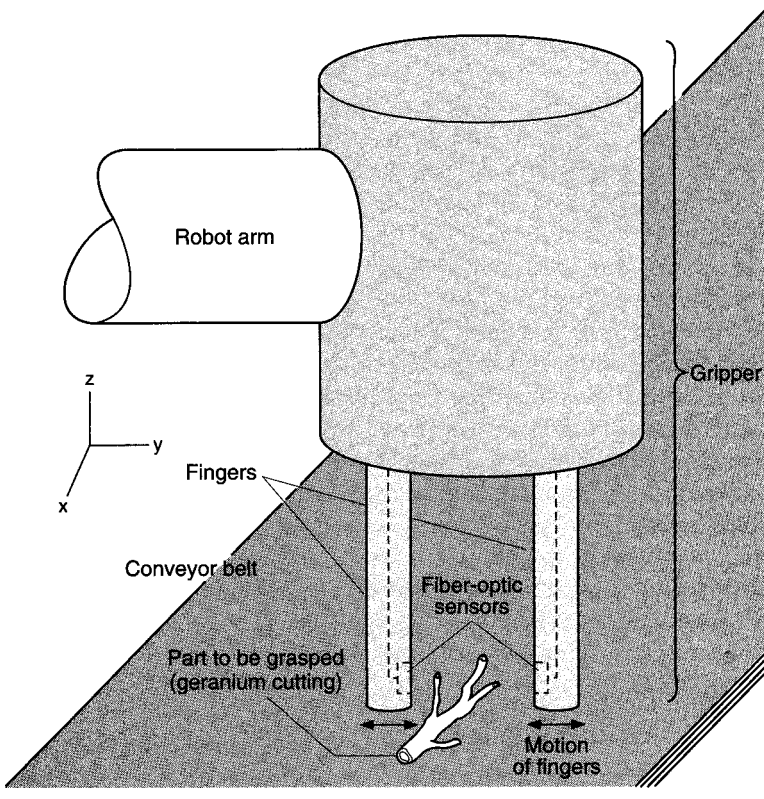


Figure 6.26 Robotic gripper for geranium cutting.

A vision system locates the cutting in the *x-y* plane lying on a conveyer belt (Fig. 6.26).

The robot uses fiber-optic sensors in the fingertip of the gripper to locate the cutting along the axis. The gripper grasps the cutting under closed-loop digital servo force control. The size (that is, the diameter of the stem) of the cutting is determined from the finger position feedback, while the cutting is being grasped under force control. The robot transports the cutting to a scale for weighing, to a trimming station, and finally to a potting station. In this manner, cuttings are sorted according to weight, length, and diameter.

The control subsystem includes a 32-bit minicomputer that processes the vision information and collects system grating data. The minicomputer communicates with a robot controller and the gripper node control. The gripper node control communicates with the scale. The robot controller communicates with the gripper node controller via discrete input/output triggering of each of the gripper functions.

The gripper subsystem includes a PC/AT-compatible industrial computer; a gripper mechanism, actuated by two DC servomotors, with an integrated load cell; discrete input/output components; and two fiber-optic analog-output distance sensors. The computer includes a discrete input/output circuit card, an 8-channel A/D converter circuit card, a motor-control circuit card with A/D components, two serial ports, and a 286 processor with coprocessor.

A geranium cutting comprises a main stem and several petioles (tiny stem leaves). The individual outputs from the fiber-optic sensors can be processed into an indication of whether a stem or a petiole is coming into view as the gripper encounters the cutting. Consequently, the gripper can be commanded to grasp a stem but not a petiole. The axial centerline of a stem can also be recognized from the outputs of the five optic sensors. Upon recognition of a centerline, the gripper signals the robot, and the robot commands the gripper to close.

The motor-controller circuit card supplies the command signals to the amplifier that drives the gripper motors. This card can be operated as a position control with digital position feedback or as a force control with analog force feedback from the load cell mounted in the gripper. A microprocessor is located on the motor control card. Buffered command programs are downloaded from the computer to this card for independent execution by the card.

Prior to a controlled force closure, the motor-control card controls the gripper in position-servo mode until a specified force threshold is sensed, indicating contact with the cutting. Thereafter, the position-servo loop is opened, and the command signal to the amplifier is calculated as the difference between the force set point and the force feedback from the load cell. This distance is multiplied by a programmable gain value, then pulse-width-modulated with a programmable duty cycle of typically 200 percent. This technique provides integral stability to the force-control loop. The force-control loop is bidirectional in the sense that, if the cutting expands between the fingertips, the fingers are made to separate and, if the cutting contracts, the fingers are made to approach each other.

6.25 Ultrasonic Stress Sensor Measuring Dynamic Changes in Materials

An *ultrasonic dynamic vector stress sensor* (UDVSS) has recently been developed to measure the changes in dynamic directional stress that occur in materials or structures at the location touched by the device when the material or structure is subjected to cyclic load. A strain gauge device previously used for the measurement of such a stress measured strain in itself, not in the part being stressed, and thus provided a secondary measurement. Other techniques, such as those that

involve thermoelasticity and shearography, have been expensive and placed demands on the measured material. The optical measurement of stress required the application of a phase coat to the object under test. The laser diffraction method required notching or sharp marking of the specimen.

A UDVSS is the first simple portable device able to determine stress directly in the specimen itself rather than in a bonded gauge attached to the specimen. As illustrated in Fig. 6.27, a typical material testing machine applies cyclic stress to a specimen. The UDVSS includes a probe, which is placed in contact with the specimen; an electronic system connected to the probe; and a source of a reference signal. The probe assembly includes a probe handle that holds the probe, a transducer mount that contains the active ultrasonic driver and receiver, an ultrasonic waveguide transmitter and ultrasonic waveguide receiver that convert the electrical signals to mechanical motion and the inverse, and a cable that connects the probe of the electronics. When in contact with the specimen, the ultrasonic waveguide transmitter causes acoustic waves to travel across the specimen to the ultrasonic waveguide receiver, wherein the wave is converted to an electrical signal.

The operation of the UDVSS is based on the physical phenomenon that the propagation of sound in the specimen changes when the stress in the specimen changes. A pulse phase-locked loop reacts to a change in propagation of sound and therefore in stress by changing

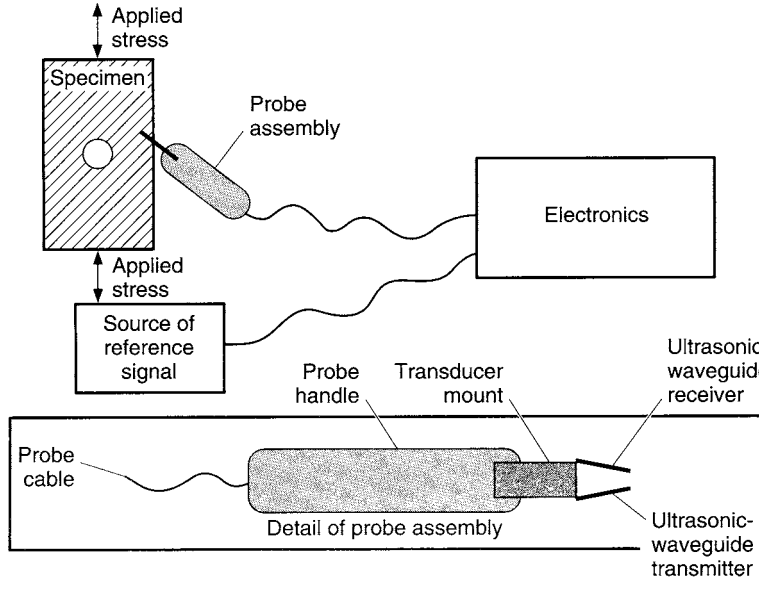


Figure 6.27 Ultrasonic dynamic stress sensor.

its operational frequency. The component of that signal represents that change in voltage needed to keep the system at quadrature to follow the system change in stress. That signal provides the information on changing stress.

The UDVSS can be moved around on the specimen to map out the stress field, and by rotating the probe, one can determine the direction of a stress. In addition, the probe is easily calibrated. The UDVSS should find wide acceptance among manufacturers of aerospace and automotive structures for stress testing and evaluation of designs.

6.26 Predictive Monitoring Sensors Serving the CIM Strategy

Computer-integrated manufacturing technology can be well-served by a predictive monitoring system that would prevent a large number of sensors from overwhelming the electronic data monitoring system or a human operator. The essence of the method is to select only a few of the many sensors in the system for monitoring at a given time and to set alarm levels of the selected sensor outputs to reflect the limit of expected normal operation at the given time. The method is intended for use in a highly instrumented system that includes many interfacing components and subsystems—for example, an advanced aircraft, an environmental chamber, a chemical processing plant, or a machining work cell.

Several considerations motivate the expanding effort in implementing the concept of predictive monitoring. Typically, the timely detection of anomalous behavior of a system and the ability of the operator or electronic monitor to react quickly are necessary for the continuous safe operation of the system.

In the absence of a sensor-planning method, an operator may be overwhelmed with alarm data resulting from interactions among sensors rather than data directly resulting from anomalous behavior of the system. In addition, much raw sensor data presented to the operator may by irrelevant to an anomalous condition. The operator is thus presented with a great deal of unfocused sensor information, from which it may be impossible to form a global picture of events and conditions in the system. The predictive monitoring method would be implemented in a computer system running artificial intelligence software, tentatively named PREMON. The predictive monitoring system would include three modules: (1) a causal simulator, (2) a sensor planner, and (3) a sensor interpreter (Fig. 6.28).

The word *event* in Fig. 6.28 denotes a discontinuous change in the value of a given quantity (sensor output) at a given time. The inputs to the causal simulator would include a causal mathematical model of the system to be monitored, a set of events that describe the initial

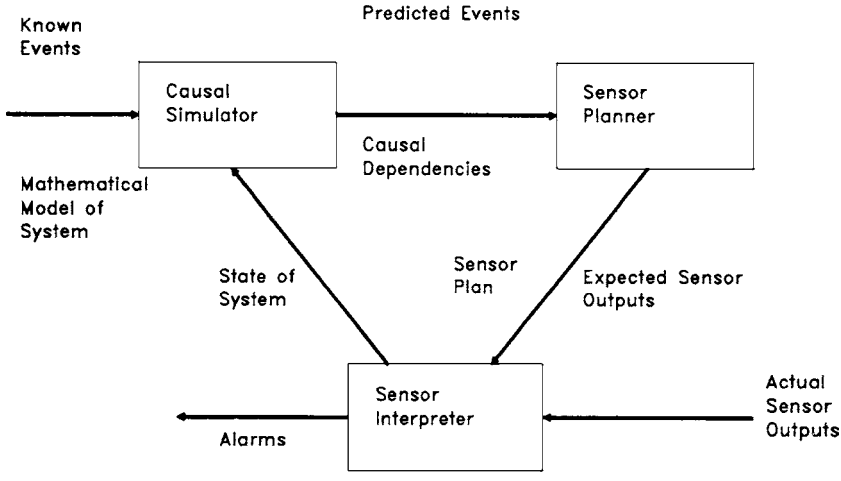


Figure 6.28 Predictive monitoring system would concentrate on the few sensor outputs that are most important at the moment.

state of the system, and perhaps some future scheduled events. The outputs of the causal simulator would include a set of predicted events and a graph of causal dependency among events.

The sensor planner would use the causal dependency graph generated by the causal simulator to determine which few of all the predicted events are important enough that they should be verified. In many cases, the most important events would be taken to be those that either caused, or are caused by, the greatest number of other events. This notion of causal importance would serve as the basis for the election of those sensors, the outputs of which should be used to verify the expected behavior of the system.

The sensor interpreter would compare the actual outputs of the selected sensors with the values of those outputs predicted by the causal simulator. Alarms would be raised where the discrepancies between predicted and actual values were significant.

6.27 Reflective Strip Imaging Camera Sensor—Measuring a 180°-Wide Angle

A proposed camera sensor would image a thin striplike portion of a field of view 180° wide. For example, it could be oriented to look at the horizon in an easterly direction from north to south or it could be rotated about a horizontal north/south axis to make a "pushbroom" scan of the entire sky proceeding from the easterly toward the westerly half of the horizon. Potential uses for the camera sensor include surveillance of clouds, coarse mapping of terrain, measurements of

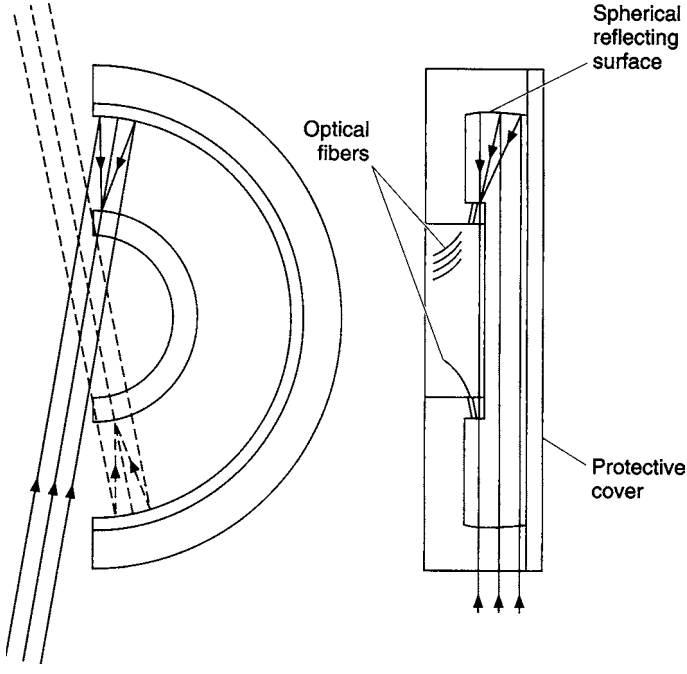


Figure 6.29 A thin segment of a hemispherical concave reflector would form an image from a 180° strip field of view onto optical fibers.

the bidirectional reflectance distribution functions of aerosols, imaging spectrometry, oceanography, and exploration of the planets.

The imaging optics would be a segment of concave hemispherical reflecting surfaces placed slightly off center (Fig. 6.29). Like other reflecting optics, it would be achromatic. The unique optical configuration would practically eliminate geometric distortion of the image. The optical structure could be fabricated and athermalized fairly easily in that it could be machined out of one or a few pieces of metal, and the spherical reflecting surface could be finished by diamond turning. In comparison, a camera sensor with a fish-eye lens, which provides a nearly hemispherical field of view, exhibits distortion, chromatism, and poor athermalization. The image would be formed on a thin semicircular strip at half the radius of the sphere. A coherent bundle of optical fibers would collect the light from this strip and transfer the image to a linear or rectangular array of photodetectors or to the entrance slit of an image spectrograph. Provided that the input ends of the fibers were properly aimed, the cones of acceptance of the fibers would act as aperture stops. Typically, the resulting width of the effective aperture of the camera sensor would be about one-third the focal length ($f/3$).

The camera sensor would operate at wavelengths from 500 to 1100 nm. The angular resolution would be about 0.5°. In the case of an effective aperture of $f/3$, the camera would provide an unvignetted view over the middle 161° of the strip, with up to 50 percent vignetting in the outermost 9.5° on each end.

The decentration of the spherical reflecting surface is necessary to make room for the optical fibers and the structure that would support them. On the other hand, the decentration distance must not exceed the amount beyond which the coma that results from decentration would become unacceptably large. In the case of an effective aperture of $f/3$, the coma would be only slightly in excess of the spherical aberration if the decentration were limited to about $f/6$. This would be enough to accommodate the fibers and supporting structure.

6.28 Optical Sensor Quantifying Acidity of Solutions

With environmental concerns increasing, a method for taking effective measurements of acidity will minimize waste and reduce both the cost and the environmental impact of processing chemicals. Scientists at Los Alamos National Laboratory (LANL) have developed an optical sensor that measures acid concentration at a higher level than does any current device. The optical high-acidity sensor, reduces the wear generated from acidity measurements and makes possible the economic recycling of acid waste.

The high-acidity sensor (Fig. 6.30) consists of a flow cell (about 87.5 mm across) in which two fused silica lenses are tightly mounted across from one another. Fiber-optic cables connect the two lenses to a spectrophotometer. One lens is coated with a sensing material consisting of a polymer that is chemically bound to the lens and an indicator that is physically entrapped with the polymer. Acidic solutions

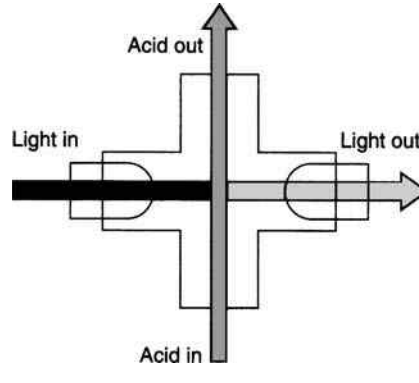


Figure 6.30 Optical high-acidity sensor.

flow vertically up through the flow cell. The light from a fixed source in the spectrophotometer is collimated by the coated lens, passes through the acidic solution, and then is focused by the second lens.

The amount of light absorbed by the indicator depends on the acidity of the solution. The absorption spectrum of the indicator reflects the concentration of hydrogen ions in the solution, which is how acidity is measured.

The LANL sensor's sensitivity range—acid concentrations of 4 to 12 molar—is unmatched by any other acidity measurement technique in current manufacturing. Present techniques are unsuitable for measuring solutions with a negative pH or for measuring chemical processes. With the LANL sensor, high acid concentrations can be measured in 3 min or less, which is 50 times faster than measuring with manual sampling and titration. The sensing material can function in highly acidic solutions for four to six months with calibration less than once a week, and needs to be replaced only once or twice a year. A sensor using similar principles has recently been developed from measuring lower acidity concentrations.

The sensor is selective in its measurements because the small pores of the polymer allow only hydrogen ions to pass through the indicator. Metal ions do not form a complex with the indicator under acidic solutions. The sensor is reusable—that is, chemically reversible. In short, no comparable product for measuring high acidity is reversible or as sensitive, fast, accurate, and selective as the LANL optical high-acidity sensor.

The prime use for the sensor is monitoring highly acidic chemical processes and waste solutions. High-acidity processes are common in preparing metals from ore-mining operations, treating fuel elements from nuclear power plants, manufacturing bulk acid, and metal finishing including passivation and electroplating.

For highly acidic processor applications, the LANL sensor will save companies thousands of dollars by improving efficiency and decreasing the time devoted to acid measurements. The sensor can be used on manufacturing lines, allowing control and waste management adjustment to be made before acidity fluctuations become a problem. The sensor will improve efficiency at least 25 percent by eliminating the need to reprocess material processed incorrectly on the first try because its true acidity was not known and controlled. Higher efficiency will mean lower cost and minimal waste products generated; the sensor itself generates no caustic waste. Finally, the sensor's waste monitoring capabilities will help ensure that any discharged waste is environmentally benign.

For the 500 acidity measurements done at LANL in 1991, the sensor saved $99,500, a 99.5 percent savings in labor costs in the first year. And, because the sensor generates no waste, 20 L a year of caustic waste was avoided, a 100 percent reduction.

6.29 Sensors for Biomedical Technology

In recent years, advanced imaging and other computer-related technology have greatly expanded the horizons of basic biological and biochemical research. Currently, such drivers as the growing needs for environmental information and increased understanding of genetic systems have provided impetus to biotechnology development. This is a relatively new specialty in the marketplace; nevertheless, the intensity of worldwide competition is escalating. Collaborative research and development projects among the U.S. government, industry, and academia constitute a major thrust for rapid deployment of research and development. The results can place the nation in a position of world leadership in biotechnology.

6.29.1 Sensor for Detecting Minute Quantities of Biological Materials

A new device based on laser-excited fluorescence provides unparalleled detection of biological materials for which only minuscule samples may be available. This device, invented at the Ames Laboratories, received a 1991 R&D 100 award.

The Ames Microfluor detector was developed to meet a need for an improved detection technique, driven by important new studies of the human genome, abuse substances, toxins, DNA adduct formation, and amino acids, all of which may be available only in minute amounts. Although powerful and efficient methods have been developed for separating biological mixtures in small volumes (i.e., capillary electrophoresis), equally powerful techniques for subsequent detection and identification of these mixtures have been lacking.

The Microfluor detector combines very high sensitivities with the ability to analyze very small volumes. The instrument design is based on the principle that many important biomaterials are fluorescent, while many other biomaterials, such as peptides and oligonucleotides, can be made to fluoresce by adding a fluorescent tag.

When a sample-filled capillary tube is inserted into the Microfluor detector and is irradiated by a laser beam, the sample will fluoresce. The detector detects, monitors, and quantifies the contents by sensing the intensity of the fluorescent light emitted. The signal is proportional to the concentration of the materials. The proportionality constant is characteristic of the material itself.

Analyses can be performed with sample sizes 50 times smaller than those required by other methods, and concentrations as low as 10^{-11} molar (1 part per trillion) can be measured. Often, the critical components in a sample are present at these minute concentrations. These two features make the Microfluor detector uniquely compatible with capillary electrophoresis. In addition, the Ames-developed detector is distinct from other laser-excited detectors in that it is not

seriously affected by stray light from the laser itself; it also allows simple alignment and operation in full room light. The Microfluor detector has already been used to determine the extent of adduct formation and base modification in DNA so that the effects of carcinogens on living cells can be studied. Future uses of the sensor will include DNA sequencing and protein sequencing. With direct or indirect fluorescence detection, researchers are using this technique to study the chemical contents of individual living cells. This capability may allow pharmaceutical products to be tested on single cells rather than a whole organism, with improved speed and safety.

6.29.2 Sensors for Early Detection and Treatment of Lung Tumors

A quick accurate method for early detection of lung cancer would raise chances of patient survival from less than 50 percent to 80 or 90 percent. Until now, small cancer cells deep in the lung have been impossible to detect before they form tumors large enough to show up in x-rays. Researchers at Los Alamos National Laboratory in collaboration with other institutions, including the Johns Hopkins school of medicine and St. Mary's Hospital in Grand Junction, Colorado, are developing methods for finding and treating lung cancer in its earliest stages.

A detection sensor involves a porphyrin, one of an unusual family of chemicals that is found naturally in the body and concentrates in cancer cells. The chemical is added to a sample of sputum coughed up from the lung. When exposed to ultraviolet or laser light, cells in the porphyrin-treated sputum glow a bright red. When the sample is viewed under the microscope, the amount and intensity of fluorescence in the cells determines the presence of cancer.

The first clinical test of a detection technique using porphyrin was done by LANL and St. Mary's Hospital in 1988. Four different porphyrins were tested on sputum samples from two former miners, one known to have lung cancer and one with no detectable cancer. One of the porphyrins was concentrated in certain cells only in the sputum of the miner with lung cancer. Other tests concluded that these were cancer cells. Later, a blind study of sputum samples from 12 patients, 8 of whom had lung cancer in various stages of development, identified all the cancer patients as well as a ninth originally thought to be free of cancer. Further tests showed that this patient also had lung cancer.

Identifying the ninth patient prompted a new study in which the procedure was evaluated for its ability to detect precancerous cells in the lung. In this study, historical sputum samples obtained from Johns Hopkins were treated with the porphyrin. Precancerous conditions that chest x-rays had not detected were identified in samples from patients who later developed the disease. Although further

testing is needed, researchers now feel confident that the technique and the detection sensor can detect precancerous conditions three to four years before onset of the disease.

Working in collaboration with industry, researchers expect to develop an instrument in several years for rapid automated computerized screening of larger populations of smokers, miners, and other people at risk of developing lung cancer. Because lung cancer is the leading cancer killer in both men and women, a successful screening procedure would dramatically lower the nation's mortality rate from the disease. A similar procedure has potential for the analysis of Pap smears, now done by technicians who screen slides one at a time and can misread them.

In addition to a successful screening program, LANL hopes to develop an effective treatment for early lung cancer. The National Cancer Institute will help investigate and implement both diagnostic and therapeutic programs. Researchers are performing basic cell studies on animals to investigate the effect of porphyrin on lung cancer cells. They are also exploring the use of the LANL-designed porphyrin attached to radioactive copper to kill early cancer cells in the lung in a single search-and-destroy mission. The porphyrin not only seeks out cancer cells but also has a molecular structure, similar to a hollow golf ball, that can take certain metals into its center. A small amount of radioactive copper 67 placed inside the porphyrin should destroy tumors the size of pinhead, as well as function as a tracer.

6.29.3 Ultrasensitive Sensors for Single-Molecule Detection

Enhancing the sensitivity of research instruments has long been a goal of physicists, biologists, and chemists. With the single-molecule detector, researchers have achieved the ultimate in this pursuit: detection of a single fluorescent molecule in a liquid. The new instrument—a thousand times more sensitive than existing commercial detectors—brings new capabilities to areas of science and technology that affect lives in many ways, from DNA sequencing, biochemical analysis, and virus studies to identifying environmental pollutants.

For some time, scientists have observed individual molecules trapped in a vacuum, where they are isolated and relatively easy to find. However, many important biological and chemical processes occur in a liquid environment, where many billions of other molecules surround the molecules of interest. Developing a single-molecule detector that operates in such an environment presented a difficult challenge.

The observation technique involves attaching fluorescent dye molecules to the molecules of interest and then exciting the dye molecule by passing them in solution to a rapidly pulsed laser beam and detecting the subsequent faint light, or photons, they emit. The fluorescent lifetimes of the molecules are much shorter than the time the

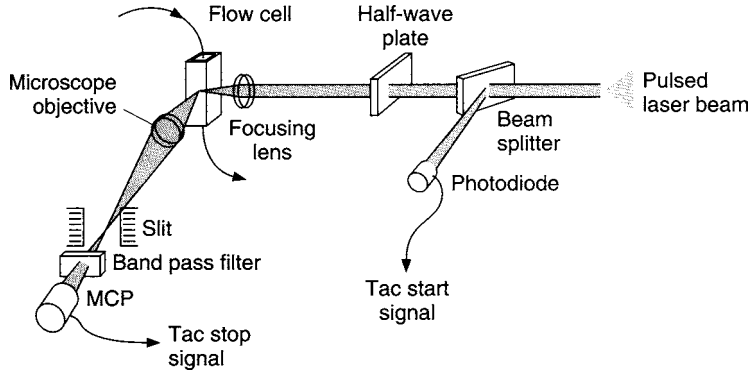


Figure 6.31 Single-molecule detector for individual fluorescent molecules in solution.

molecules spend in the laser beam; therefore each molecule is reexcited many times and yields many fluorescent photons. The signature of the passing molecule is the burst of photons that occurs when the molecule is passing the laser beam (Fig. 6.31).

A lens, or microscope objective, and a slit are arranged to image the photons from a small region around the laser beam waist onto a microchannel plate photomultiplier (MCPP) that counts individual photons. The intense excitation light from the laser is blocked from reaching the MCPP by a bandpass spectral filter, which is centered near the peak fluorescent wavelength. The excitation light consists of extremely short pulses, each about 70 trillionths of a second. The dye molecule does not emit light until a billionth of a second after excitation, so the flash of laser light fades before the feeble molecular glow occurs. For reliable identification of individual molecules, the technique maximizes the number of the detected photons and minimizes the number of background photons. Although some background remains, the technique registers over 85 percent of the fluorescent molecules.

Developed to aid in sequencing chromosomes for the Human Genome Project, the sensor detector's high activity would allow DNA sequencing rates hundreds of times faster than those obtainable with present techniques. It would also eliminate the need for radioactive materials, gels, and electrophoresis solutions, which often create disposal problems, and it is expected to help make DNA sequencing a routine diagnostic and clinical tool. One eventual benefit of the DNA research may be rapid screening for any of 3500 known genetic diseases such as diabetes, cystic fibrosis, and Alzheimer's disease.

The ultrasensitive detector can be used to find and quantify minute amounts of chemicals, enzymes, and viruses in the blood and monitor the dispersal of extremely low concentrations of environmental pollutants. The device may also be useful in studying the interaction of viruses and their binding sites. It may be possible to

develop a procedure for rapidly evaluating the efficiency of vaccines; such a procedure could quickly expedite the search for an effective vaccine against the AIDS virus.

Further Reading

Barks, R. E., "Optical High Acidity Sensor," Los Alamos National Laboratory, June 1991.

Bicknel, T. J., and W. H. Farr, "Acousto Optical/Electronic Processor for SAR," *NASA Tech Briefs*, 16, May 1992.

Bonnert, R., "Design of High Performance Digital Tachometer with Digital Microcontroller," *IEEE Trans. Instrum.*, 38, 1104–1108 (1989).

Buser, R. A., and N. F. Rooij, "Resonant Silicon Structures," *Sensors and Actuators*, 17, 145–154 (1989).

D'Amico A., and E. Verona, "SAW Sensors," *Sensors and Actuators*, 17, 66–66 (1989).

Fleming Dias, J., "Physics Sensors Using SAW Devices," *Hewlett-Packard J.*, 18–20, December 1981.

Gast, T., "Sensors with Oscillating Elements," *J. Phys. E. Sci. Instrum.*, 18, 783–789 (1985).

Hayman, J. S., and M. Frogatt, "Ultrasonic Dynamic Vector Stress Sensor (UDVSS)," Langley Research Center, 1992.

Hewlett-Packard, "Design and Operational Considerations for the HELDS-5000 Incremental Shaft Encoder," Application Note 1011, Palo Alto, 1981.

Higbie, B. N., "Automatic Detection of Faults in Turbomachinery Bearings," NASA Tech Briefs, 16, May 1992.

Higbie, N., Technology Integration and Development Group Inc., *Technical Report*, May 1992.

Huner, B., P. Klinkhachorn, and E. B. Everton, "Hybrid Clock Oscillator Modules as Deposition Monitors," *Rev. Sci. Instrum.*, 59, 983–986 (1988).

Ito, H., "Balanced Absorption Quartz Hygrometer," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, 34, 136–141 (1987).

Lokshin, A. M., "Hand/Eye Coordination for Microrobotic Motion—Utilizing Vision Technology," Caltech, 1991.

Los Alamos National Laboratory, Research Team, "Single Molecule Detection," January 1991.

Montgomery, J. L., "Force and Optical Sensors Controlling Robotic Gripper for Agriculture," Martin Marietta Corp., 1992.

Noble, M. N., D. N. Mark, and R. Blue, "Tracking Retroreflective Targets on a Structure," *NASA Tech Briefs*, 16, May 1992.

"Optical Sensor to Quantify Highly Acidic Solution—Sensing High Acidity without Generating Waste," LANL, *Technology '91*, 57–58 (1991).

Tien-Hsin Chao, "Experimental Optoelectronic Associative Memory," *NASA Tech Briefs*, 16, May 1992.

"Ultrasensitive Detection for Medical and Environmental Analysis—Single Molecule Detector," LANL, *Technology '91*, 80–81 (1991).

Vaughan, A. H., "Reflective Strip Imaging Camera Sensor—Measuring a 180° Wide Angle," Caltech, NASA's Jet Propulsion Laboratory, 1992.

Williams, D. E., "Laser Sensor Detecting Microfluor," Ames Laboratory, July 1991.

This page intentionally left blank

CHAPTER 7
Industrial Sensors and Control

7.1 Introduction

Current manufacturing strategy defines manufacturing systems in terms of sensors, actuators, effectors, controllers, and control loops. Sensors provide a means for gathering information on manufacturing operations and processes being performed. In many instances, sensors are used to transform a physical stimulus into an electrical signal that may be analyzed by the manufacturing system and used for making decisions about the operations being conducted. Actuators convert an electrical signal into a mechanical motion. An actuator acts on the product and equipment through an effector. Effectors serve as the "hand" that achieves the desired mechanical action. Controllers are computers of some type that receive information from sensors and from internal programming, and use this information to operate the manufacturing equipment (to the extent available, depending on the degree of automation and control). Controllers provide electronic commands that convert an electrical signal into a mechanical action. Sensors, actuators, effectors, and controllers are linked to a control loop.

In limited-capability control loops, little information is gathered, little decision making can take place, and limited action results. In other settings, "smart" manufacturing equipment with a wide range of sensor types can apply numerous actuators and effectors to achieve a wide range of automated actions.

The purpose of sensors is to inspect work in progress, to monitor the work-in-progress interface with the manufacturing equipment, and to allow self-monitoring of manufacturing by the manufacturing system's own computer. The purpose of the actuator and effector is to transform the work in progress according to the defined processes of the manufacturing system. The function of the controller is to allow for varying degrees of manual, semiautomated, or fully automated control over the processes. In a fully automated case, such as in computer-integrated manufacturing, the controller is completely adaptive and

functions in a closed-loop manner to produce automatic system operation. In other cases, human activity is involved in the control loop.

In order to understand the ways in which the physical properties of a manufacturing system affect the functional parameters associated with the manufacturing system, and in order to determine the types of physical manufacturing system properties that are necessary to implement the various desired functional parameters, it is important to understand the technologies available for manufacturing systems that use automation and integration to varying degrees.

The least automated equipment makes use of detailed operator control over all equipment functions. Further, each action performed by the equipment is individually directed by the operator. Manual equipment thus makes the maximum use of human capability and adaptability. Visual observations can be enhanced by the use of microscopes and cameras, and the actions undertaken can be enhanced by the use of simple effectors. The linkages between the sensory information (from microscopes or through cameras) and the resulting actions are obtained by placing the operator in the loop.

This type of system is clearly limited by the types of sensors used and their relationship to the human operator, the types of the effectors that can be used in conjunction with the human operator, and the capabilities of the operator. The manufacturing equipment that is designed for a manual strategy must be matched to human capabilities. The human–manufacturing equipment interface is extremely important in many manufacturing applications. Unfortunately, equipment design is often not optimized as a sensor-operator-actuator/effector control loop.

A manufacturing system may be semiautomated, with some portion of the control loop replaced by a computer. This approach will serve the new demands on manufacturing system design requirements. Specifically, sensors now must provide continuous input data for both the operator and computer. The appropriate types of data must be provided in a timely manner to each of these control loops. Semiautomated manufacturing systems must have the capability for a limited degree of self-monitoring and control associated with the computer portion of the decision making loop. An obvious difficulty in designing such equipment is to manage the computer- and operator-controlled activities in an optimum manner. The computer must be able to recognize when it needs operator support, and the operator must be able to recognize which functions may appropriately be left to computer control. A continuing machine-operator interaction is part of normal operations.

Another manufacturing concept involves fully automated manufacturing systems. The processing within the manufacturing system itself is fully computer-controlled. Closed-loop operations must exist between sensors and actuators/effectors in the manufacturing system. The manufacturing system must be able to monitor its own performance and decision making for all required operations. For effective

automated operation, the mean time between operator interventions must be large when compared with the times between manufacturing setups.

$$\text{MTOI} = \left(\sum_i^n \tau_1 + \tau_2 + \tau_3 + \tau_4 + \cdots + \tau_4 \right) / n \tag{7.1}$$

where τ = setup time
i = initial setup
n = number of setups

The processes in use must rarely fail; the operator will intervene only when such failures occur. In such a setting, the operator's function is to ensure the adequate flow of work in progress and respond to system failure.

Several types of work cells are designed according to the concept of total manufacturing integration. The most sophisticated cell design involves fully automated processing and materials handling. Computers control the feeding of work in progress, the performance of the manufacturing process, and the removal of the work in progress. Manufacturing systems of this type provide the opportunity for the most advanced automated and integrated operations. The manufacturing system must be modified to achieve closed-loop operations for all of these functions.

Most manufacturing systems in use today are not very resourceful. They do not make use of external sensors that enable them to monitor their own performance. Rather, they depend on internal conditioning sensors to feed back (to the control system) information regarding manipulator positions and actions. To be effective, this type of manufacturing system must have a rigid structure and be able to determine its own position based on internal data (largely independent of the load applied). This leads to large, heavy, and rigid structures.

The more intelligent manufacturing systems use sensors that enable them to observe work in progress and a control loop that allows corrective action to be taken. Thus, such manufacturing systems do not have to be as rigid because they can adapt.

The evolution toward more intelligent and adaptive manufacturing systems has been slow, partly because the required technologies have evolved only in recent years and partly because it is difficult to design work cells that effectively use the adaptive capabilities. Enterprises are not sure whether such features are cost-effective and wonder how to integrate smart manufacturing systems into the overall strategy.

The emphasis must be on the building-block elements necessary for many types of processing. If the most advanced sensors are combined with the most advanced manufacturing systems, concepts, and state-of-the-art controllers and control loops, very sophisticated manufacturing systems can result. On the other hand, much more rudimentary sensors, effectors, and controllers can produce simple types of actions.

In many instances today, sensors are analog (they involve a continuously changing output property), and control loops make use of digital computers. Therefore, an analog-to-digital converter between the pre-processor and the digital control loop is often required.

The sensor may operate either passively or actively. In the passive case, the physical stimulus is available in the environment and does not have to be provided. For an active case, the particular physical stimulus must be provided. Machine vision and color identification sensors are an active means of sensing, because visible light must be used to illuminate the object before a physical stimulus can be received by the sensor. Laser sensors are also active-type sensors. Passive sensors include infrared devices (the physical stimulus being generated from infrared radiation associated with the temperature of a body) and sensors to measure pressure, flow, temperature, displacement, proximity, humidity, and other physical parameters.

7.2 Sensors in Manufacturing

Many types of sensors have been developed during the past several years, especially those for industrial process control, military uses, medicine, automotive applications, and avionics. Several types of sensors are already being manufactured by commercial companies.

Process control sensors in manufacturing will play a significant role in improving productivity, qualitatively and quantitatively, throughout the coming decades. The main parameters to be measured and controlled in industrial plants are temperature, displacement, force, pressure, fluid level, and flow. In addition, detectors for leakage of explosives or combustible gases and oils are important for accident prevention.

Optical-fiber sensors may be conveniently divided into two groups: (1) intrinsic sensors and (2) extrinsic sensors.

Although intrinsic sensors have, in many cases, an advantage of higher sensitivity, almost all sensors used in process control at present belong to the extrinsic type. Extrinsic-type sensors employ light sources such as LEDs, which have higher reliability, longer life, and lower cost than semiconductor lasers. They also are compatible with multimode fibers, which provide higher efficiency when coupled to light sources and are less sensitive to external mechanical and thermal disturbances.

As described in Chapter 3, objects can be detected by interrupting the sensor beam. Optical-fiber interrupters are sensors for which the principal function is the detection of moving objects. They may be classified into two types: reflection and transmission.

In the reflection-type sensor, the light beam emitted from the fiber is reflected back into the same fiber if the object is situated in front of the sensor.

In the transmission-type sensor, the emitted light from the input fiber is interrupted by the object, resulting in no received light in the output fiber located at the opposite side. Typical obstacle interrupters employ low-cost large-core plastic fibers because of the short transmission distance. The minimum detectable size of the object is typically limited to 1 mm by the fiber core diameter and the optical beam. The operating temperature range of commercially available sensors is typically –40 to +70°C. Optical-fiber sensors have been utilized in industry in many ways, such as:

- Detection of lot number and expiration dates—for example, in the pharmaceutical and food industries
- Color difference recognition—for instance, colored objects on a conveyer
- Defect detection, such as with missing wire leads in electronic components
- Counting discrete components—for example, bottles or cans
- Detecting absence or presence of labels—for instance, packaging in the pharmaceutical and food industries

Fiber-optic sensors for monitoring process variables such as temperature, pressure, flow, and liquid level are also classified into two types: (1) the normally OFF type in which the shutter is inserted between the fibers in the unactivated state. Thus, this type of sensor provides high and low levels as the light output corresponds to ON and OFF states, respectively, and (2) the normally ON type where the shutter is retracted from the gap in the inactivated state.

In both types, the shutter is adjusted so it does not intercept the light beam completely but allows a small amount of light to be transmitted, even when fully closed. This transmitted light is used to monitor the cable (fiber) continuity for faults and provides an intermediate state. Commercially available sensors employ fibers of 200-μm core diameter. The typical differential attenuation, which determines the ON-OFF contrast ratio, is about 20 dB. According to manufacturers' specifications, these sensors operate well over the temperature range –40 to +80°C with 2-dB variation in light output.

7.3 Temperature Sensors in Process Control

Temperature is one of the most important parameters to be controlled in almost all industrial plants, since it directly affects material properties and thus product quality. During the past few years, several temperature sensors have been developed for use in electrically or chemically hostile environments. Among these, the practical temperature sensors, which are now commercially available, are classified

into two groups: (1) low-temperature sensors with a range of –100 to +400°C using specific sensing materials such as phosphors, semiconductors, and liquid crystals, and (2) high-temperature sensors with a range of 500 to 2000°C based on blackbody radiation.

7.3.1 Semiconductor Absorption Sensors

Many of these sensors can be located up to 1500 m away from the optoelectronic instruments. The operation of semiconductor temperature sensors is based on the temperature-dependent absorption of semiconductor materials. Because the energy and gap of most semiconductors decrease almost linearly with increasing temperature T, the band-edge wavelength $\lambda_g(T)$ corresponding to the fundamental optical absorption shifts toward longer wavelengths at a rate of about 3 Å/°C [for gallium arsenide (GaAs)] with T. As illustrated in Fig. 7.1, when a light-emitting diode with a radiation spectrum covering the wavelength $\lambda_g(T)$ is used as a light source, the light intensity transmitted through a semiconductor decreases with T.

Figure 7.2 shows the reflection-type sensing element. A polished thin GaAs chip is attached to the fiber end and mounted in a stainless-steel capillary tube of 2-mm diameter. The front face of the GaAs is antireflection-coated, while the back face is gold-coated to return the light into the fiber.

The system configuration of the thermometer is illustrated in Fig. 7.3. In order to reduce the measuring errors caused by variations in parasitic losses, such as optical fiber loss and connector loss, this thermosensor employs two LED sources [one aluminum gallium arsenide (AlGaAs), the other indium gallium arsenide (InGaAs)] with different wavelengths. A pair of optical pulses with different wavelengths $\lambda_s = 0.88$ μm and $\lambda_r = 1.3$ μm are guided from the AlGaAs LED and the InGaAs LED to the sensing element along the fiber. The light

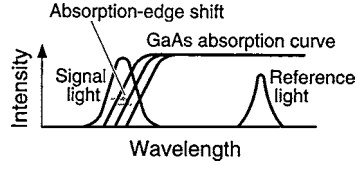


Figure 7.1 Operating principle of optical-fiber thermometer based on temperature-dependent GaAs light absorption.

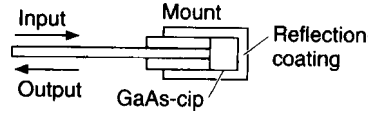


Figure 7.2 Sensing element of the optical-fiber thermometer with GaAs light absorber.

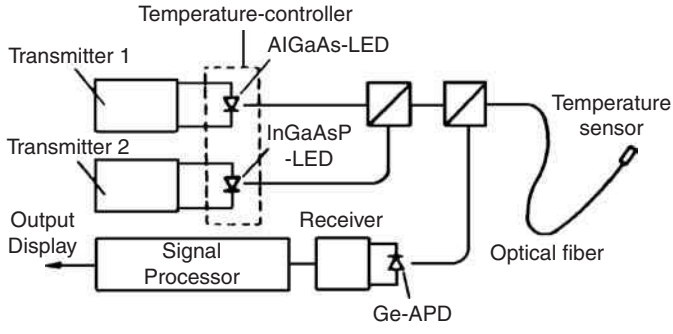


Figure 7.3 System configuration of the optical-fiber thermometer with GaAs light absorber.

of λ_s is intensity-modulated by temperature. On the other hand, GaAs is transparent for the light of λ_r, which is then utilized as a reference light. After detection by a *germanium avalanche photodiode* (GeAPD), the temperature-dependent signal λ_s is normalized by the reference signal λ_r in a microprocessor.

The performance of the thermometer is summarized in Table 7.1. An accuracy of better than ±2°C is obtained within a range of –20 to +150°C. The principle of operation for this temperature sensor is based on the temperature-dependent direct fluorescent emission from phosphors.

7.3.2 Semiconductor Temperature Detector Using Photoluminescence

The sensing element of this semiconductor photoluminescence sensor is a double-heterostructure GaAs epitaxial layer surrounded by two $Al_x\ Ga_{1-x}$ As layers. When the GaAs absorbs the incoming exciting light, the electron-hole pairs are generated in the GaAs layer. The electron-hole pairs combine and reemit the photons with a wavelength determined by temperature. As illustrated in Fig. 7.4, the luminescent wavelength shifts monotonically toward longer wavelengths as the temperature T increases. This is a result of the decrease in the energy gap E_g with T. Therefore, analysis of the luminescent spectrum yields the required temperature information. The double heterostructure of the sensing element provides excellent quantum efficiency for the luminescence because the generated electron-hole pairs are confined between the two potential barriers (Fig. 7.5).

The system is configured as shown in Fig. 7.6. The sensing element is attached to the end of the silica fiber (100-μm core diameter). The excitation light from an LED, with a peak wavelength of about 750 nm, is coupled into the fiber and guided to a special GRIN lens mounted to a block of glass. A first optical inference filter IF_1, located between the GRIN lens and the glass block, reflects the excitation

Property	Semiconductor Absorption Sensor (Mitsubishi, Japan)	Semiconductor Photoluminescence Sensor (ASEA Innovation, Sweden)	Phosphor A Sensor (Luxton, U.S.)	Phosphor B Sensor (Luxton, U.S.)
Range, °C	–20 to +150	0 to 200	20 to 240	20 to 400
Accuracy, °C	±2.0	±1.0	±2.0	±2.0
Diameter, m	2	From 0.6	0.7	1.6
Time constant, s	0.5	From 0.3	From 0.25	
Fiber type	100-μm silica core	100-μm silica core	400-μm polymer clad	Fiber silica
Fiber length, m	300	500	100	300
Light source	AlGaAs LED	AlGaAs LED	Halogen lamp	Halogen lamp

Table 7.1 Characteristics of Semiconductor Sensors (Thermometer Performance)

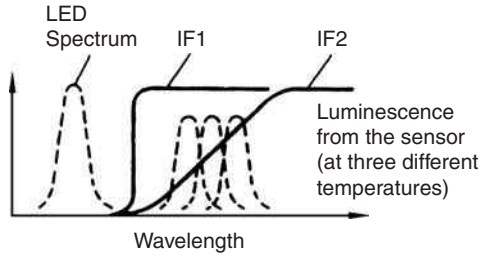
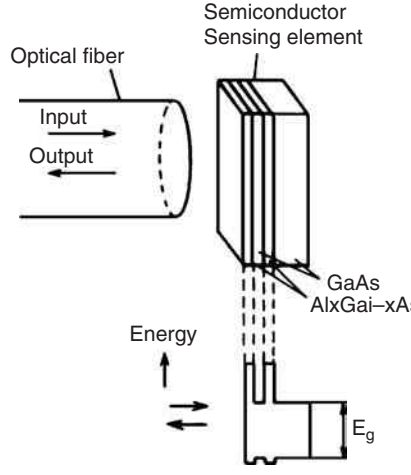


Figure 7.4 Operating principle of optical-fiber thermometer based on temperature-dependent photoluminescence from a GaAs epitaxial film.

Figure 7.5
Sensing element of an optical-fiber thermometer based on temperature-dependent photoluminescence.



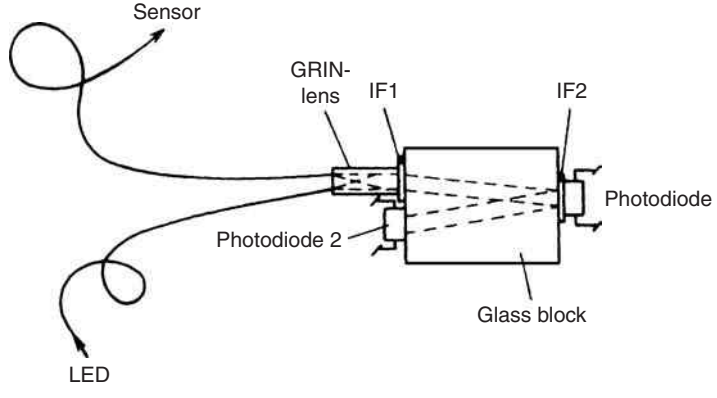


Figure 7.6 Optical system of an optical-fiber thermometer based on temperature-dependent photoluminescence.

light that is guided to the sensing element along the fiber. However, this optical filter is transparent to the returned photoluminescent light. The reflectivity of the second interference filter, IF_2, changes at about 900 nm. Because the peak wavelength of the luminescence shifts toward longer wavelengths with temperature, the ratio between the transmitted and reflected light intensifies if IF_2 changes. However, the ratio is independent of any variation in the excitation light intensity and parasitic losses. The two lights separated by IF_2 are detected by photodiodes 1 and 2. The detector module is kept at a constant temperature in order to eliminate any influence of the thermal drift of IF_2.

The measuring temperature range is 0 to 200°C, and the accuracy is ±1°C. According to the manufacturer's report, good long-term stability, with a temperature drift of less than 1°C over a period of nine months, has been obtained.

7.3.3 Temperature Detectors Using Point-Contact Sensors in Process Manufacturing Plants

Electrical sensors are sensitive to microwave radiation and corrosion. The need for contact-type temperature sensors have lead to the development of point-contact sensors that are immune to microwave radiation, for use in: (1) electric power plants using transformers, generators, surge arresters, cables, and bus bars; (2) industrial plants utilizing microwave processes; and (3) chemical plants utilizing electrolytic processes.

The uses of microwaves include drying powder and wood; curing glues, resins, and plastics; heating processes for food, rubber, and oil; device fabrication in semiconductor manufacturing; and joint welding of plastic packages, for example.

Semiconductor device fabrication is currently receiving strong attention. Most semiconductor device fabrication processes are now performed in vacuum chambers. They include plasma etching and stripping, ion implantation, plasma-assisted chemical vapor deposition, radio-frequency sputtering, and microwave-induced photoresist baking. These processes alter the temperature of the semiconductors being processed. However, the monitoring and controlling of temperature in such hostile environments is difficult with conventional electrical temperature sensors. These problems can be overcome by the contact-type optical-fiber thermometer.

7.3.4 Noncontact Sensors—Pyrometers

Because they are noncontact sensors, pyrometers do not affect the temperature of the object they are measuring. The operation of the pyrometer is based on the spectral distribution of blackbody radiation, which is illustrated in Fig. 7.7 for several different temperatures. According to the Stefan-Boltzmann law, the rate of the total radiated

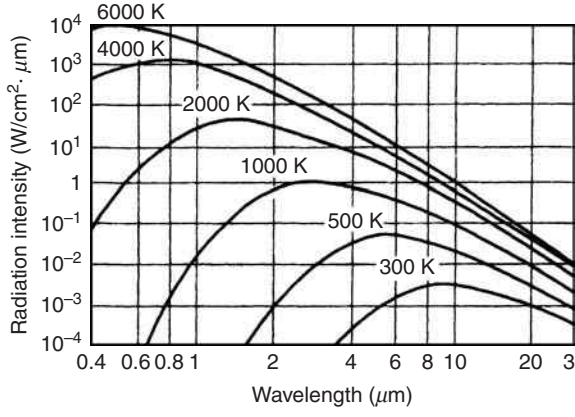


Figure 7.7 Spectral distribution of blackbody radiation.

energy from a blackbody is proportional to the fourth power of absolute temperature and is expressed as:

$$W_t = \sigma T^4 \tag{7.2}$$

where σ is the Stefan-Boltzmann constant and has the value of 5.6697×10^{-8} W/m$^2\hbar K^4$.

The wavelength at which the radiated energy has its highest value is given by Wien's displacement law,

$$\lambda_m T = 2.8978 \times 10^{-3} \text{m}\hbar\text{K} \tag{7.3}$$

Thus, the absolute temperature can be measured by analyzing the intensity of the spectrum of the radiated energy from a blackbody. A source of measurement error is the emissivity of the object, which depends on the material and its surface condition. Other causes of error are deviation from the required measurement distance and the presence of any absorbing medium between the object and the detector.

Use of optical fibers as signal transmission lines in pyrometers allows remote sensing over long distances, easy installation, and accurate determination of the position to be measured by observation of a focused beam of visible light from the fiber end to the object. The sensing head consists of a flexible bundle with a large number of single fibers and lens optics to pick up the radiated energy (Fig. 7.8).

The use of a single silica fiber instead of a bundle is advantageous for measuring small objects and longer distance transmission of the picked-up radiated light. The lowest measurable temperature is 500°C, because of the unavoidable optical loss in silica fibers at wavelengths longer than 2 μm. Air cooling of the sensing head is usually necessary when the temperature exceeds 1000°C.

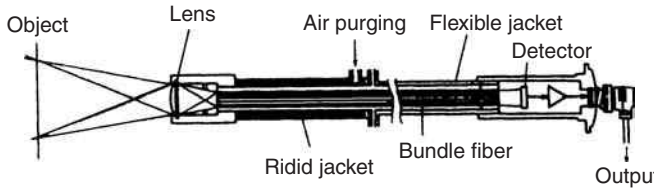


Figure 7.8 Schematic diagram of an optical-fiber pyrometer.

Optical-fiber pyrometers are one of the most successful optical-fiber sensors in the field of process control in manufacturing. Typical applications are:

- Casting and rolling lines in steel and other metal plants
- Electric welding and annealing
- Furnaces in chemical and metal plants
- Fusion, epitaxial growth, and sputtering processes in the semiconductor industry
- Food processing, paper manufacturing, and plastic processing

Figure 7.9 is a block diagram of the typical application of optical-fiber pyrometers for casting lines in a steel plant, where the temperature distribution of the steel slab is measured. The sensing element consists of a linear array of fused-silica optical rods, thermally protected by air-purge cooling. Light radiated from the heated slabs is collected by the optical rods and coupled into a 15-m-long bundle of fibers, which transmits light to the optical processing unit. In this system, each fiber in the bundle carries the signal from a separate lens, which provides the temperature information at the designated spot of the slabs. An optical scanner in the processing unit scans the bundle, and the selected light signal is analyzed in two wavelength bands by using two optical interference filters.

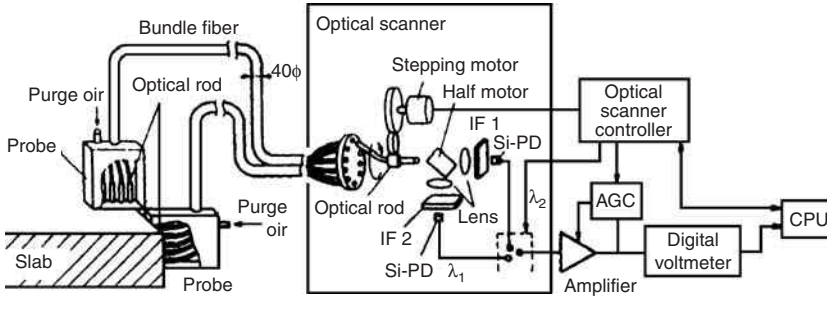


Figure 7.9 Temperature distribution measurement of steel slabs by an optical-fiber pyrometer using the two-wavelength method.

7.4 Pressure Sensors

If a pressure P acting on a diaphragm compresses a spring until an equilibrium is produced, the pressure can be represented as:

$$F(\text{kg}) = A(\text{m}^2) \times P(\text{kg}/\text{m}^2) \tag{7.4}$$

In this equation, F represents the force of the spring and A represents a surface area of the diaphragm. The movement of the spring is transferred via a system of levers to a pointer whose deflection is a direct indication of the pressure (Fig. 7.10). If the measured value of the pressure must be transmitted across a long distance, the mechanical movement of the pointer can be connected to a variable electrical resistance (potentiometer). A change in the resistance results in a change in the measured voltage, which can then easily be evaluated by an electronic circuit or further processed. This example illustrates the fact that a physical quantity is often subject to many transformations before it is finally evaluated.

7.4.1 Piezoelectric Crystals

Piezoelectric crystals may be utilized to measure pressure. Electrical charges are produced on the opposite surfaces of some crystals when they are mechanically loaded by deflection, pressure, or tension. The electrical charge produced in the process is proportional to the effective force. This change in the charge is very small. Therefore, electrical amplifiers are used to make it possible to process the signals (Fig. 7.11).

Pressure in this situation is measured by transforming it into a force. If the force produced by pressure on a diaphragm acts on a piezoelectric crystal, a signal which is proportional to the pressure measured can be produced by using suitable amplifiers.

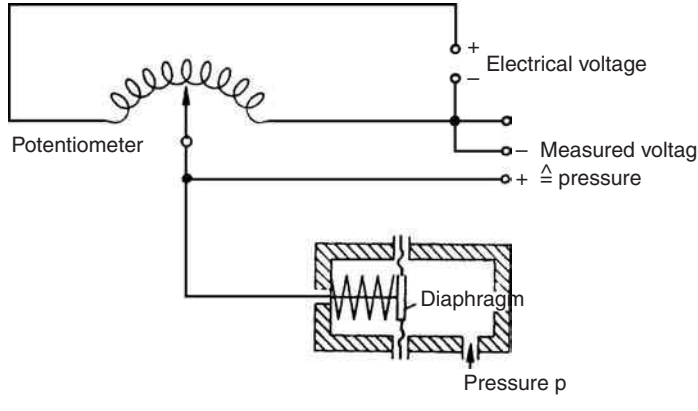


Figure 7.10 Deflection as a direct indication of pressure.

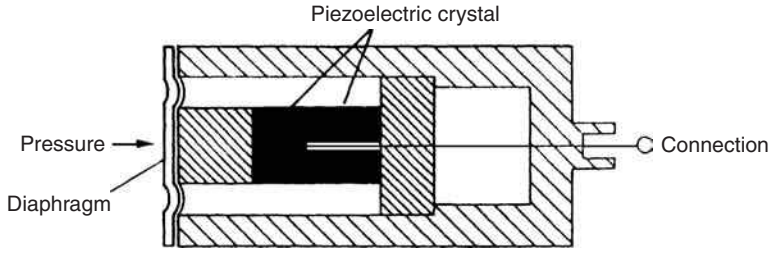


Figure 7.11 Electrical amplifiers are connected to a piezoelectric crystal.

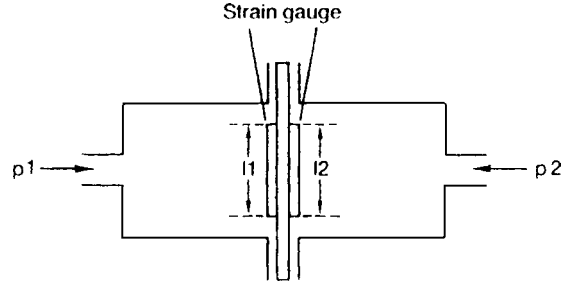


Figure 7.12 Strain gauge for measurement of pressure.

7.4.2 Strain Gauges

Strain gauges can also measure pressure. The electrical resistance of a wire-type conductor is dependent, to a certain extent, on its cross-sectional area. The smaller the cross section (i.e., the thinner the wire), the greater the resistance of the wire. A strain gauge is a wire that conducts electricity and stretches as a result of the mechanical influence (tension, pressure, or torsion) and thus changes its resistance in a manner that is detectable. The wire is attached to a carrier, which in turn is attached to the object to be measured. Conversely, for linear compression, which enlarges the cross-sectional area of a strain gauge, resistance is reduced. If a strain gauge is attached to a diaphragm (Fig. 7.12), it will follow the movement of the diaphragm. It is either pulled or compressed, depending on the flexure of the diaphragm.

7.5 Fiber-Optic Pressure Sensors

A Y-guide probe can be used as a pressure sensor in process control if a reflective diaphragm, moving in response to pressure, is attached to the end of the fiber (Fig. 7.13). This type of pressure sensor has a significant advantage over piezoelectric transducers since it works as a noncontact sensor and has a high frequency response. The pressure signal is transferred from the sealed diaphragm to the sensing diaphragm,

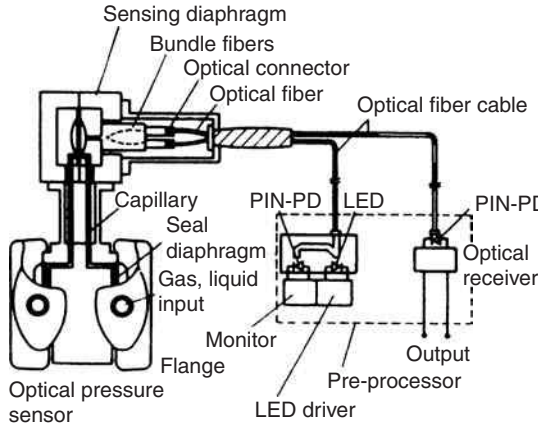


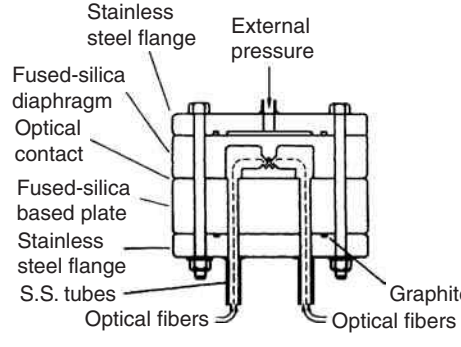
Figure 7.13 Schematic diagram of a fiber-optic pressure sensor using a Y-guide probe with a diaphragm attached.

which is attached to the end of the fiber. With a stainless-steel diaphragm about 100 μm thick, hysteresis of less than 0.5 percent and linearity within ±0.5 percent are obtained up to the pressure level of 3×10^5 kg/m^2 (2.94 MPa) in the temperature range of –10 to +60°C.

The material selection and structural design of the diaphragm are important to minimize drift. Optical-fiber pressure sensors are expected to be used under severe environments in process control. For example, process slurries are frequently highly corrosive, and the temperature may be as high as 500°C in coal plants. The conventional metal diaphragm exhibits creep at these high temperatures. In order to eliminate these problems, an all-fused-silica pressure sensor based on the microbending effect in optical fiber has been developed (Fig. 7.14). This sensor converts the pressure applied to the fused silica diaphragm into an optical intensity modulation in the fiber.

A pressure sensor based on the wavelength filtering method has been developed. The sensor employs a zone plate consisting of a

Figure 7.14 Fiber-optic microbend sensor.



reflective surface, with a series of concentric grooves at predetermined spacing. This zone plate works as a spherical concave mirror whose effective radius of curvature is inversely proportional to the wavelength. At the focal point of the concave mirror, a second fiber is placed which transmits the returned light to two photodiodes with different wavelength sensitivities. When broadband light is emitted from the first fiber to the zone plate, and the zone plate moves back and forth relative to the optical fibers in response to the applied pressure, the wavelength of the light received by the second fiber is varied, causing a change in the ratio of outputs from the two photodiodes. The ratio is then converted into an electrical signal, which is relatively unaffected by any variations in parasitic losses.

7.6 Nano-Positioning Capacitive Metrology Sensors

The nano-positioning capacitive sensor (Fig. 7.15) provides fast response, and precise trajectory control. It is capable of providing digital controllers with a fast fiber-link interface and an ID-chip for automatic calibration functions. It provides a resolution of 0.1 nm.

7.6.1 Nano-Capacitive Positioning Sensors

Single-electrode capacitive sensors are direct metrology devices. They use an electric field to measure the change in capacitance between the probe and a conductive target surface, without any physical contact. This makes them free of friction, and hysteresis, and provides

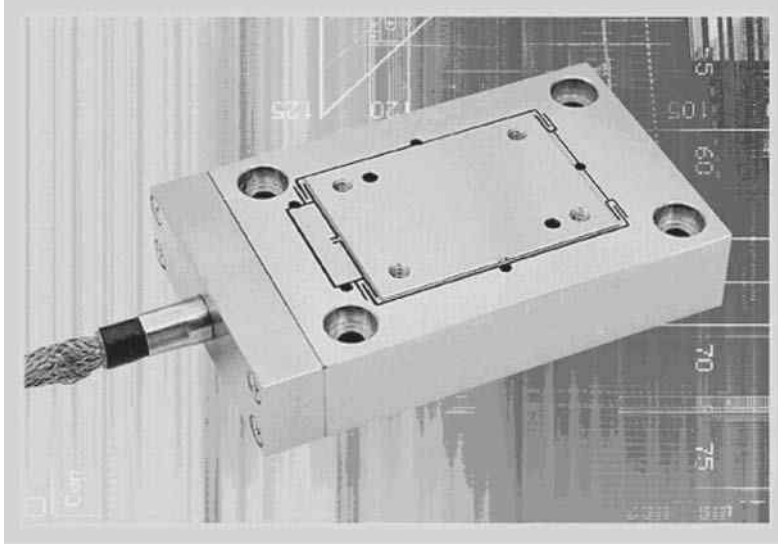


Figure 7.15 A nano-positioning capacitive sensor.

high-phase fidelity and bandwidth. The selectable bandwidth setting allows the user to adapt the system to different applications. For the highest accuracy and sub-nanometer resolution, the bandwidth can be limited to 10 Hz.

For high-dynamics measurements, a bandwidth of up to 6.6 kHz is possible, with a resolution still in the 1-nm range.

The user can choose a measurement range from 20 to 500 µm, depending on the nominal measurement range of the selected sensor. The ten-channel system provides different extended measuring ranges for each selected sensor.

The capacitive sensor's measuring capacitance is the metrology system for most nano-positioning applications. They are absolute-measuring high-dynamics devices. The capacitive sensors / control electronics use a high-frequency AC excitation signal for enhanced bandwidth and drift-free measurement stability (Fig. 7.16).

The electronics of the capacitive position sensor incorporate an innovative design providing superior linearity, low sensitivity to cable capacitance, low background noise, and low drift.

The *Integrated Linearization System* (ILS) compensates for influences caused by errors, such as non-parallelism of the plates. A comparison between a conventional capacitive position sensor system and the results obtained with the ILS is shown in Fig. 7.17. When nano-capacitance sensors are used with digital controllers, which add polynomial linearization techniques, a positioning linearity of up to 0.003 percent is achievable.

The travel range is 15 µm; the gain 1.5 µm/V. Linearity is better than 0.02 percent; even higher linearity is achievable with digital



Figure 7.16 A 15-µm piezo nano-positioning stage with an integrated capacitive position sensor.

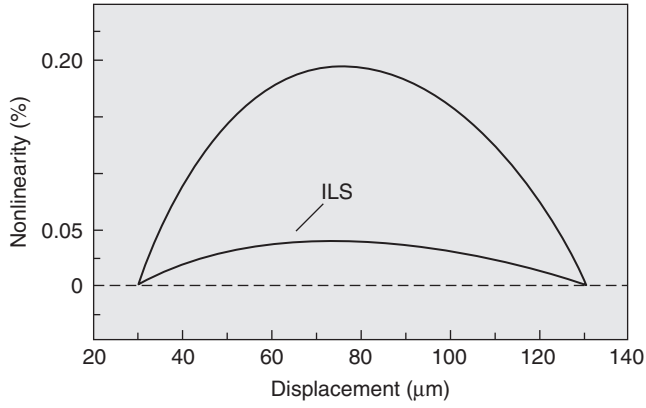


Figure 7.17 The linearity of a conventional capacitive position sensor system versus ILS (Integrated Linearization System), shown before digital linearization.

controllers. Figure 7.18 shows the linearity of a piezo flexure nano-positioning stage with an integrated capacitive position sensor operated in closed-loop mode with an analog controller. All errors contributed by the mechanics, PZT drive, sensors and electronics are included in the resulting linearity of better than 0.02 percent. Even higher linearity is achievable with digital controllers.

The long-term stability of the nano-capacitive position sensor and electronics design is shown in Fig. 7.19. They enable direct measurement of the moving platform and are especially well-suited for multiaxis measurements in parallel metrology configurations. Parallel metrology means that the controller monitors all controlled degrees of freedom relative to "ground," a prerequisite for active trajectory control. The capacitive sensors also provide high linear accuracy.

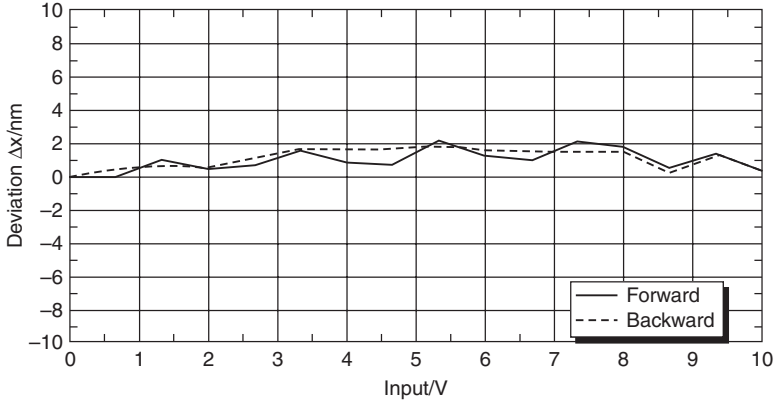


Figure 7.18 Linearity of a 15-μm piezo nano-positioning stage operated with advanced control electronics.

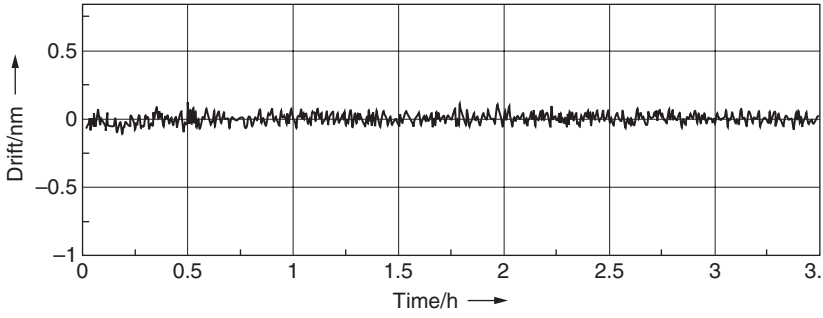


Figure 7.19 Measurement stability of a capacitive position sensor control board with a 10-pF reference capacitor over 3.5 hours.

7.7 Electrode Geometry, Sensor Surface Flatness, and Finish

During sensor production, great care must be taken to maintain critical mechanical tolerances. Measuring surfaces are diamond-tool machined using sophisticated process control techniques. The result is the smooth, ultra-flat, mirrored surfaces required to obtain the highest resolution commercially available. The target and probe plates are configured in such a way that the sensor capacitance in air is 10 pF at the nominal measuring distance. This means that one sensor electronics module can operate all sensor range models without modification.

7.8 Special Design Eliminates Cable Influences

When measuring distance by detection of capacitance changes, fluctuations in the cable capacitance can have an adverse effect on accuracy. This is why most capacitive measurement systems only provide satisfactory results with short well-defined cable lengths. Nanocapacitance systems use a special design that eliminates cable influences, permitting use of cable lengths of up to 3 m without difficulty. For optimum results, it is recommended that calibration of the sensor-actuator system be done by Metrology Laboratories. Longer distances between sensor and electronics can be spanned with special loss-free digital transmission protocols.

7.9 Materials Achieving Greater Accuracy

The best measurements are obtained when the coefficient of thermal expansion of the sensor and the substrate to which it is affixed are as

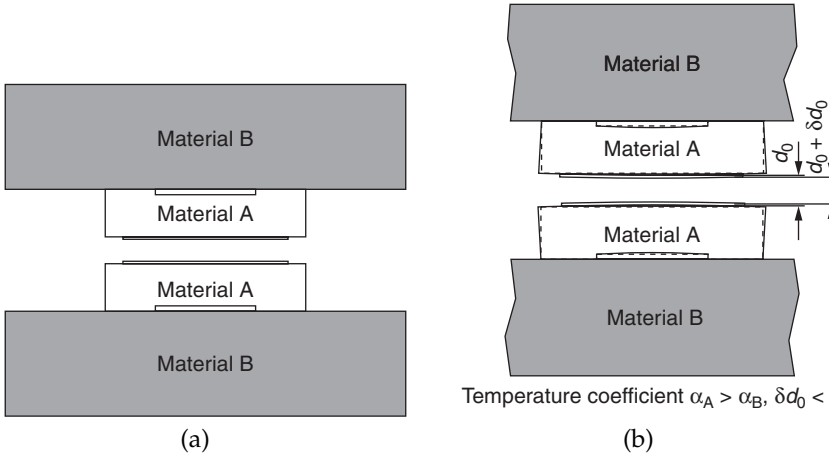


Figure 7.20 (a) The optimum match between sensor material and environment. (b) The CTE of sensor material and environment are different. Buckling can reduce sensor flatness and accuracy.

nearly equal as possible (Fig. 7.20a). If they differ, temperature changes lead to mechanical stress and deformation, which can compromise accuracy (Fig. 7.20.b). It is the material choice that affects performance, not the absolute value of the temperature coefficient. A low-temperature coefficient in a sensor mounted on a stainless steel stage will thus give poorer results than a steel sensor (Fig. 7.20b).

7.10 Mounting, Calibration, and Measuring Ranges

The sensors should be mounted with a mid-range distance between the plates equal to the nominal measuring range (Fig. 7.21). The measuring range then comprises 50 to 150 percent of this distance. The corresponding range at the output of the sensor electronics is 10 volts in width. The probe/target capacitance at a distance equal to the nominal measuring range equals that of the 10-pF reference capacitor

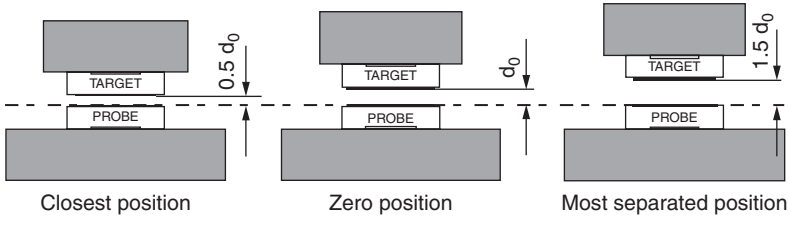


Figure 7.21 The distance between the probe and target at the center position, with closest spacing and maximum separation.

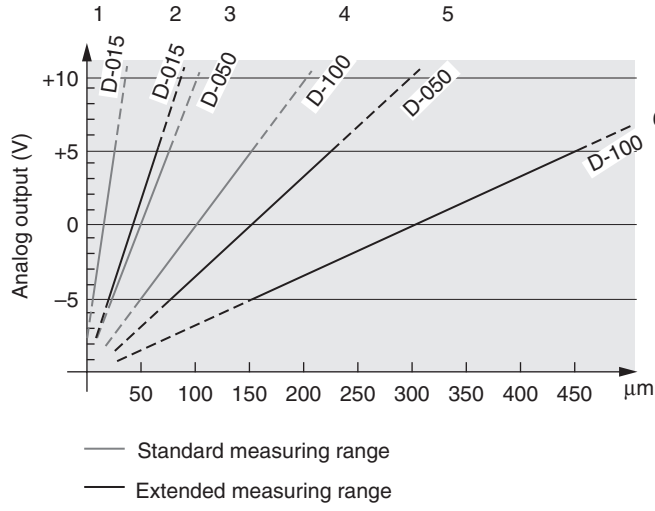


Figure 7.22 Measuring ranges of different capacitive position sensors, standard ranges are 1- D-015, 3- D-050, and 4- D-100, while the extended ranges are 2- D016, 5- D050, and 6- D-100.

in the electronics. The nominal range itself can be increased or decreased by selecting a different reference capacitor in the electronics (Fig. 7.22).

7.11 Parallelism of Measuring Surfaces

For optimum results, target and probe plates must remain parallel to each other during measurement (Figs. 7.23a and 7.23b). Positioning systems with multilink flexure guidance reduce tip and tilt to negligible levels (Fig. 7.24) and achieve outstanding accuracy.

This effect is negligible in nano-positioning systems with micro-radian-range guiding accuracy (Fig. 7.24).

Flexure-guided nano-positioning systems offer sub-micro-radian guiding accuracy and are optimally suited for capacitive sensors (Fig. 7.24).

7.12 Electronics Support

Standard nano-capacitive position sensor electronics are supported by all digital controllers. Inputs are developed on I/O controller cards and on several OEM controllers for high-speed parallel port controllers (Fig. 7.25). In addition, nano-capacitive sensors may be developed to accommodate up to seven channels to meet specific needs.

Various sensor probes are available at 100 µm, 50 µm, and 20 µm and are developed for nominal measurement range (Fig. 7.26).

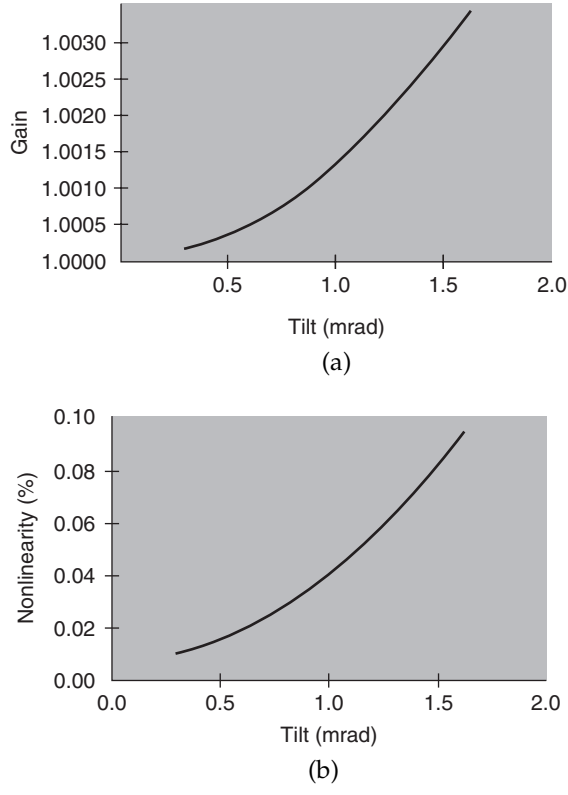


Figure 7.23 (a) Relative sensor gain versus tilt. (b) Nonlinearity versus tilt.

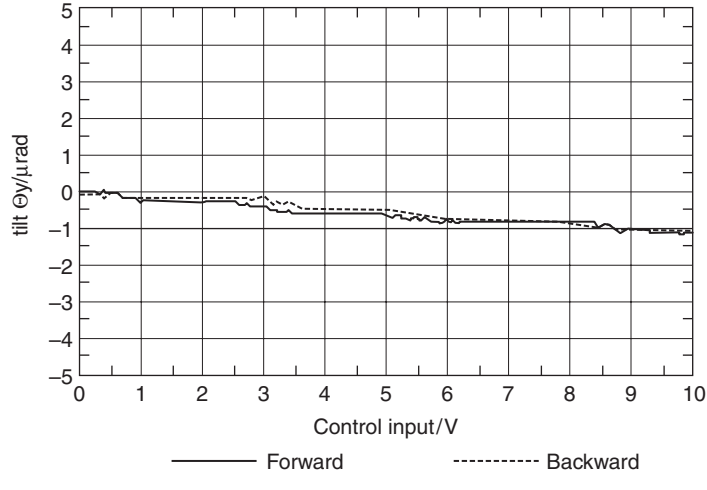


Figure 7.24 Nano-positioning systems with a micro-radian-range.

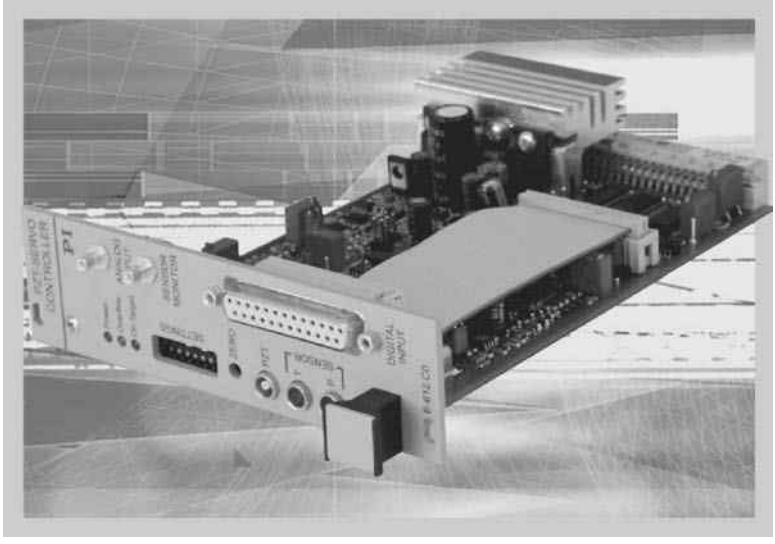


Figure 7.25 A high-speed nano automation controller board.

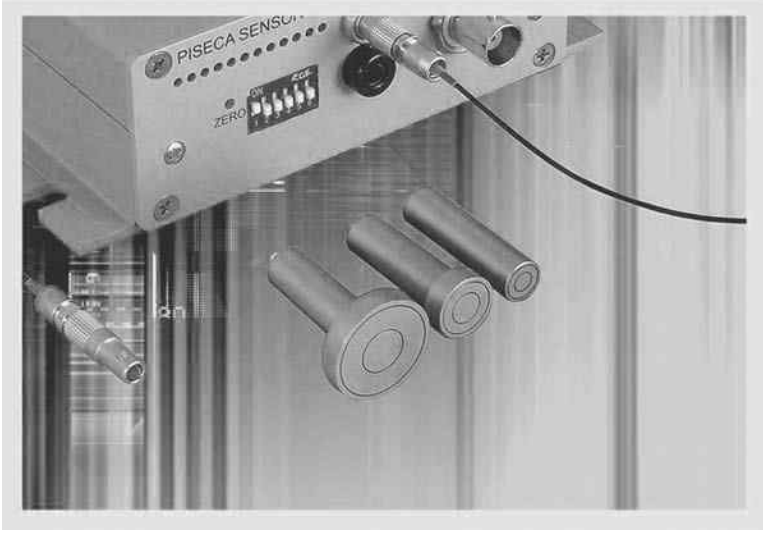


Figure 7.26 High-precision capacitive sensor probes with signal conditioner electronics.

High-speed nano automation signal conditioner electronics are specially designed for single-electrode capacitive position sensor probes. They provide analog output with very high linearity, exceptional long-term stability, sub-nanometer position resolution, and have bandwidths of up to 6.6 kHz (Fig. 7.27).

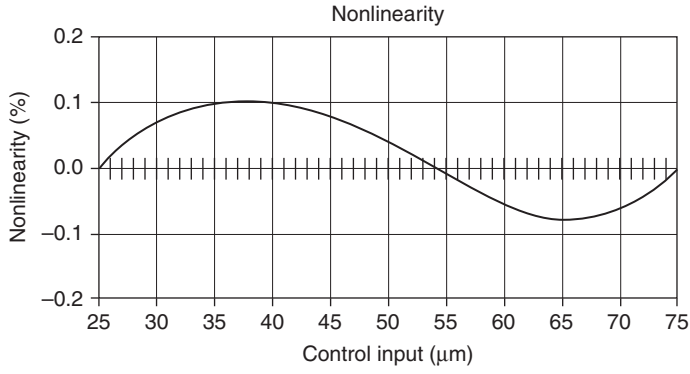


Figure 7.27 The output linearity error of a high-speed nano controller signal conditioner.

7.13 Sensor Installation

The simple installation of the single-electrode probes is facilitated by the LED-bar indicating the optimum gap between probe and target.

7.13.1 Factory Calibration for Improved Linearity

The highest possible linearity and accuracy are achieved with factory calibration of the sensor probe, together with the signal conditioner electronics. Two measurement ranges can be calibrated at the same time for one particular sensor probe. Factory calibration also optimizes parameters like ILS (linearization), gain, and offset and eliminates cable capacitance influences.

7.14 Using the Integrated Linearization System (ILS) for Highest Accuracy

A linearization circuit compensates the influences of parallelism errors between sensor and target to ensure excellent measuring linearity to 0.1 percent.

7.14.1 Multichannel Measurements

The sensor electronics are equipped with I/O lines for the synchronization of multiple sensor systems.

7.15 Displacement Sensors for Robotic Applications

The operating principle of a displacement sensor using Y-guide probes is illustrated in Fig. 7.28. The most common Y-guide probe is

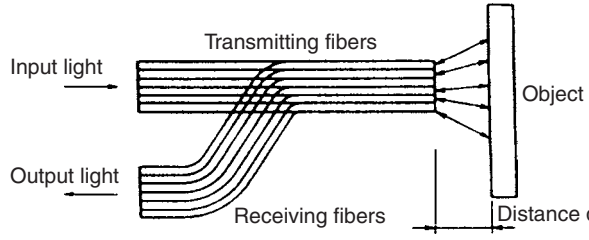


Figure 7.28 The principle of operation of a fiber-optic mechanical sensor using a Y-guide probe.

a bifurcated fiber bundle. The light emitted from one bundle is back-reflected by the object to be measured and collected by another bundle (receiving fibers). As a result, the returned light at the detector is intensity-modulated to a degree dependent on the distance between the end of the fiber bundle and the object. The sensitivity and the dynamic range are determined by the geometrical arrangement of the array of fiber bundles and by both the number and type of the fibers. Figure 7.29 shows the relative intensity of the returned light as a function of distance for three typical arrangements: random, hemispherical, and concentric circle arrays. The intensities increase with distance and reach a peak at a certain discrete distance. After that, the intensities fall off very slowly. Most sensors use the high-sensitivity regions in these curves. Among the three arrangements, the random array has the highest sensitivity but the narrowest dynamic range. The displacement sensor using the Y-guide probe provides a resolution of 0.1 μm,

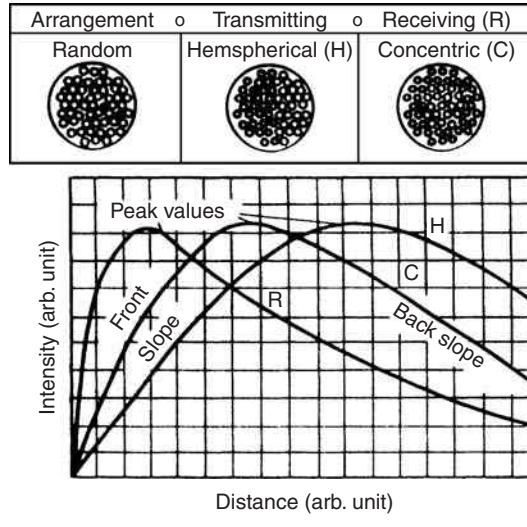


Figure 7.29 Relative intensity of returned light for three fiber-optic arrangements

linearity within 5 percent, and a dynamic range of 100 μm displacement. Y-guide probe displacement sensors are well-suited for robotics applications as position sensors and for gauging and surface assessment since they have high sensitivity to small distances.

One profound problem of this type of displacement sensor is the measuring error arising from the variation in parasitic losses along the optical transmission line. Recalibration is required if the optical path is interrupted, which limits the range of possible applications. In order to overcome this problem, a line-loss-independent displacement sensor with an electrical subcarrier phase encoder has been implemented. In this sensor, the light from an LED modulated at 160 MHz is coupled into the fiber bundle and divided into two optical paths. One of the paths is provided with a fixed retroreflector at its end. The light through the other is reflected by the object. The two beams are returned to the two photodiodes separately. Each signal, converted into an electric voltage, is electrically heterodyned into an intermediate frequency at 455 kHz. Then, the two signals are fed to a digital phase comparator, the output of which is proportional to the path distance. The resolution of the optical path difference is about 0.3 mm, but improvement of the receiver electronics will provide a higher resolution.

7.16 Process Control Sensors Measuring and Monitoring Liquid Flow

According to the laws of fluid mechanics, an obstruction inserted in a flow stream creates a periodic turbulence behind it. The frequency of shedding the turbulent vortices is directly proportional to the flow velocity. The flow sensor in Fig. 7.30 has a sensing element consisting of a thin metallic obstruction and a downstream metallic bar attached to a multimode fiber-microbend sensor. As illustrated in Fig. 7.31, the vortex pressure produced at the metallic bar is transferred, through a diaphragm at the pipe wall that serves as both a seal and a pivot for the bar, to the microbend sensor located outside the process line pipe. The microbend sensor converts the time-varying mechanical force caused by the vortex shedding into a corresponding intensity modulation of the light. Therefore, the frequency of the signal converted into the electric voltage at the detector provides the flow-velocity information. This flow sensor has the advantage that the measuring

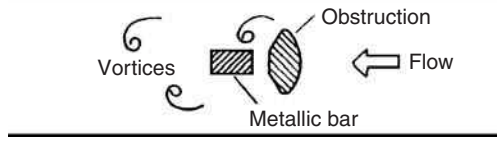


Figure 7.30 Principle of operation of a vortex-shedding flow sensor.

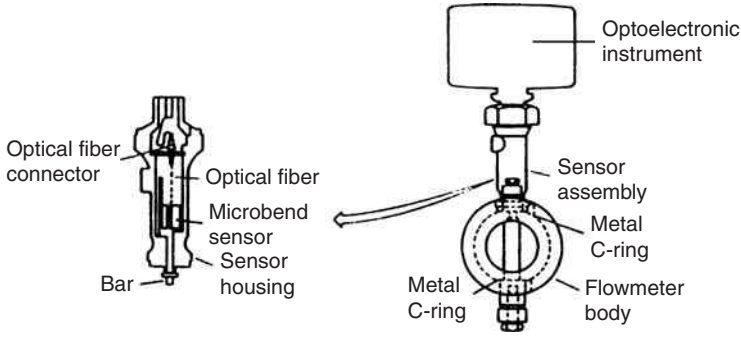


Figure 7.31 Schematic diagram of a vortex-shedding flow sensor.

accuracy is essentially independent of any changes in the fluid temperature, viscosity, or density, and in the light source intensity. According to the specifications for typical optical vortex-shedding flow sensors, flow rate can be measured over a Reynolds number range from 5×10^3 to 6000×10^3 at temperatures from –100 to +600°C. This range is high compared to that of conventional flow meters. In addition, an accuracy of ±0.4 and ±0.7 percent, respectively, is obtained for liquids and gases with Reynolds numbers above 10,000.

7.16.1 Flow Sensors Detecting Small Air Bubbles for Process Control in Manufacturing

Another optical-fiber flow sensor employed in manufacturing process control monitors a two-fluid mixture (Fig. 7.32). The sensor can distinguish between moving bubbles and liquid in the flow stream and display the void fraction, namely, the ratio of gas volume to the

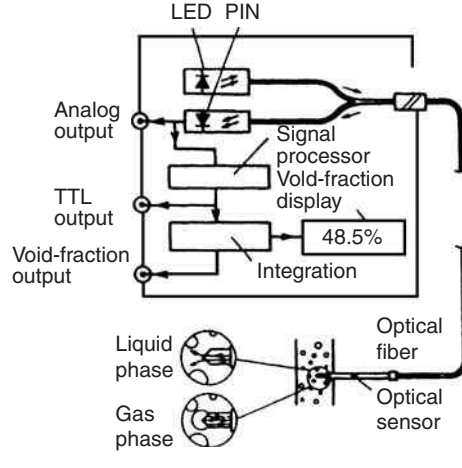


Figure 7.32 Flow sensor for two-phase mixtures.

total volume. The principle of operation is quite simple. The light from the LED is guided by the optical fiber to the sensing element, in which the end portion of the fiber is mounted in a stainless steel needle with a 2.8-mm outer diameter. When liquid is in contact with the end of the fiber, light enters the fluid efficiently and very little light is returned. However, when a gas bubble is present, a significant fraction of light is reflected back. With this technique, bubbles as small as 50 μm may be detected with an accuracy of better than 5 percent and a response time of only 10 μs.

Potential applications of this flow sensor for the control of processes in manufacturing systems are widespread—for example, detection of gas plugs in production wells in the oil industry and detection of fermenters and distillers in the blood-processing and pharmaceutical industries.

An optical-fiber flow sensor for a two-phase mixture based on Y-guide probes is shown in Fig. 7.33. Two Y-guide probes are placed at different points along the flow stream to emit the input light and pick up the retroreflected light from moving solid particles in the flow. The delay time between the signals of the two probes is determined by the average velocity of the moving particles. Therefore, measurement of the delay time by a conventional correlation technique provides the flow velocity. An accuracy of better than ±1 percent and a dynamic range of 20:1 are obtained for flow velocities up to 10 m/s. A potential problem of such flow sensors for two-phase mixtures is poor long-term stability, because the optical fibers are inserted into the process fluid pipes.

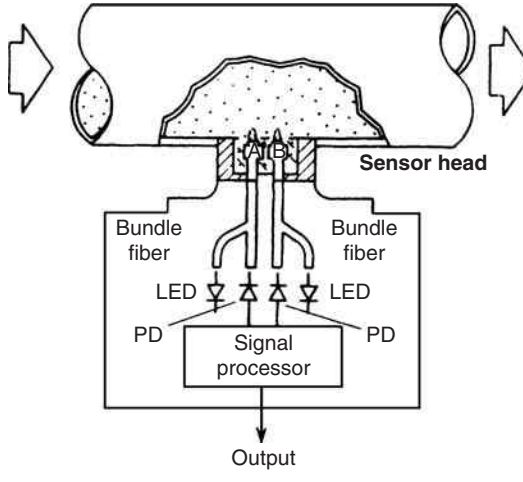


Figure 7.33 Flow sensor using two Y-guided probes based on a correlation technique.

7.16.2 Liquid Level Sensors in Manufacturing Process Control for Petroleum and Chemical Plants

Several optical-fiber liquid level sensors developed in recent years have been based on direct interaction between the light and liquid. The most common method in commercial products employs a prism attached to the ends of two single optical fibers (Fig. 7.34). The input light from an LED is totally internally reflected and returns to the output fiber when the prism is in air. However, when the prism is immersed in liquid, the light refracts into the fluid with low reflection, resulting in negligible returned light. Thus, this device works as a liquid level switch. The sensitivity of the sensor is determined by the contrast ratio, which depends on the refractive index of the liquid. Typical examples of signal output change for liquids with different refractive indices are indicated in Table 7.2.

The output loss stays at a constant value of 33 dB for refractive indices higher than 1.40. The signal output of a well-designed sensor can be switched for a change in liquid level of only 0.1 mm.

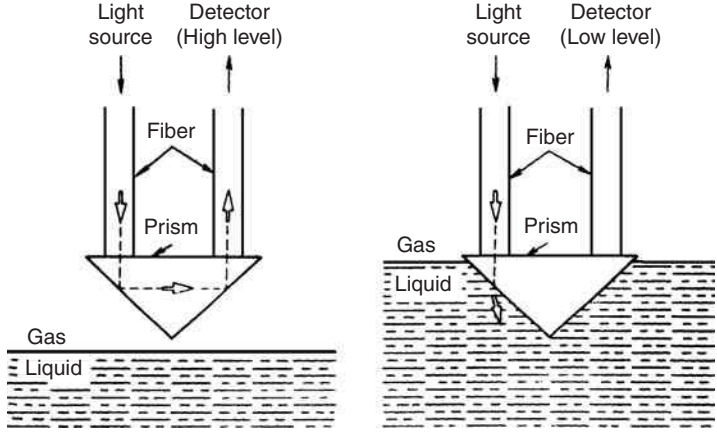


Figure 7.34 Principle of operation of a liquid level sensor with a prism attached to two optical fibers.

Refractive Index, *n*	Loss Change, dB
1.333	2.1
1.366	4.6
1.380	6.0
1.395	31.0

Table 7.2 Refractive Index versus Output

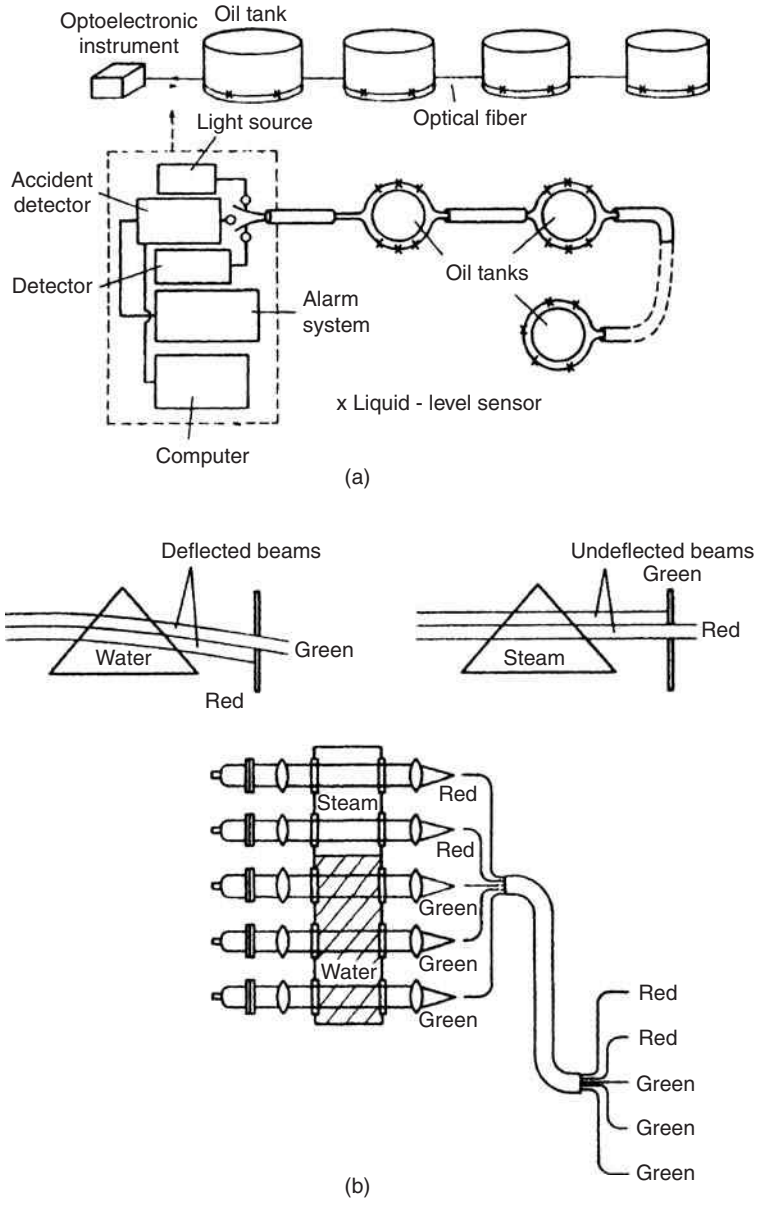


Figure 7.35 (a) Principle of operation of a liquid level sensor for the measurement of boiler-drum water. (b) Liquid level sensor for the measurement of boiler-drum water with five-port sensors.

Problems to be solved for this sensor are dirt contamination on the prism surface and bubbles in the liquid. Enclosing the sensing element with a fine filter helps keep it clean and simultaneously reduces level fluctuations caused by bubbles. Since optical-fiber liquid level sensors have the advantages of low cost and electrical isolation, their use is widespread in petroleum and chemical plants, where the hazardous environment causes difficulties with conventional sensors. They are used, for example, to monitor storage tanks in a petroleum plant.

Another optical-fiber liquid level sensor, developed for the measurement of boiler-drum water levels, employs a triangularly shaped gauge through which red and green light beams pass. The beams are deflected as it fills with water so that the green light passes through an aperture. In the absence of water, only red light passes through. Optical fibers transmit red or green light from individual gauges to a plant control room located up to 150 m from the boiler drum (Fig. 7.35). The water level in the drum is displayed digitally.

This liquid level sensor operates at temperatures up to 170°C and pressures up to 3200 lb/in^2 gauge. Many sensor units are installed in the boiler drum, and most have been operating for seven years. This sensor is maintenance-free, fail-safe, and highly reliable.

7.17 Sensory MEMS Enable Certain Molecules to Signal Breast Cancer's Spread

Metastasis is responsible for approximately 90 percent of all cancer-related deaths. Sensory MEMS technology enabled researchers to uncover how breast tumors use a particular type of molecule to promote metastasis. It also revealed that tumor cells learn to exploit cytokines to promote the spread of breast cancer.

The work of several scientists examine how cells in the body communicate with each other through cytokines, signaling molecules that direct a wide range of activities such as cell growth and movement. One important cytokine *transforming growth factor* (TGF) normally suppresses tumor development. However, according to the findings, cancer cells in humans are able to misuse these cytokines for their own gain by compelling TGF to enhance a tumor's ability to spread instead of suppressing it.

Using computer-based analysis to classify patient tumor samples based on their levels of TGF, it was observed that about half of all breast tumors contained active TGF. The affected tumors were found to be more aggressive and more likely to metastasize to the lung during the course of the patients' disease.

Using mice for the next set of experiments, it was discovered that TGF prompts breast cancer cells to make a second cytokine, known as *angiopoietin-like 4* (ANGPTL4), which enhances the ability of the cancer to spread to the lungs through the blood circulation. The results show that the breast cancer cells use ANGPTL4 to break down the thin capillaries of the lung, thus facilitating their escape into the lung tissue.

TGF enhances human breast cancer metastasis and reveals how tumor cells learn to exploit cytokines by making them work as a relay system to promote the spread of breast cancer. The researchers are currently seeking to determine whether TGF and ANGPTL4 may also be active in other types of tumors, and are evaluating ways to interfere with the action of these cytokines to prevent metastasis in cancer patients. Deciphering how cancer cells take advantage of these cytokines is essential for developing therapies that can prevent this process since cytokines that act outside of cells can be more easily targeted by drugs that block their activity.

The study provides support for developing agents that interfere with TGF in order to prevent and treat cancer metastasis. It points at ANGPTL4 as a possible target to interrupt the TGF stimulus of metastasis without interfering with the molecule's beneficial effects. Several pharmaceutical companies are currently testing TGF-blocking compounds in clinical trials as candidate drugs against breast cancer, melanoma, and other types of cancer.

7.18 On-Line Measuring and Monitoring of Gas by Spectroscopy

An optical spectrometer or optical filtering unit is often required for chemical sensors because the spectral characteristics of absorbed, fluorescent, or reflected light indicate the presence, absence, or precise concentration of a particular chemical species (Fig. 7.36).

Sensing of chemical parameters via fibers is usually done by monitoring changes in a suitably selected optical property—absorbance, reflectance, scattering (turbidity), or luminescence (fluorescence or phosphorescence), depending on the particular device. Changes in parameters such as the refractive index may also be employed for sensing purposes. The change in light intensity due to absorption is determined by the number of absorbing species in the optical path, and is related to the concentration C of the absorbing species by the Beer-Lambert relationship. This law describes an exponential reduction of light intensity with distance (and also concentration) along the optical path. Expressed logarithmically,

$$A = \log I_0 / I = \eta l C$$

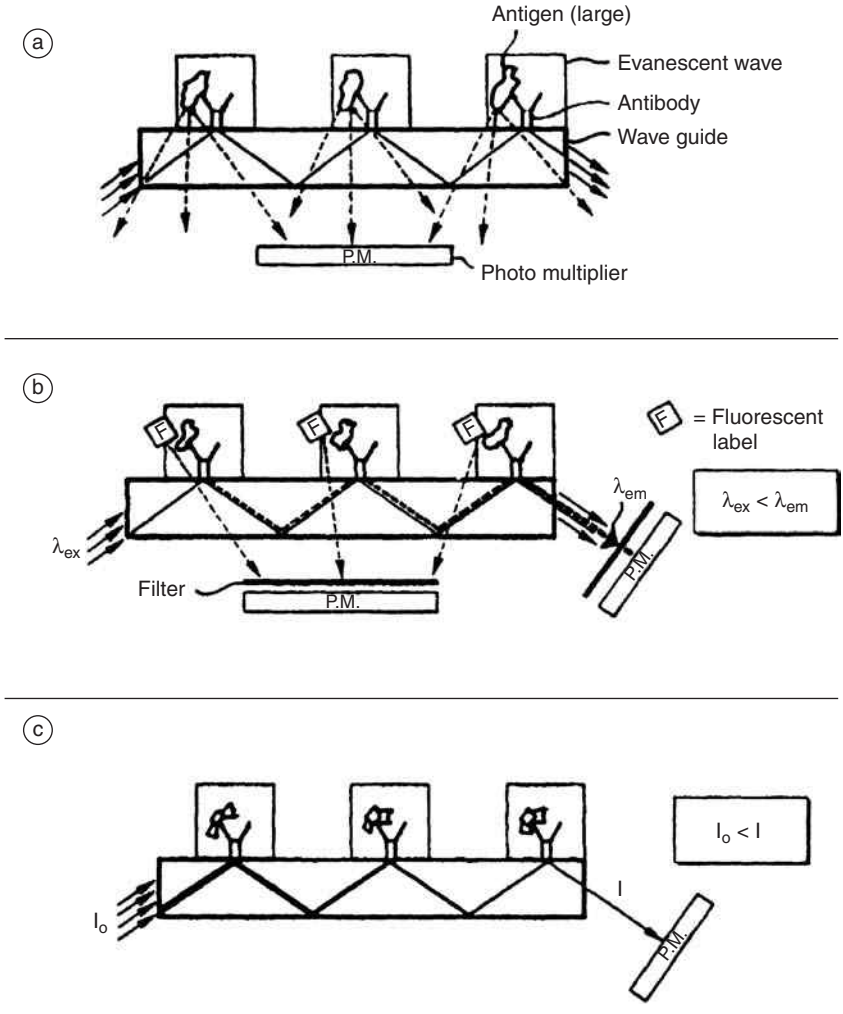


Figure 7.36 (a) Light scattering from the waveguide surface increases as the antigen-antibody complexes are formed, and is detected by a side-mounted photodetector. (b) Fluorescence is excited in a fluorescent marker, attached to an antigen molecule, by light passing through the waveguide. The fluorescent light can be collected either sideways from the guide, or as light that is retrapped by the guide and directed to a photodetector. (c) Absorption of light by antibody-antigen complexes on the surface attenuates light traveling down the waveguide.

where A is the optical absorbance, l is the path length of the light, η is the molar absorptivity, and I_0 and I are the incident and transmitted light, respectively. For absorption measurements via optical fibers, the medium normally must be optically transparent.

An accurate method for the detection of leakage of flammable gases such as methane (CH_4), propane (C_3H_8), and ethylene (C_2H_4) is vital in gas and petrochemical plants in order to avoid serous accidents. The recent introduction of low-loss fiber into spectroscopic measurements of these gases offers many advantages for process control in manufacturing:

- Long-distance remote sensing
- On-line measurement and monitoring
- Low cost
- High reliability

The most commonly used method at present is to carry the sample back to the measuring laboratory for analysis; alternatively, numerous spectrometers may be used at various points around the factory. The new advances in spectroscopic measurements allow even CH_4 to be observed at a distance of 10 km with a detection sensitivity as low as 5 percent of the *lower explosion limit* (LEL) concentration. The optical-fiber gas measuring system employs an absorption spectroscopy technique, with the light passing through a gas-detection cell for analysis. The overtone absorption bands of a number of flammable gases are located in the near-infrared range (Fig. 7.37).

The optical gas sensing system can deal with a maximum of 30 detection cells (Fig. 7.38). The species to be measured are CH_4, C_3H_8, and C_2H_4 molecules. Light from a halogen lamp (an infrared light source) is distributed into a bundle of 30 single optical fibers. Each of

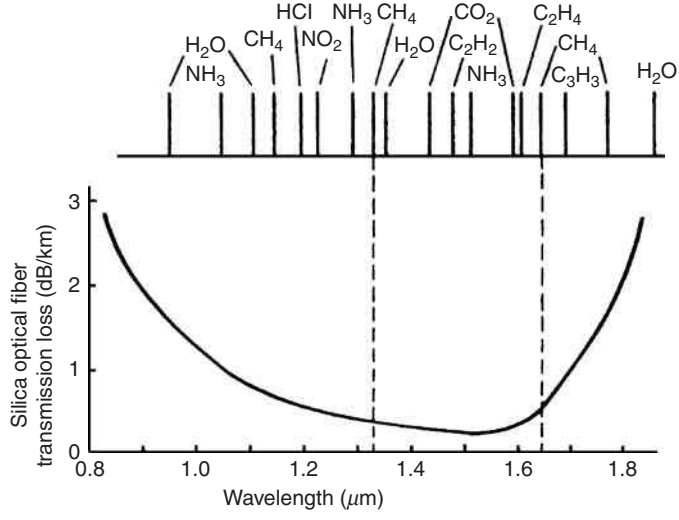


Figure 7.37 Absorption lines of typical flammable gases in the near-infrared, and the transmission loss of silica fiber.

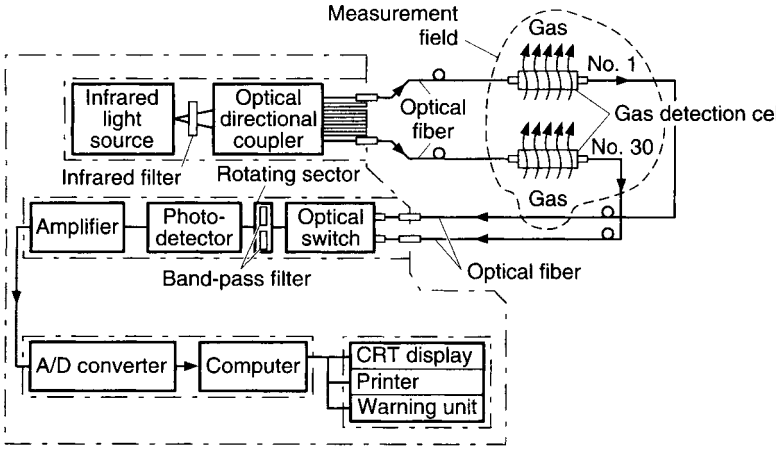


Figure 7.38 A gas detection system with 30 detection cells.

the distributed beams is transmitted through a 1-km length of fiber to a corresponding gas detection cell. The receiving unit is constructed of three optical switches, a rotating sector with four optical interference filters, and three Ge photodiodes. Each optical switch can select any ten returned beams by specifying the number of the cell. The peak transmission wavelength of the optical filter incorporated in the sensor is 1.666 μm for CH_4, 1.690 μm for C_3H_8, 1.625 μm for C_2H_2, and 1.600 μm for a reference beam. After conversion to electrical signals, the signal amplitudes for the three gases are normalized by the reference amplitude. Then, the concentration of each gas is obtained from a known absorption-concentration calibration curve stored in a computer.

An intrinsic distributed optical-fiber gas sensor for detecting the leakage of cryogenically stored gases such as CH_4, C_2H_4, and N_2 has also been developed. The sensor's operation is based on the temperature-dependent transmission loss of optical fiber—that is, the optical fiber is specially designed so that the transmission loss increases with decreasing temperature by choosing the appropriate core and cladding materials. Below the critical temperature, in the region of –55°C, most of the light has transferred to the cladding layer, and the light in the core is cut off. By connecting this temperature-sensitive fiber between a light source and a detector and monitoring the output light level, the loss of light resulting from a cryogenic liquid in contact with the fiber can be detected directly.

7.19 Using Avalanche Photodiodes to Improve System Performance

Avalanche photodiode detectors (APD) are used in many diverse applications, such as laser range finders (Fig. 7.39), data communications,

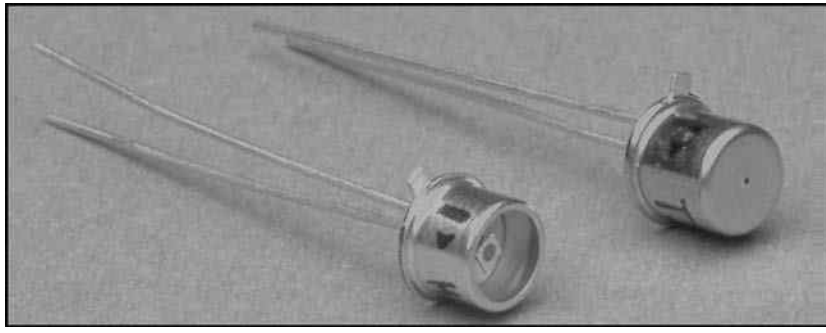


Figure 7.39 Avalanche photodiode detectors (APD)

and photon correlation studies. This section of the chapter discusses APD structures, critical performance parameters, and excess noise factors. For low-light detection in the 200- to 1150-nm range, the designer has three basic detector choices: the silicon PIN detector, the *silicon avalanche photodiode* (APD), and the *photomultiplier tube* (PMT).

APDs are widely used in instrumentation and aerospace applications, offering a combination of high speed and high sensitivity that is unmatched by PIN detectors, and quantum efficiencies at >400 nm that is unmatched by PMTs. Providing a guideline to defining and selecting avalanche photodiodes should offer the user of spectroscopy sensing devices the fundamental tools needed to apply minimum signal-to-noise spectra and would include the following:

- Explanation of an avalanche photodiode
- How to select an APD
- Discussion of excess noise factors
- Explanation of Geiger Mode
- Possible applications

7.20 Structures of Avalanche Photodiodes—APD Structures

In order to understand why more than one APD structure exists, it is important to appreciate the design tradeoffs that must be accommodated by the APD designer. The ideal APD would have zero dark noise, no excess noise, a broad spectral and frequency response, a gain range from 1 to 106 or more, and a low cost. More simply, an ideal APD would be a good PIN photodiode with gain! In reality, however, this is difficult to achieve because of the need to trade off

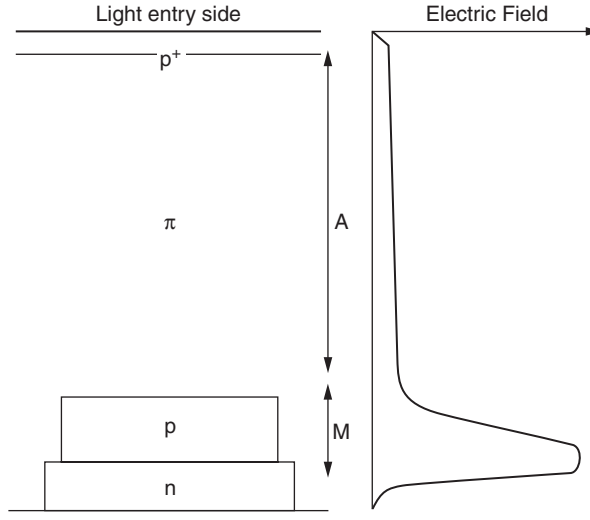


Figure 7.40 Schematic cross-section for a typical APD structure.

conflicting design requirements. What some of these tradeoffs are, and how they are optimized in commercially available APDs, are listed next.

Consider the schematic cross-section for a typical APD structure (Fig. 7.40). The basic structural elements provided by the APD designer include an absorption region A, and a multiplication region M. Present across region A is an electric field E that serves to separate the photogenerated holes and electrons, and sweeps one carrier towards the multiplication region. The multiplication region M is designed to exhibit a high electric field so as to provide internal photocurrent gain by impact ionization.

This gain region must be broad enough to provide a useful gain, M, of at least 100 for silicon APDs, or 10–40 for germanium or InGaAs APDs. In addition, the multiplying electric field profile must enable effective gain to be achieved at field strength below the breakdown field of the diode.

Figure 7.40 shows the "reach-through" structure, which provides a combination of high speed, low noise, capacitance, and extended IR response.

The multiplication region M is designed to exhibit a high electric field so as to provide internal photocurrent gain by impact ionization.

A variation of the reach-through structure is *epitaxial silicon APD* (EPI-APD). The EPI-APD allows simpler and cheaper manufacturing methods, resulting in a lower-cost device relative to the standard reach-through device. The tradeoff is a higher k factor, resulting in a higher excess noise factor for equivalent gain (Table 7.3).

Detector Type	Ionization Ratio	X-Factor	Typical Gain	Excess Noise (Factor at Typical Gain)
	(k)	—	(M)	(F)
Silicon (reach-through structure)	0.02	0.3	150	4.9
Silicon Epitaxial APDs	0.06	0.45	100	7.9
Silicon (SLiK low-k structure)	0.002	0.17	500	3.0
Germanium	0.9	0.95	10	9.2
InGaAs	0.45	0.7–0.75	10	5.5

Table 7.3 Typical Values of k, X, and F for Si, Ge, and InGaAs APDs

7.21 Critical Performance Parameters

An APD differs from a PIN photodiode by providing internal photoelectronic signal gain. Therefore, output signal current, I_S, from an APD equals I_s = M.Ro (λ) × P_s, where Ro(λ) is the intrinsic responsivity of the APD at a gain M = 1 and wavelength l, M is the gain of the APD, and P_s is the incident optical power. The gain is a function of the APD's reverse voltage, V_R, and will vary with applied bias. A typical gain-voltage curve for a silicon APD is shown in Fig. 7.41.

The gain as a function of the bias voltage varies with the structure of the APD.

One of the key parameters to consider when selecting an APD is the detector's spectral noise. Like other detectors, an APD will normally be operating in one of two noise-limited detection regimes; either detector noise limited at low power levels, or photon shot noise limited at higher powers. Since an APD is designed to be operated under a reverse bias, sensitivity at low light levels will be limited by the shot noise and the APD's leakage current. Shot noise derives from the random statistical Poissonian fluctuations of the dark current, I_D (or signal current). Dark current shot noise (IN - SHOT) is normally given by $I_{N(SHOT)} = (2.q.B.I_D)^{½}$, for a PIN detector, where B is the system bandwidth. This differs for an APD, however, as bulk leakage current, I_{DB}, is multiplied by the gain, M, of the APD[4]. Total leakage current I_D is therefore equal to:

$$I_D = I_{DS} + I_{DB} \cdot M \tag{1}$$

where I_{DS} is the surface leakage current.

In addition, the avalanche process statistics generate current fluctuations, and APD

performance is degraded by an "excess noise factor" (F) compared to a PIN, equation 4. The total spectral noise current for an APD in "dark" conditions is thus given by equation (1), ID = IDS + IDB · M: where I_{DS} is the surface leakage current.

$$i_n = [2.q \cdot I_{DS} + I_{DB} \cdot M \cdot F \cdot B]^{0.5} \quad (2)$$

where q is the electron charge.

At higher signal light levels, the detector transitions to the photon shot noise limited regime where sensitivity is limited by photon shot noise on the current generated by the optical signal. Total noise from the APD in "illuminated" conditions will therefore equal the quadratic sum of the detector noise plus the signal shot noise. For a given optical signal power, P_s, this is given by:

$$i_n = [2.q \cdot I_{DS} + I_{DB} \cdot M + R(\lambda) \cdot M \cdot P_s \cdot F \cdot B]^{0.5} \quad (3)$$

In the absence of other noise sources, an APD therefore provides a signal-to-noise ratio (SNR), which is worse than a PIN detector with the same quantum efficiency. Noise equivalent power (NEP) cannot be used as the only measure of a detector's relative performance, but rather detector signal-to-noise (SNR) at a specific wavelength and bandwidth should be used to determine the optimum detector type for a given application. The optimum signal-to-noise occurs at a gain M, where total detector noise equals the input noise of the amplifier or load resistor. The optimum gain depends in part on the excess noise factor, F, of the APD, and ranges from M = 50 to 1000 for silicon APDs, and is limited to M = 10 to 40 for germanium and InGaAs APDs.

7.22 Selecting an APD

APDs are generally recommended for high-bandwidth applications, or where internal gain is needed to overcome high preamplifier noise.

The following is a simple guide that can be used to decide if an APD is the most appropriate for one's light detection requirements.

- Determine the wavelength range to be covered. (See the following section "Types of APDs," to determine the specific APD type useful for the wavelength range to be covered.)
- Determine the minimum size of the detector that can be used in the optical system. Effective optics can often be more cost-effective than the use of an overly large PIN or avalanche photodetectors.
- Determine the required electrical frequency bandwidth of the system; again, over-specifying bandwidth will degrade the SNR of the system.

7.22.1 Types of APDs

Avalanche photodiodes are commercially available that span the wavelength range from 300 nm to 1700 nm. Silicon APDs can be used between 300 nm to 1100 nm, germanium between 800 nm and 1600 nm, and InGaAs from 900 nm to 1700 nm. Although significantly more expensive than germanium APDs, InGaAs APDs are typically available with much lower noise currents, exhibit an extended spectral response to 1700 nm, and provide higher frequency bandwidth for a given active area.

A germanium APD is recommended for applications in high electromagnetic interference (EMI) environments, where amplifier noise is significantly higher than the noise from an InGaAs APD, or for applications where cost is a primary consideration.

7.22.2 Understanding the Specifications Responsivity and Gain

APD gain will vary as a function of applied reverse voltage, as shown in Fig. 7.41. In addition, for many APDs, it is not possible, or practical, to make an accurate measurement of the intrinsic responsivity: Ro(λ), at a gain M = 1.

It is therefore inappropriate to state typical gain and diode sensitivity at M = 1 as a method for specifying diode responsivity at a given operating voltage.

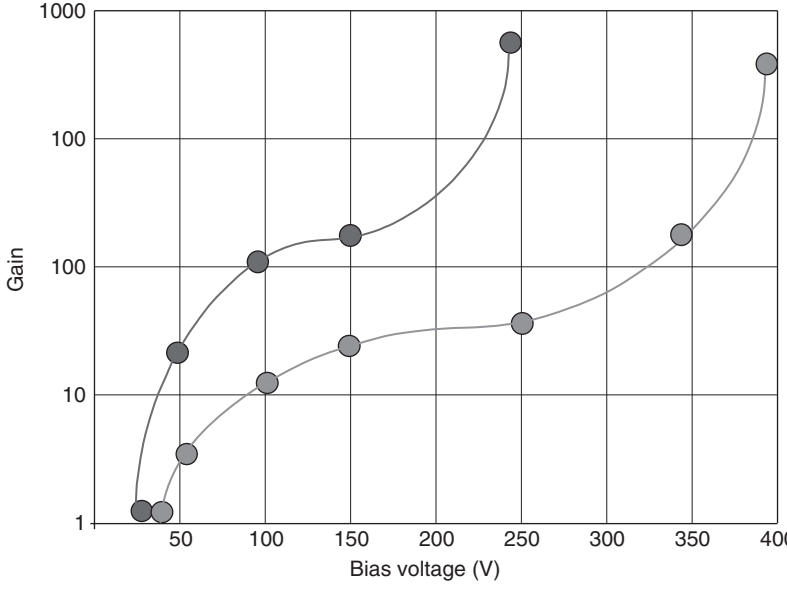


Figure 7.41 A typical gain-voltage curve for a silicon APD.

In order to characterize APD response, one must specify APD responsivity (in amps/watt) at a given operating voltage. However, because of diode-to-diode variations in the exact gain voltage curve of each APD, the specific operating voltage for a given responsivity will vary from one APD to another.

Manufacturers should therefore specify a voltage range within which a specific responsivity will be achieved.

An example of a typically correct specification for diode responsivity, in this case for an InGaAs APD, is as follows:

$$R_{MIN}\ (1300\ nm) = 9.0\ A/W,$$
$$V_{OP} = 50\ V \text{ to } 90\ V,$$
$$M = 10$$

7.23 Dark Current and Noise Current

As can be seen from the noise equation (Equation 2), the total APD dark current (and the corresponding spectral noise current) is only meaningful when specified at a given operating gain.

Dark current at M = 1 is dominated by surface current, and may be significantly less than

$$I_{DB} \times M$$

Since APD dark and spectral noise currents are a strong function of APD gain, these should be specified at a stated responsivity level.

An example of a typically correct specification for diode dark current and noise current—in this case, for an InGaAs APD—is as follows:

$$I_D\ (R = 9.0A/W) = 10\ nA\ (max),$$
$$M = 10$$
$$i_N\ (R = 6.0\ A/W,\ 1\ MHz,\ 1\ Hz\ BW) = 0.8\ pA/\sqrt{Hz}\ (max),$$
$$M > 5$$

7.23.1 Excess Noise Factor

All avalanche photodiodes generate excess noise due to the statistical nature of the avalanche process. This excess noise factor is generally denoted as *F*.

As shown in the noise equation (Equation 2), $\sqrt{F}$ is the factor by which the statistical noise on the APD current (equal to the sum of the multiplied photocurrent plus the multiplied APD bulk dark current)

exceeds that which would be expected from a noiseless multiplier on the basis of Poissonian statistics (shot noise) alone.

The excess noise factor is a function of the carrier ionization ratio, k,

where (k) is usually defined as the ratio of the hole to electron ionization probabilities. The excess noise factor may be calculated using the model developed by McIntyre (3), which considers the statistical nature of

avalanche multiplication. The excess noise factor is given by:

$$F = k_{EFF} \cdot M + (1 - k_{EFF})(1 - 1/M) \quad (4)$$

Therefore, the lower the values of k and M, the lower the excess noise factor.

The effective k factor (k_{EFF}) for an APD can be measured experimentally by fitting the McIntyre formula to the measured dependence of the excess noise factor on gain. This is best done under illuminated conditions.

It may also be theoretically calculated from the carrier ionization coefficients and the electric field profile of the APD structure. The ionization ratio k is a strong function of the electric field across the APD structure, and takes its lowest value at low electric fields (only in silicon). Since the electric field profile depends upon the doping profile, the k factor is also a function of the doping profile. Depending on the APD structure, the electric field profile traversed by a photogenerated carrier and subsequent avalanche-ionized carriers may therefore vary according to photon absorption depth. For indirect band gap semiconductors such as silicon, the absorption coefficient varies slowly at the longer wavelengths, and the "mean" absorption depth is therefore a function of wavelength.

The value of k_{EFF}, and gain, M, for a silicon APD is thus a function of wavelength for some doping profiles.

The McIntyre formula can be approximated for a $k < 0.1$ and $M > 20$ without significant loss of accuracy as:

$$F = 2 + k \cdot M \quad (5)$$

Also often quoted by APD manufacturers is an empirical formula used to calculate the excess noise factor, given as:

$$F = M^x \quad (6)$$

where the value of X is derived as a log-normal linear fit of measured F-values for given values of gain M.

This approximation is sufficiently appropriate for many applications, particularly when used with APDs with a high k factor, such as InGaAs and germanium APDs. Table 7.3 provides typical values of k, X, and F for silicon, germanium, and InGaAs APDs.

The silicon APDs have three different values of ionization ration:

- The Super-Low k (SLiK) APDs used in photon counting modules
- The high-performance reach-through structures with k = 0.02 for applications requiring extremely low noise APD and high gain
- The low-cost silicon epitaxial APD with k = 0.06 ideal for high SNR applications.

In germanium and InGaAs APDs, the k-value is generally quoted at M = 10, which somewhat overestimates F at M < 10 and underestimates F at M > 10.

7.24 The Geiger Mode

In the Geiger mode, an APD is biased above its breakdown voltage ($V_R > V_{BR}$) for operation at very high gain (typically 10^5 to 10^6).

When biased above breakdown, an APD will normally conduct a large current. However, if this current is limited to less than the APD's "latching" current, there is a strong statistical probability that the current will fluctuate to zero in the multiplication region, and the APD will then remain in the "off" state until an avalanche pulse is triggered by either a bulk or photogenerated carrier. If the number of bulk carrier–generated pulses is low, the APD can therefore be used to count individual current pulses from incident photons. The value of the bulk dark current is therefore a significant parameter in selecting an APD for photon-counting, and can be reduced exponentially by cooling.

7.25 Crack Detection Sensors for Commercial, Military, and Space Industry Use

Accurate and precise detection of crack propagation in aircraft components is of vital interest for commercial and military aviation and the space industry. A system has been recently developed to detect cracks and crack propagation in aircraft components. This system uses optical fibers of small diameter (20 to 100 μm), which can be etched to increase their sensitivity. The fibers are placed on perforated adhesive foil to facilitate attachment to the desired component for testing. The fiber is in direct contact with the component (Fig. 7.42). The foil is removed after curing of the adhesive. Alternatively, in glass-fiber-reinforced plastic (GFRP) or carbon-fiber-reinforced plastic (CFRP), materials that are used more and more in aircraft design, the fiber can be easily inserted in the laminate without disturbing the normal fabrication process. For these applications, bare single fiber or

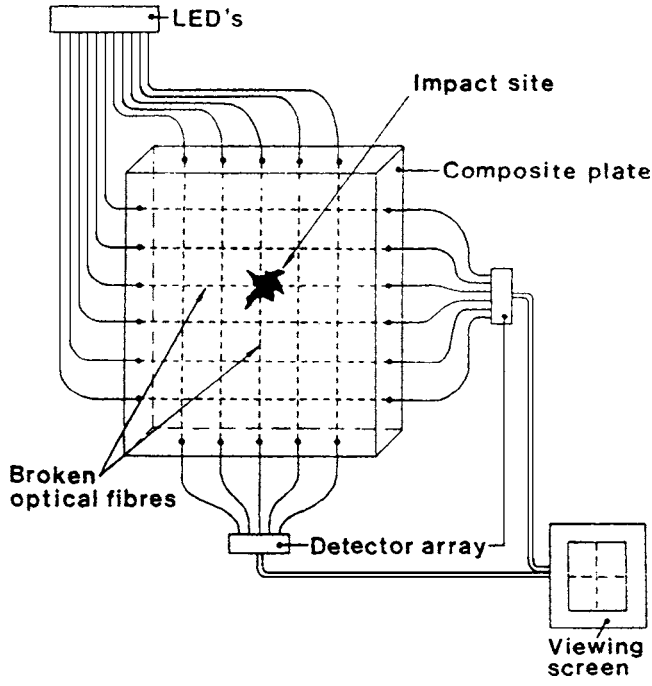


Figure 7.42 An illustration of fiber to be in direct contact with the component.

prefabricated tape with integrated bundles of fibers is used. The system was initially developed for fatigue testing of aircraft components such as frames, stringers, and rivets. In monitoring mode, the system is configured to automatically interrupt the fatigue test. The system has also been applied to the inspection of the steel rotor blades of a 2-MW wind turbine. A surveillance system has been developed for the centralized inspection of all critical components of the Airbus commercial jetliner during its lifetime. This fiber nervous system is designed for in-flight monitoring and currently is accessible to flight and maintenance personnel.

An optical-fiber mesh has been tested for a damage assessment system for a GFRP submarine sonar dome. Two sets of orthogonally oriented fibers are nested in the laminate during the fabrication process. When the fibers of the mesh are properly connected to LEDs and the detectors, the system can be configured to visualize the location of a damaged area.

As an alternative, a video camera and image processing are applied to determine the position of the damaged area. The fiber end faces at the detection side of the mesh are bundled and imaged into the camera tube. Two images are subtracted: the initial image before the occurrence of damage and the subsequent image. If fibers are broken, their location is highlighted as a result of this image subtraction.

7.26 Control of the Input/Output Speed of Continuous Web Fabrication Using Laser Doppler Velocity Sensor

A laser Doppler velocimeter (LDV) can be configured to measure any desired component velocity, perpendicular or parallel to the direction of the optical axis. An LDV system has been constructed with a semiconductor laser and optical fibers and couplers to conduct the optical power. Frequency modulation of the semiconductor laser (or, alternatively, an external fiber-optic frequency modulator) is used to introduce an offset frequency. Some commercial laser Doppler velocimeters are available with optical-fiber leads and small sensing heads. However, these commercial systems still use bulk optical components such as acoustooptic modulators or rotating gratings to introduce the offset frequency.

With an LDV system, the velocity can be measured with high precision in a short period of time. This means that the method can be applied for real-time measurements to monitor and control the velocity of objects as well as measure their vibration. Because the laser light can be focused to a very small spot, the velocity of very small objects can be measured, or if scanning techniques are applied, high spatial resolution can be achieved. This method is used for various applications in manufacturing, medicine, and research. The demands on system performance with respect to sensitivity, measuring range, and temporal resolution are different for each of these applications.

In manufacturing processes, for example, LDV systems are used to control continuous roll milling of metal (Fig. 7.43), control the rolling speed of paper and films, and monitor fluid velocity and turbulence in mixing processes. Another industrial application is vibration analysis. With a noncontact vibrometer, vibration of machines, machine tools, and other structures can be analyzed without disturbing the vibrational behavior of the structure.

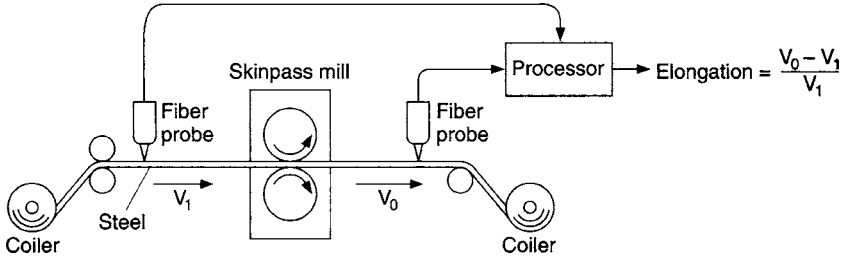


Figure 7.43 A fiber-optic laser Doppler velocimeter at a rolling mill controls pressure by measuring input speeds.

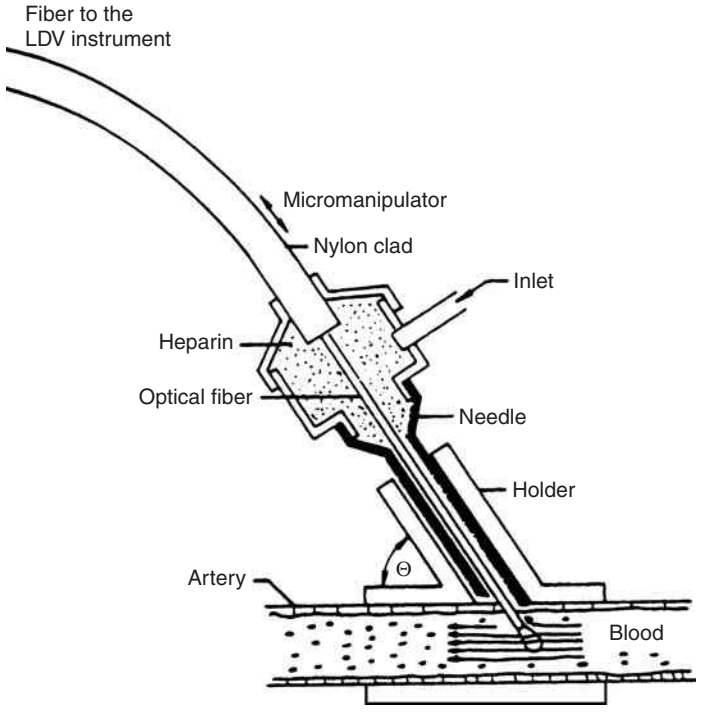


Figure 7.44 Special probe for measurement of blood velocity.

Interestingly, the LDV system proved useful in the measurement of arterial blood velocity (Fig. 7.44), thereby providing valuable medical information. Another application in medical research is the study of motion of the tympanic membrane in the ear.

7.27 Ultrasonic/Laser Nondestructive Evaluation Sensor

Ultrasonic/laser optical inspection is a relatively new noncontact technique. A laser system for generating ultrasound pulses without distortion of the object surface is shown in Fig. 7.45. A laser pulse incident on a surface will be partly absorbed by the material and will thus generate a sudden rise in temperature in the surface layer of the material. This thermal shock causes expansion of a small volume at the surface, which generates thermoelastic strains. Bulk optical systems have been used previously to generate the laser pulse energy. However, the omnidirectionality of bulk sources is completely different from other well-known sources, and is regarded as a serious handicap to laser generation.

To control the beamwidth and beam direction of the optically generated ultrasonic waves, a fiber phased array has been developed.

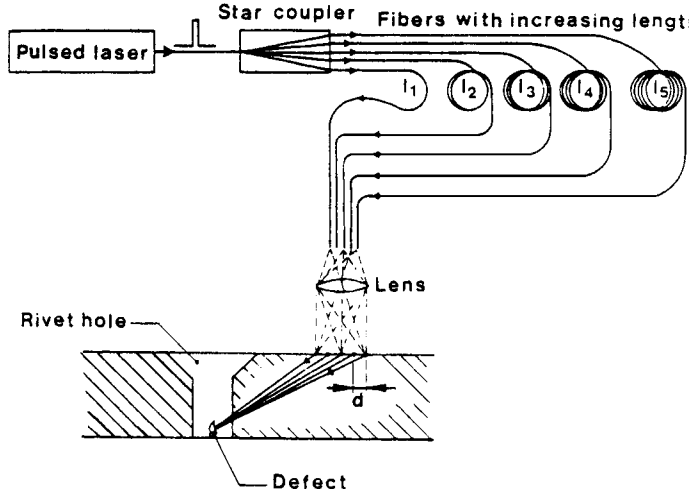


Figure 7.45 Setup for beam steering of a laser-generated ultrasound by fiber-optic phased array.

In this way, the generated ultrasonic beam can be focused and directed to a particular inspection point below the surface of an object (Fig. 7.45). This system has been optimized for the detection of fatigue cracks at rivet holes in aircraft structures.

The combination of laser-generated ultrasound and an optical-fiber interferometer for the detection of the resultant surface displacement has led to a technique that is useful for a wide variety of inspection tasks in manufacturing, including high-temperature objects and areas that are difficult to access, as well as more routine inspection and quality control in various industrial environments. Such a system can be applied to the measurement of thickness, velocity, flaws, defects, and grain size in a production process.

7.28 Process Control Sensors for Acceleration

The principle of operation of the process control acceleration sensor is illustrated in Fig. 7.46. The sensor element, consisting of a small cantilever and a photoluminescent material, is attached to the end of a single multimode fiber. The input light of wavelength λ_s is transmitted along the fiber from a near-infrared LED source to the sensor element. The sensor element returns light at two different wavelengths, one of which serves as a signal light and the other as a reference light, into the same fiber. The signal light at wavelength λ_s is generated by reflection from a small cantilever. Since the relative angle of the reflected light is changed by the acceleration, the returned light is intensity-modulated. The reference light of wavelength λ_r is generated

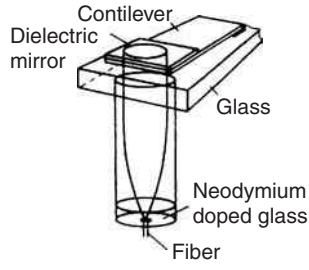


Figure 7.46 Cantilever-type acceleration sensor.

by photoluminescence of a neodymium-doped glass element placed close to the sensor end of the fiber.

The optoelectronic detector module has two optical filters to separate the signals λ_s and λ_r and two photodiodes to convert the signal and the reference light into separate analog voltages. The signal processing for compensation is then merely a matter of electrical division. A measuring range of 0.1 to 700 m/s^2 and a resolution of 0.1 m/s^2 is obtained over the frequency range of 5 to 800 Hz.

7.29 An Endoscope as Image Transmission Sensor

An imaging cable consists of numerous optical fibers, typically 3000 to 100,000, each of which has a diameter of 10 μm and constitutes a picture element (pixel). The principle of image transmission through the fibers is shown in Fig. 7.47. The optical fibers are aligned regularly and identically at both ends of the fibers. When an image is projected on one end of the image fiber, it is split into multiple picture elements. The image is then transmitted as a group of light dots with different intensities and colors, and the original picture is reduced at the far end. The image fibers developed for industrial use are made of silica glass with low transmission loss over a wide wavelength band from visible to near infrared, and can therefore transmit images over distances in excess of 100 m without significant color changes. The basic structure of the practical optical-fiber image sensing system (endoscope) is illustrated in Fig. 7.48. It consists of the image fiber, an objective lens to project the image on one end, an eyepiece to magnify the received image on the other end, a fiber protection tube, and additional fibers for illumination of the object.

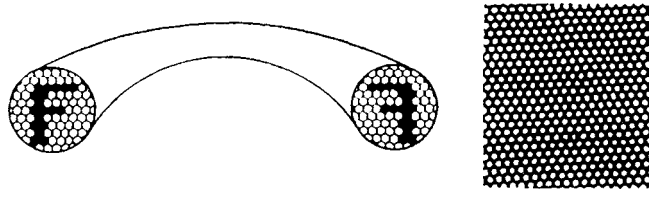


Figure 7.47 Image transmission through an image fiber.

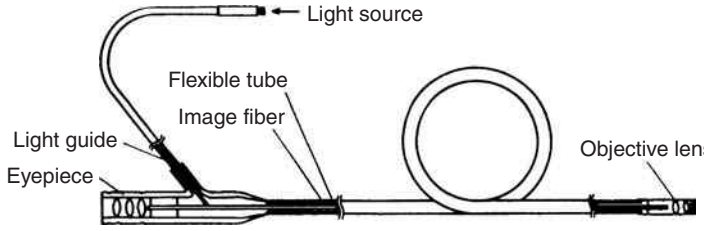


Figure 7.48 Basic structure of a fiber scope.

Many examples have been reported of the application of image fibers in process control. Image fibers are widely employed to observe the interior of blast furnaces and the burner flames of boilers, thereby facilitating supervisory control. Image fibers can operate at temperatures up to 1000°C, when provided with a cooling attachment for the objective lens and its associated equipment. Another important application of the image fiber bundles is observation, control, and inspection of nuclear power plants and their facilities. Conventional image fibers cannot be used within an ionizing radiation environment because ordinary glass becomes colored when exposed to radiation, causing increasing light transmission loss. A high-purity silica core fiber is well-known as a radiation-resistant fiber for nuclear applications.

The endoscope has demonstrated its vital importance in medical and biochemical fields such as:

- Angioplasty
- Laser surgery
- Gastroscopy
- Cystoscopy
- Bronchoscopy
- Cardioscopy

7.30 Sensor Network Architectures in Manufacturing

In fiber-optic sensor networks, the common technological base with communication is exploited by combining the signal generating ability of sensors and the signal transmitting capability of fiber optics. This combination needs to be realized by a suitable network topology in various manufacturing implementations. The basic topologies for sensor networking are illustrated in Fig. 7.49. The basic network topologies are classified into six categories:

- Linear array network with access-coupled reflective sensors (Fig. 7.49*a*).
- Ring network with in-line transmissive sensors (Fig. 7.49*b*).

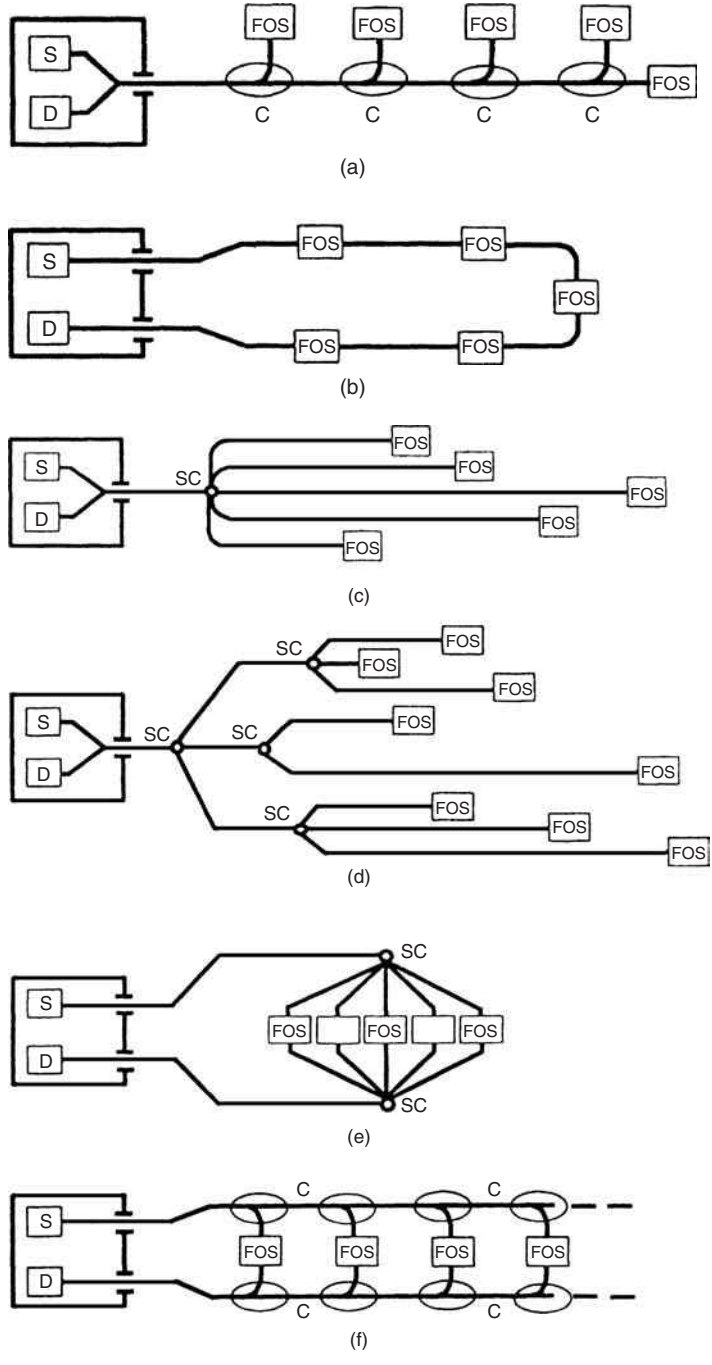


Figure 7.49 Basic network topologies: (a) linear array, (b) ring, (c) reflective star, (d) reflective tree, (e) transmissive star, and (f) ladder network.

- Star network with reflective sensors (Fig. 7.49*c*).
- Star network with reflective sensors; one or more sensors can be replaced by a separate star network, in order to obtain a tree network (Fig. 7.49*d*).
- Star network that can also be operated with transmissive sensors (Fig. 7.49*e*).
- Ladder network with two star couplers. A star coupler is replaced by several access couplers, the number required being equal to the number of sensors (Fig. 7.49*f*).

Topological modifications, especially of sensor arrays and ladder networks, may be desirable in order to incorporate reference paths of transmissive (dummy sensors) or reflective sensors (splices, open fiber end).

The transmit and return fibers, or fiber highway, generally share a common single-fiber path in networks using reflective sensors.

When a suitable fiber-optic network topology is required, various criteria must be considered:

- The sensor type, encoding principle, and topology to be used
- The proposed multiplexing scheme, required number of sensors, and power budget
- The allowable cross-communication level
- The system cost and complexity constraints
- The reliability (i.e., the effect of component failure on system performance)

7.31 Power Line Fault-Detection Systems for Power Generation and Distribution Industries

In power distribution lines, faults such as short circuits, ground faults, and lightning strikes on the conductors must be detected in a very short time to prevent damage to equipment and power failure, and to enable quick repair. If the transmission line is divided in sections and a current or magnetic-field sensor is mounted in each section, a faulty section can be determined by detection of a change of the level and phase of the current on the power line. A system was developed as a hybrid optical approach to a fault-locating system that detects the phase and current difference between two current transformers on a composite fiber-optic ground wire (OPGW) wherein, due to induction, current is constantly passing (Fig. 7.50). The signal from a local electrical sensor, powered by solar cells and batteries, is transmitted over a conventional optical-fiber communication link. By three-wavelength

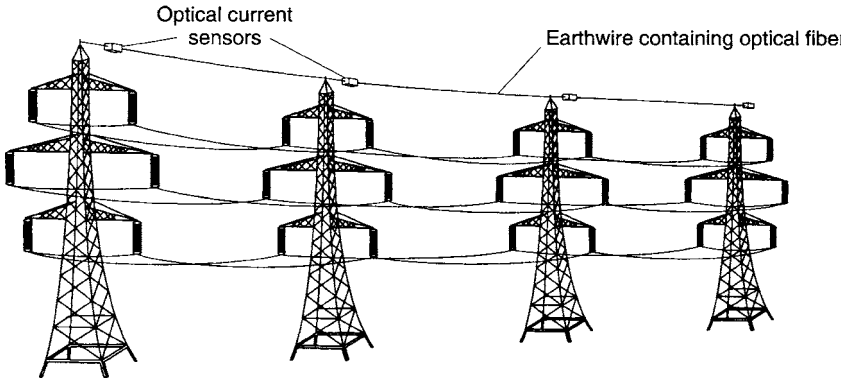


Figure 7.50 Fault locating system based on current and phase in ground wire.

multiplexing, three sensor signals can be transmitted over a single fiber. Seven sensors, three at each side of a substation and one at the substation itself, can monitor one substation on the power line, using one fiber in the OPGW.

Another system uses current transformers to pick up lightning current and thus detect lightning strikes. The signal is transmitted to a central detection point using the OPGW. Every sensor has its own OPGW fiber. This system is on a 273-kV power line in Japan.

The OPGW opens the possibility for using all kinds of sensors along the transmission line. These sensors may not only be used for locating faults, but also for monitoring structural integrity. The use of optical time-domain reflectometry (OTDR) combined with passive intrinsic (distributed) sensors along the OPGW has future potential for providing a convenient and powerful monitoring method for power lines.

Further Reading

Bailey Control Systems, Wickliffe, Ohio.

Bartman, R. K., B. R. Youmans, and N. M. Nerheim. "Integrated Optics Implementation of a Fiber Optic Rotation Sensor: Analysis and Development," *Proc. SPIE*, 719, 122–134.

Berthold, J. W., "Industrial Applications of Optical Fiber Sensors," Fiber Optic and Laser Sensors III, *Proc. SPIE*, 566, 37–44.

Carrol, R., C. D. Coccoli, D. Cardelli, and G. T. Coate, "The Passive Resonator Fiber Optic Gyro and Comparison to the Interferometer Fiber Gyro," *Proc. SPIE*, 719, 169–177 (1986).

Chappel, A. (ed.), *Optoelectronics—Theory and Practice*, McGraw-Hill, New York, 1978.

Crane, R. M., A. B. Macander, D. W. Taylor, and J. Gagorik, "Fiber Optics for a Damage Assessment System for Fiber Reinforced Plastic Composite Structures," *Rev. Progress in Quantitative NDE*, 2B, Plenum Press, New York, 1419–1430, 1982.

Doeblin, E. O., *Measurement Systems—Application and Design*, 4th ed., McGraw-Hill, New York, 1990.

Fields, J. N., C. K. Asawa, O. G. Ramer, and M. K. Barnoski, "Fiber Optic Pressure Sensor," *J. Acoust. Soc. Am.*, 67, 816 (1980).

Finkelstein, L., and R. D. Watts, "Fundamental of Transducers—Description by Mathematical Models," *Handbook of Measurement Science*, vol. 2, P. H. Sydenham (ed.), Wiley, New York, 1983.

Friebele, E. L. and M. E. Gingerich, "Radiation-Induced Optical Absorption Bands in Low Loss Optical Fiber Waveguides," *J. Non-Crust. Solids*, 38(39), 245–250 (1980).

Henze, M., "Fiber Optics Temperature and Vibration Measurements in Hostile Environments," Technical Material, *ASEA Research and Innovation*, CF23-1071E (1987).

Hofer, B., "Fiber Optic Damage Detection in Composite Structure," *Proc. 15th Congress. Int. Council Aeronautical Science*, ICAS-86-4.1.2, 135–143 (1986).

Kapany, N. S., *Fiber Optics, Principles and Applications*, Academic Press, London, 1976.

Lagakos, N., et al., "Multimode Optical Fiber Displacement Sensor," *Appl. Opt.*, 20, 167 (1981).

Liu, K., "Optical Fiber Displacement Sensor Using a Diode Transceiver," Fiber Optic Sensors II, A. M. Sheggi (ed.), *Proc. SPIE*, 798, 337–341 (1987).

Mizuno, Y., and T. Nagai, "Lighting Observation System on Aerial Power Transmission Lines by Long Wavelength Optical Transmission," *Applications of Fiber Optics in Electrical Power Systems in Japan*, C.E.R.L. Letterhead, paper 5.

Mori, S., et al., "Development of a Fault-Locating System Using OPGW," *Simitomo Electric Tech. Rev.* (25), 35–47.

Norton, H. N., *Sensors and Analyzer Handbook*, Prentice-Hall, Englewood Cliffs, N.J., 1982.

Neubert, H.K.P., *Instrument Transducers*, 2d ed., Clarendon Press, Oxford, 1975.

Ogeta, K., *Modern Control Engineering*, 2d ed. Prentice-Hall, Englewood Cliffs, N.J., 1990.

Petrie, G. R., K. W. Jones, and R. Jones, "Optical Fiber Sensors in Process Control," *4th Int. Conf. Optical Fiber Sensors, OFS'86*, Informal Workshop at Tsukuba Science City, VIII I–VIII 19, 1986.

Place, J. D., "A Fiber Optic Pressure Transducer Using A Wavelength Modulation Sensor," *Proc. Conf. Fiber Optics '85 (Sira)*, London, 1985.

Ramakrishnan, S., L. Unger, and R. Kist, "Line Loss Independent Fiberoptic Displacement Sensor with Electrical Subcarrier Phase Encoding," 5*th Int. Conf. Optical Fiber Sensors, OFS '88*, New Orleans, 133–136, 1988.

Sandborn, V. A., *Resistance Temperature Transducers*, Metrology Press, Fort Collins, Colo., 1972.

Scruby, C. B., R. J. Dewhurst, D. A. Hutchins, and S. B. Palmer, "Laser Generation of Ultrasound in Metals," *Research Techniques in Nondestructive Testing*, vol. 15, R. S. Sharpe (ed.), Academic Press, London, 1982.

Tsumanuma, T., et al., "Picture Image Transmission-System by Fiberscope," *Fujikura Technical Review* (15), 1–10 (1986).

Vogel, J. A., and A. J. A. Bruinsma, "Contactless Ultrasonic Inspection with Fiber Optics," Conf. *Proc. 4th European Conf. Non-Destructive Testing*, Pergamon Press, London, 1987.

Yasahura, T., and W. J. Duncan, "An Intelligent Field Instrumentation System Employing Fiber Optic Transmission," *Advances in Instrumentation*, ISA, Wiley, London, 1985.

Cova, S. et al., " Avalanche Photodiodes for Near Infrared Photon-Counting," *SPIE Proc.*, vol. 2388 (1995).

Dautet, H., et al., "Photon-Counting Techniques with Silicon Avalanche Photodiodes," *Applied Optics*, 32 (21), 3894–3900 (1993).

Mclntyre, R. J., "Multiplication Noise in Uniform Avalanche Diodes," *IEEE Trans. Electron Devices*, ED13, 164–168 (1966).

Product Datasheet ED-0017/03/88: C30902 - 921 E/S, EG&G Canada Ltd.

This page intentionally left blank

CHAPTER 8

Sensors in Flexible Manufacturing Systems

8.1 Introduction

Flexibility has become a key goal in manufacturing, hence the trend toward flexible manufacturing systems. These are designed to produce a variety of products from standard machinery with a minimum of workers. In the ultimate system, raw material in the form of bars, plates, and powder would be used to produce any assembly required without manual intervention in manufacture. Clearly, this is a good breeding ground for robots.

However, it should be emphasized that the early FMSs are in fact *direct numerical control* (DNC) systems for machining. And it must be acknowledged that an NC machine tool is really a special-purpose robot. It manipulates a tool in much the same way as a robot handles a tool or welding gun. Then, with no more than a change in programming, it can produce a wide range of products. Moreover, the controllers for robots and NC machines are almost the same. But for an NC machine to be converted into a self-supporting flexible system, it needs some extra equipment, including a handling device. It then forms an important element in an FMS.

The principle of flexible manufacturing for machining operations is that the NC machining cells are equipped with sensors to monitor tool wear and tool breakage. Such cells are able to operate unmanned so long as they can be loaded by a robot or similar device since the

sensors will detect any fault and shut the operation down if necessary. The requirements can be summarized as:

- A CNC system with sufficient memory to store many different machining programs.
- Automatic handling at the machining tool either by a robot or other material handling system.
- Workpieces stored near the machine to allow unmanned operation for several hours. A guided vehicle system may be employed if workpieces are placed at a designated storage and retrieval system away from the machine.
- Various sensors to monitor, locate, and/or diagnose any malfunction.

8.2 The Role of Sensors in FMS

The monitoring sensor devices are generally situated at the location of the machining process, measuring workpiece surface textures, cutting-tool vibrations, contact temperature between cutting tool and workpiece, flow rate of cooling fluid, electrical current fluctuations, and so on. Data in the normal operating parameters are stored in memory with data on acceptable manufacturing limits. As the tool wears, the tool changer is actuated. If the current rises significantly, along with other critical signals from sensors, a tool failure is indicated. Hence, the machine is stopped. Thus, with the combination of an NC machine, parts storage and retrieval, handling devices, and sensors, the unmanned cell becomes a reality. Since the control system of the NC machine, robot, and unmanned guided vehicles are similar, central computer control can be used effectively.

Systems based on these principles have been developed. In Japan, Fanauc makes numerical controllers, small NC and EDM machines, and robots; Murata makes robot trailers as well as a variety of machinery including automated textile equipment; and Yamazaki makes NC machines. In France, Renault and Citroen use FMS to machine gear boxes for commercial vehicles and prototype engine components, respectively, while smaller systems have been set up in many other countries.

The central element in establishing an error-free production environment is the availability of suitable sensors in manufacturing. The following represents a summary of sensing requirements in manufacturing applications:

- Part identification
- Part presence or absence
- Range of object for handling

- Single-axis displacement of measurement
- Two-dimensional location measurement
- Three-dimensional location measurement

8.2.1 Current Available Sensor Technology for FMS

The currently available sensors for manufacturing applications can be classified into four categories:

- *Vision sensors*
  - Photodetector
  - Linear array
  - TV camera
  - Laser triangulation
  - Laser optical time-domain reflectometry
  - Optical fiber
- *Tactile sensors*
  - Probe
  - Strain gauges
  - Piezoelectric
  - Carbon material
  - Discrete arrays
  - Integrated arrays
- *Acoustic sensors*
  - Ultrasonic detectors and emitters
  - Ultrasonic arrays
  - Microphones (voice control)
- *Passive sensors*
  - Infrared
  - Magnetic proximity
  - Ionizing radiation
  - Microwave radar

Integrating vision sensors and robotics manipulators in flexible manufacturing systems presents a serious challenge in production. Locating sensors on the manipulator itself, or near the end effector, provides a satisfactory solution to the position of sensors within the FMS. Locating an image sensor above the work area of a robot may cause the manipulator to obscure its own work area. Measurement of the displacement of the end effector also may suffer distortion since

the destination will be measured in relative terms, not absolute. Placing the sensor on the end effector allows absolute measurement to be taken, reducing considerably the need for calibration of mechanical position and for imaging linearity. Image sensory feedback in this situation can be reduced to the simplicity of range finding in some applications.

Extensive research and development activities were conducted recently to find ways to integrate various sensors close to the gripper jaws of robots. The promise of solid-state arrays for this particular application has not entirely materialized, primarily because of diversion of effort resulting from the commercial incentives associated with the television industry. It might be accurate to predict that, over the next decade, imaging devices manufactured primarily for the television market will be both small and affordable enough to be useful for robotics applications. However, at present, array cameras are expensive and, though smaller than most thermionic tube cameras, they are still far too large to be installed in the region of a gripper. Most of the early prototype arrays of modest resolution (developed during the mid-1970s) have been abandoned.

Some researchers have attacked the problem of size reduction by using coherent fiber optics to retrieve the image from the gripper array, which imposes a cost penalty on the total system. This approach can, however, exploit a fundamental property of optical fiber in that a bundle of coherent fibers can be subdivided to allow a single high-resolution imaging device to be used to retrieve and combine a number of lower-resolution images from various paths of the work area including the gripper with each subdivided bundle associated with its own optical arrangement.

Linear arrays have been used for parts moving on a conveyer in such a way that mechanical motion is used to generate one axis of a two-dimensional image. The same technique can be applied to a robot manipulator by suing the motion of the end effector to generate a two-dimensional image.

Tactile sensing is required in situations involving placement. Both active and passive compliant sensors have been successfully applied in the field. This is not the case for tactile array sensors because they are essentially discrete in design, are inevitably cumbersome, and have very low resolution.

Acoustic sensors, optical sensors, and laser sensors are well developed for effective use in manufacturing applications. Although laser range-finding sensors are well developed, they are significantly underused in FMS, especially in robotic applications. Laser probes placed at the end effector of an industrial robot will form a natural automated inspection system in manufacturing.

Sensing for robot applications does not depend on a relentless pursuit for devices with higher resolution; rather, the fundamental

consideration is selecting the optimum resolution for the task to be executed. There is a tendency to assume that the higher the resolution, the greater the application range for the system. However, considerable success has been achieved with a resolution as low as 50 × 50 picture elements. With serial processing architectures, this resolution will generate sufficient gray-scale data to test the ingenuity of image processing algorithms. Should its processing time fall below 0.5 s, an algorithm can be used for robots associated with handling. However, in welding applications, the image processing time must be faster.

8.3 Robot Control through Vision Sensors

An increasing number of manufacturing processes rely on machine-vision sensors for automation. The tasks for which vision sensors are used vary widely in scope and difficulty. Robotic applications in which vision sensing has been used successfully include inspection, alignment, object identification, and character recognition.

Human vision involves transformation, analysis, and interpretation of images. Machine-vision sensing can be explained in terms of the same functions: image transformation, image analysis, and image interpretation.

8.3.1 Image Transformation

Image transformation involves acquiring camera images and converting them to electrical signals that can be used by a vision computer (Fig. 8.1). After a camera image is transformed into an electronic (digitized) image, it can be analyzed to extract useful information in

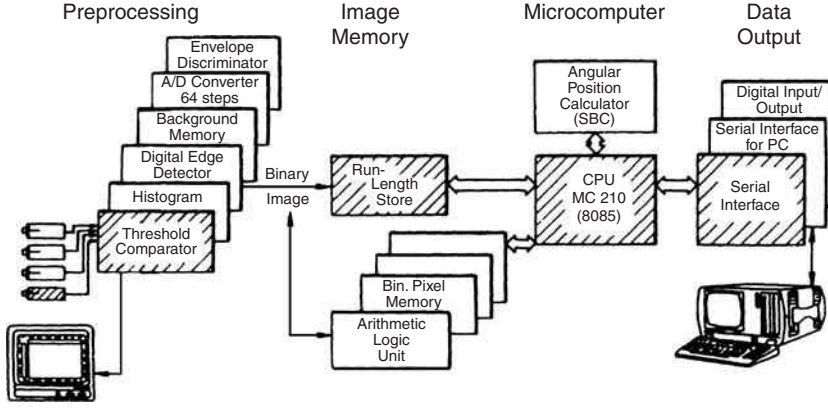


Figure 8.1 Image transformation involves the acquisition and conversion of camera images to electrical signals that can be used by a vision controlled CPU.

the image such as object edges, alignment, regions, boundaries, colors, and the absence or presence of vital components.

Once the image is analyzed, the vision sensing system can interpret what the image represents so that the robot can continue its task. In robot vision execution, design considerations entail cost, speed, accuracy, and reliability.

8.3.2 Robot Vision and Human Vision

Given that robot vision systems typically execute only part of what is normally called *seeing*, some similarities with human vision nevertheless arise. Other similarities arise in the "hardware" of human and robot visual systems.

The similarities in hardware could include an analogy between eyes and video cameras—both have lenses to focus an image on a sensitive "retina" that produces a visual signal interpreted elsewhere. In both human and robot vision this signal is passed to a device that can remember important aspects of the image for a time, perform specialized image processing functions to extract important information from the raw visual signal, and analyze the image in a more general way.

Some similarities in performance follow. Human and robot vision work well only where lighting is good. Both can be confused by shadows, glare, and cryptic color patterns. Both combine size and distance judgments, tending to underestimate size when they underestimate distance, for instance.

However, humans and machines have far more differences than similarities. A human retina contains several million receptors, constantly sending visual signals to the brain. Even the more limited video camera gathers over 7 Mb of visual information per second. Many of the surprising aspects of machine vision arise from the need to reduce this massive flow of data so it can be analyzed by a computer system.

Machine-vision systems normally only detect, identify, and locate objects, ignoring many of the other visual functions. However, they perform this restricted set of functions very well, locating and even measuring objects in a field of view more accurately than any human can.

8.3.3 Robot Vision and Visual Tasks

Several standard visual tasks are performed by robot-vision systems. These tasks include recognizing when certain objects are in the field of view, determining the location of visible objects, assisting a robot hand with pickup and placement, and inspecting known objects for the presence of certain characteristics (usually specific flaws in manufacture) (Fig. 8.2a). Figure 8.2b illustrates advanced vision inspection and measurement system capturing a protein like string consisted of

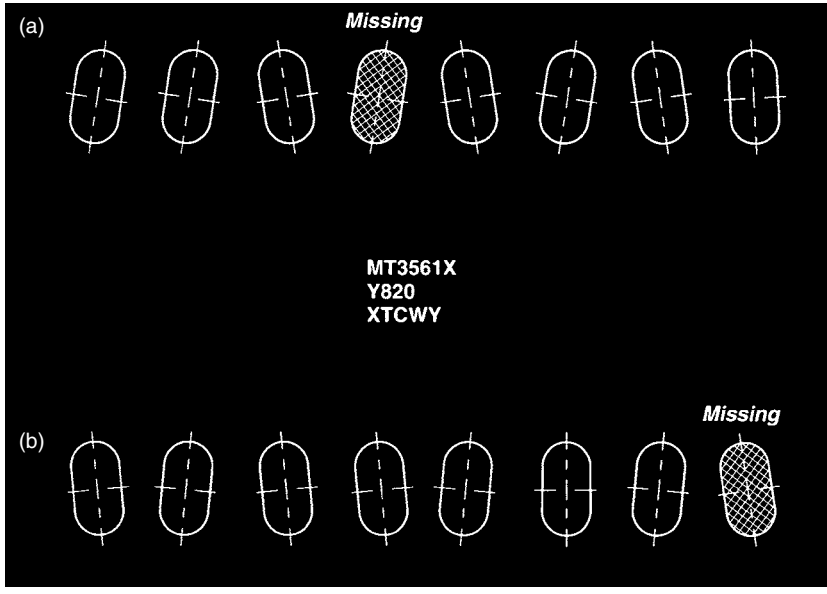


FIGURE 8.2 (a) Image flaw detection and analysis. (b) Captured image of protein like strings measuring maximum length of fibers in microns. (*Curtsey American SensoRx, inc.*)

a matrix of arranged fibers. This string protein structure is used in growing human tissue of a wounded patient.

A robot-vision system must exercise some judgment in performing *visual tasks*—those for which the input is a visual image (normally obtained from an ordinary video camera). Which visual tasks are relatively easy for machine vision, and which are hard? The distinction is not so much that some tasks are hard and others easy; rather, it is the detail within a task that distinguishes easy problems from hard ones.

What makes a problem hard? Some of the contributing factors are:

- Objects that vary widely in detail. (Examining stamped or milled product may be easy, while molded or sculpted items may be more difficult. Natural objects are by far the hardest with which to deal.)
- Lighting variations, including reflections and shadows, as well as fluctuations in brightness (as found in natural sunlight). These variations may go unnoticed by human inspectors, but they can make otherwise easy problems difficult or impossible for robot vision.
- In general, ignoring "unimportant" variations in an image while responding to "significant" ones is very hard. (Most hard problems can be placed in this category.)

8.3.4 Robot Visual Sensing Tasks

Robots will work in unpleasant locations. Health hazards are of no concern to them. Special safety equipment is not required for a robot spraying paint, welding, or handling radioactive materials or chemicals. All this adds up to reduced production costs. As the day progresses, the tired worker has a tendency to pay less attention to details, and the quality of the finished product may suffer. This is especially noticeable in automobiles where spray paint can run or sag and weld joints may not be made perfectly. The panels of the car may not be aligned, and the finished product may not operate properly, with predictable customer dissatisfaction. In pharmaceutical production, too, an operator inspecting and verifying lot number and expiration date may become fatigued and fail to inspect for sensitive information.

Robots, on the other hand, do not tire or change their work habits unless they are programmed to do so. They maintain the same level of operation throughout the day. With vision-sensing systems and robots, it is possible for American manufacturers to compete against lower labor costs in foreign countries. The initial investment is the only problem. After such an investment, the overall operation costs of the production line are reduced or held constant. Small educational robots can be used to retrain humans to operate and maintain robots.

The roles of robots with machine vision can be summarized as follows:

- *Handling and assembly.* Recognizing position/orientation of objects to be handled or assembled, determining presence or absence of parts, and detecting parts not meeting required specifications
- *Part classification.* Identifying objects and recognizing characters
- *Inspection.* Checking for assembly and processing, surface defects, and dimensions
- *Fabrications.* Making investment castings, grinding, deburring, water-jet cutting, assembling wire harnesses, gluing, sealing, puttying, drilling, fitting, and routing
- *Welding.* automobiles, furniture, and steel structures
- *Spray painting.* automobiles, furniture, and other objects

8.3.5 Robots Utilizing Vision Systems to Recognize Objects

A basic use of vision is recognizing familiar objects. It is easy to see that this ability should be an important one for robot vision. It can be a task in itself, as in counting the number of each kind of bolt in a mixed lot on a conveyer belt. It can also be an adjunct to other tasks—for example, recognizing a particular object before trying to locate it precisely, or before inspecting it for defects.

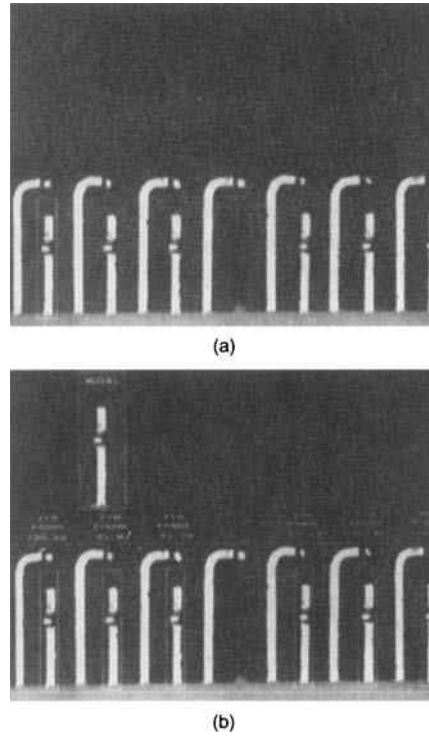


Figure 8.3
Recognition of a learned object.

It is important to note that this task actually has two distinct parts: first, object familiarization—that is, learning what an object looks like, then object recognition (Fig. 8.3).

There are many ways of learning to recognize objects. Humans can learn from verbal descriptions of the objects, or they can be shown one or more typical items. A brief description of a *pencil* is enough to help someone identify many unfamiliar items as *pencil*. Shown a few *pencils*, humans can recognize different types of *pencils*, whether they look exactly like the samples or not.

Robot-vision systems are not so powerful, but both these approaches to training still apply to them. Robots can be given descriptions of what they are to recognize, perhaps derived from CAD data to guide a machine tool (Fig. 8.4) or they can be shown samples, then be expected to recognize objects more or less like the samples.

Recognizing objects once they have been learned is the second, and more difficult, part of the task. Several basic questions arise in virtually every recognition task. Among them are "What are the choices?" and "What can change?"

Specifying the actual task completely is normally the hardest part of the application. When this has been accomplished, the basic question is, for what particular vision features should the search begin to

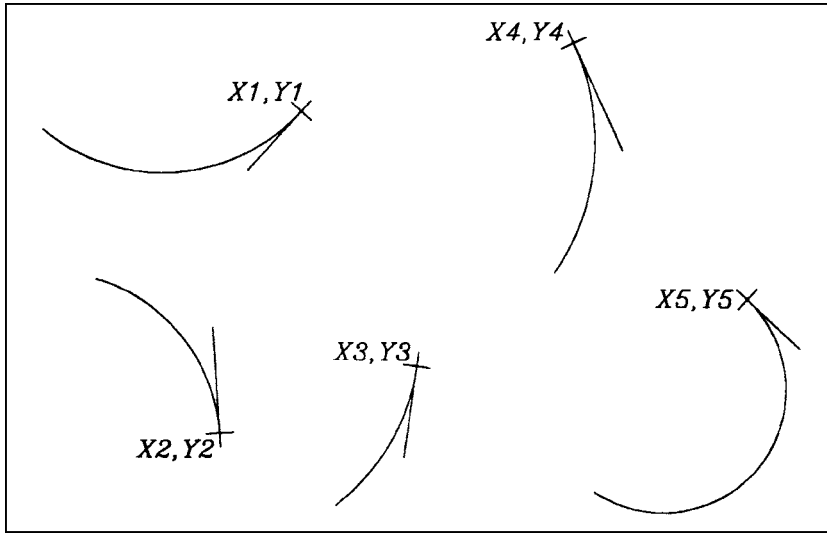


Figure 8.4 Machine guidance.

recognize an object, and, when the features are found, how should they be analyzed to decide which object (if any) has been found?

8.4 Robot Vision Locating Position

Humans use several techniques for gauging distances, especially triangulation on the left- and right-eye views, feedback from the eye's focusing mechanism, and the apparent motion produced by small head movements. This kind of object location may make use of knowledge about the object being located. By knowing how large the object really is, one can judge how far away it is from the size of its retinal image.

Few robot-vision systems use binocular vision, autofocus feedback, or moving cameras to estimate distances. However, with a rigidly mounted camera, it is possible to interpret each visible point as lying on a particular line of sight from the camera. Accurate information about the true distances between visible points on an object allows the robot-vision system to accurately calculate its distance. Similarly, if an object is resting on a platform at an accurately known distance, the robot vision system can interpret distances in the camera image as accurate distances on the object.

However locations are determined, the visual location task for a robot-vision system normally includes calibration as well as object location. Calibration normally is carried out by providing the system with a view that has easily identifiable points at known spatial locations.

Once calibrated, the system can then locate objects in its own coordinate system (pixels) and translate the position into work cell coordinates (inches, millimeters, etc.).

8.5 Robot Guidance with Vision System

Another use of machine vision is in robot guidance—helping a robot to handle and place parts and providing it with the visual configuration of an assembly after successive tasks. This can involve a series of identification and location tasks. The camera can be attached to a mobile arm, making the location task seem somewhat more like normal vision. However, the camera typically is mounted on a fixed location to reduce system complexity.

While each image can give the location of certain features with respect to the camera, this information must be combined with information about the current location and orientation of the camera to give an absolute location of the object. However, the ability to move the camera for a second look at an object allows unambiguous location of visible features by triangulation. Recognition is a useful tool for flexible manufacturing systems within a CIM environment. Any of several parts may be presented to a station where a vision system determines the type of part and its exact location. While it may be economical and simpler to send the robot a signal giving the part type when it arrives, the ability to detect what is actually present at the assembly point, not just what is supposed to be there, is of real value for totally flexible manufacturing.

8.5.1 Robot Vision Performing Inspection Tasks

Visual inspection can mean any of a wide variety of tasks, many of which can be successfully automated. Successful inspection tasks are those in which a small number of reliable visual cues (features) are to be checked and a relatively simple procedure is used to make the required evaluation from those cues.

The differences between human and robot capabilities are most evident in this kind of task, where the requirement is not simply to distinguish between good parts and anything else—a hard enough task—but usually to distinguish between good parts or parts with harmless blemishes, and bad parts. Nevertheless, when inspection can be done by robot vision, it can be done very predictably.

Many inspection tasks are well suited to robot vision. A robot-vision system can dependably determine the presence or absence of particular items in an assembly (Fig. 8.5), providing accurate information on each of them. It can quickly gauge the approximate area of each item passing before it, as long as the item appears somewhere in its field of view (Fig. 8.6).



Figure 8.5 Presence or absence of items in an assembly.

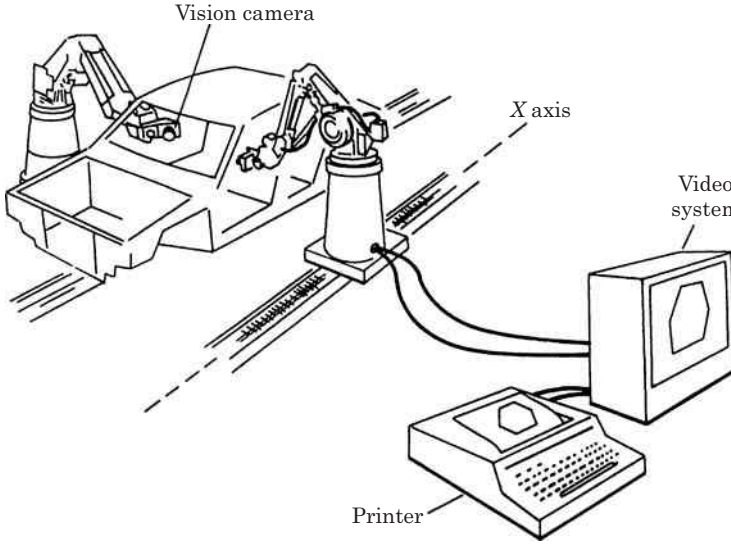


Figure 8.6 Object in field of view.

8.5.2 Components of Robot Vision

Figure 8.7 is a schematic diagram of the main components of a robot in a typical vision process for manufacturing. A fixed camera surveys a small, carefully lighted area where the objects to be located or inspected are placed. When visual information is needed (as signaled by some external switch or sensors), a digitizer in a robot vision system converts the camera image into a "snapshot": a significant array of integer brightness values (called gray levels). This array is sorted in a large *random access memory* (RAM) array in the robot-vision system called an *image buffer* or a *frame buffer*. Once sorted, the image can be displayed on a monitor at any time. More importantly, the image can be analyzed or manipulated by a vision computer, which can be programmed to solve robot vision problems. The vision computer is often connected to a separate general-purpose (host) computer,

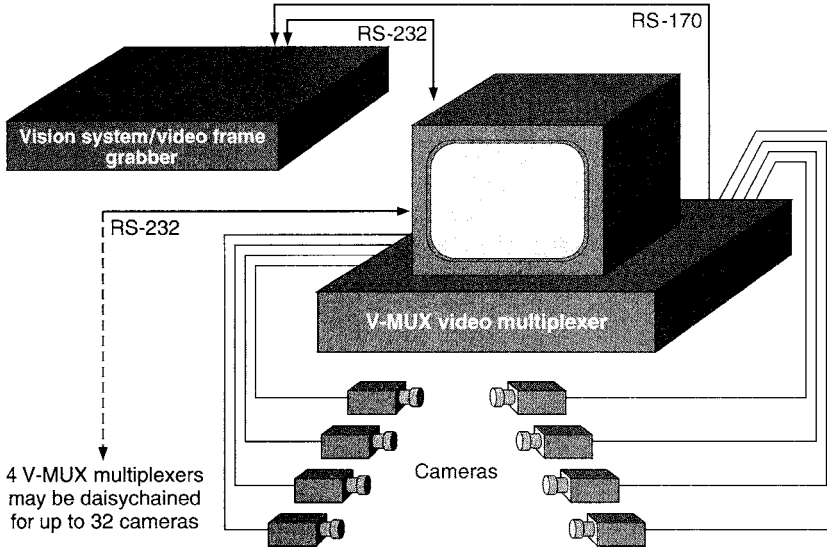


Figure 8.7 Schematic diagram of typical vision process.

which can be used to load programs or to perform tasks not directly related to vision.

Once an image is acquired, vision processing operations follow a systematic path. Portions of the image buffer may first be manipulated to suppress information that will not be valuable to the task at hand and to enhance the information that will be useful. Next, the vision program extracts a small number of cues from the image—perhaps allowing the region of interest to be reduced to exclude even more extraneous data.

At this stage, the vision program calculates, from the selected image region, the cues (features) of direct importance to the task at hand and makes a decision about the presence of a known part or its location in the field of view, or perhaps about the presence of specific defects in the object being inspected. Finally, the robot-vision system activates control lines based on the decisions, and (perhaps) transfers a summary of the conclusion to a data storage device or another computer.

8.6 End Effector Camera Sensor for Edge Detection and Extraction

A considerable amount of development of synchronized dual camera sensors at a strategic location on a robot end effector has been conducted for processing two-dimensional images stored as binary matrices. A large part of this work has been directed toward solving problems of character recognition. While many of these techniques

are potentially useful in the present context, it is valuable to note some important differences between the requirements of character recognition and those associated with visual feedback for mechanical assembly.

8.6.1 Shape and Size

All objects presented to the assembly machine are assumed to match an exact template of the reference object. The object may have an arbitrary geometric shape, and the number of possible different objects is essentially unlimited. Any deviation in shape or size, allowing for errors introduced by the visual input system, is a ground for rejection of the object (though this does not imply the intention to perform 100 percent inspection of components). The derived description must therefore contain all the shape and size information originally presented as a stored image. A character recognition system must tolerate considerable distortion, or style, in the characters to be recognized, the most extreme example being handwritten characters. The basic set of characters, however, is limited. The closest approach to a template-matching situation is achieved with the use of a type font specially designed for machine reading, such as optical character recognition (Fig. 8.8).

8.6.2 Position and Orientation

A component may be presented to the assembly machine in any orientation and any position in the field of view. Though a position- and orientation-invariant description is required in order to recognize the component, the measurement of these parameters is also an important function of the visual system to enable subsequent manipulation. While a line character may sometimes be skewed or bowed, individual

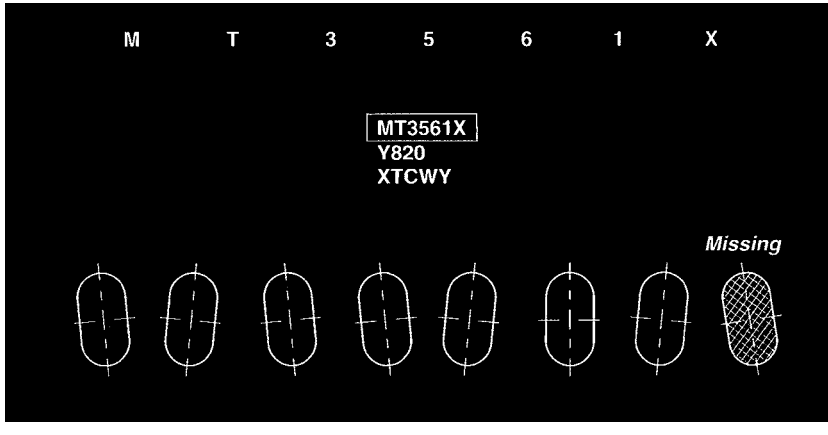


Figure 8.8 Optical character recognition.

characters are normally presented to the recognition system in a relatively constrained orientation, a measurement of which is not required.

8.6.3 Multiple Objects

It is a natural requirement that the visual system for an assembly machine should be able to accommodate a number of components randomly positioned in the field of view. The corresponding problem of segmentation in character recognition is eased (for printed characters) by *a priori* knowledge of character size and pitch. Such information has fostered techniques for the segmentation of touching characters. No attempt is made to distinguish between touching objects. Their combined image will be treated by the identification procedures as that of a single, supposedly unknown object.

The essentially unlimited size of the set of objects that must be accommodated by the recognition system demands that a detailed description of shapes be extracted for each image. There are, however, a number of basic parameters which may be derived from an arbitrary shape to provide valuable classification and position information. These include:

- Area
- Perimeter
- Minimum enclosing rectangle
- Center of the area
- Minimum radius vector (length and direction)
- Maximum radius vector (length and direction)
- Holes (number, size, position)

Measurements of area and perimeter provide simple classification criteria that are both position- and orientation-invariant. The dimensionless shape factor area/perimeter has been used as a parameter in object recognition. The coordinates of the minimum enclosing rectangle provide some information about the size and shape of the object, but this information is orientation-dependent. The center of area is a point that may be readily determined for any object, independent of orientation, and is thus of considerable importance for recognition and location purposes. It provides the origin for the radius vector, defined as a line in the center of the area to a point on the edge of an object. The radius vectors of maximum and minimum length are potentially useful parameters for determining both identification and orientation. Holes are common features of engineering components, and the number present in a part is a further suitable parameter. The holes themselves may also be treated as objects, having shape, size, and position relative to the object in which they are found.

The requirements for the establishment of connectivity in the image and the derivation of detailed descriptions of arbitrary geometric shapes are most appropriately met by an edge-following technique. The technique starts with the location of an arbitrary point on the black/white edge of an object in the image (usually by a raster scan). An algorithm is then applied that locates successive connected points on the edge until the complete circumference has been traced and the starting point is reached. If the direction of each edge point relative to the previous point is recorded, a one-dimensional description of the object is built up, which contains all the information present in the original shape. Such chains of directions have been extensively studied by Freeman. Measurements of area, perimeter, center of area, and enclosing rectangle may be produced while the edge is being traced, and the resulting edge description is in a form convenient for the calculation of radius vectors.

Edge following establishes connectivity for the object being traced. Continuing the raster scan in search of further objects in the stored image then presents the problem of the rediscovery of an already traced edge.

A computer plot of the contents of the frame with the camera viewing a square and a disk is illustrated in Fig. 8.9, and the result of applying the edge-extracting operation is illustrated in Fig. 8.10. The edge-following procedure may now be applied to the image in the same way it was to the solid object. The procedure is arranged,

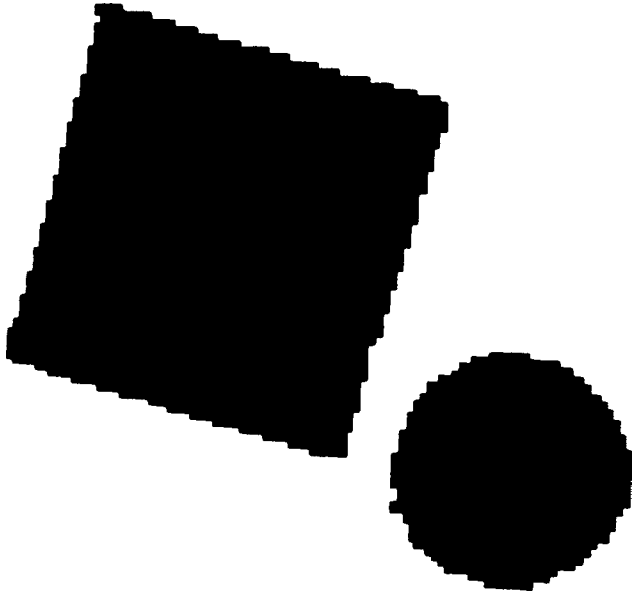


Figure 8.9 Computer plot of the contents of a frame.

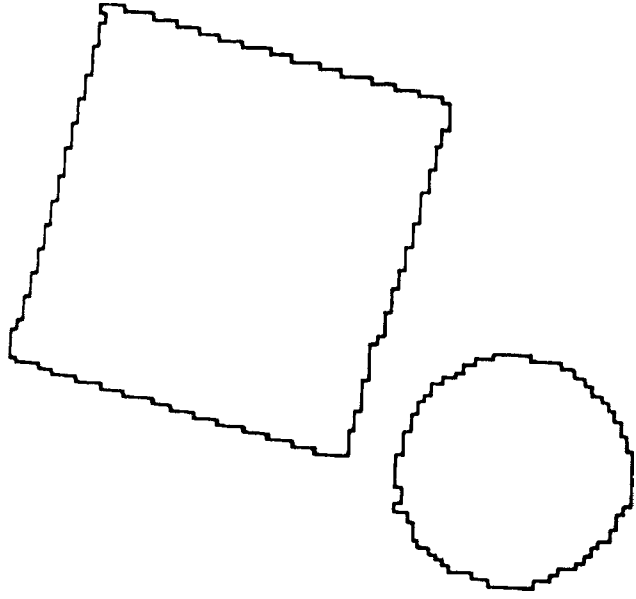


Figure 8.10 Result of applying edge extraction operation.

however, to reset each edge point as it is traced. The tracing of a complete object thus removes it from the frame and ensures it will not be subsequently retraced.

8.7 End Effector Camera Sensor Detecting Partially Visible Objects

A new method of locating partially visible two-dimensional objects has been developed. The method is applicable to complex industrial parts that may contain several occurrences of local features, such as holes and corners. The matching process utilizes clusters of mutually consistent features to hypothesize objects and uses templates of the objects to verify these hypotheses. The technique is fast because it concentrates on key features that are automatically selected on the basis of detailed analysis of CAD-type models of the objects. The automatic analysis applies general-purpose routines for building and analyzing representations of clusters of local features that could be used in procedures to select features for other locational strategies. These routines include algorithms to compute the rotational and mirror symmetries of objects in terms of their local features. The class of tasks that involve the location of the partially visible object ranges from relatively easy tasks, such as locating a single two-dimensional object, to the extremely difficult task of locating three-dimensional

objects jumbled together in a pallet. In two-dimensional tasks, the uncertainty is in the location of an object in a plane parallel to the image plane of the camera sensor. This restriction implies a simple one-to-one correspondence between sizes and orientations in the image, on the one hand, and sizes and orientations in the plane of the object, on the other.

This class of two-dimensional tasks can be partitioned into four subclasses that are defined in terms of the complexity of the scene:

- A portion of one of the objects
- Two or more objects that may touch one another
- Two or more objects that may overlap one another
- One or more objects that may be defective

This list is ordered roughly by the increasing amount of effort required to recognize and locate the object.

Figure 8.11 illustrates a portion of an aircraft frame member. A typical task might be to locate the pattern of holes for mounting purposes. Since only one frame member is visible at a time, each feature appears at most once, which simplifies feature identification. If several objects can be in view simultaneously and can touch one another, as in Fig. 8.12, the features may appear several times. Boundary features such as corners may not be recognizable, even though they are in the picture, because the objects are in mutual contact. If the objects can lie on one another (Fig. 8.13), even some of the internal holes may

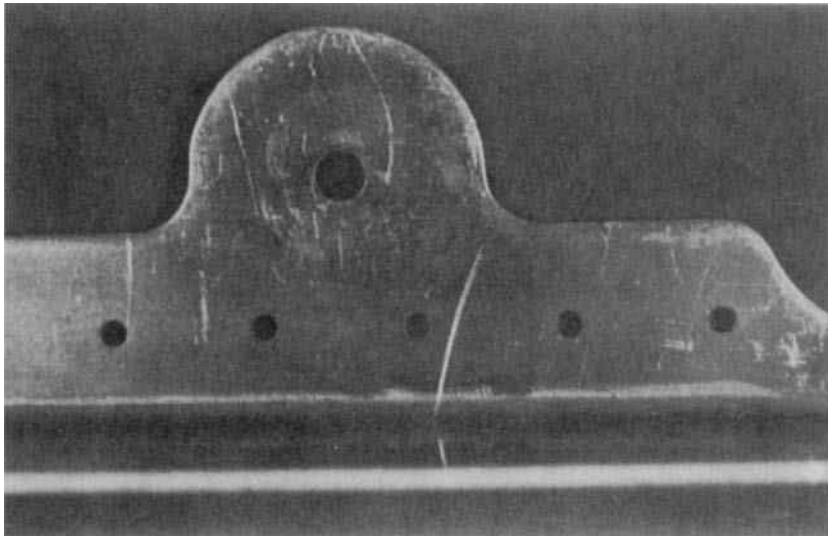


Figure 8.11 Portion of an aircraft frame member.

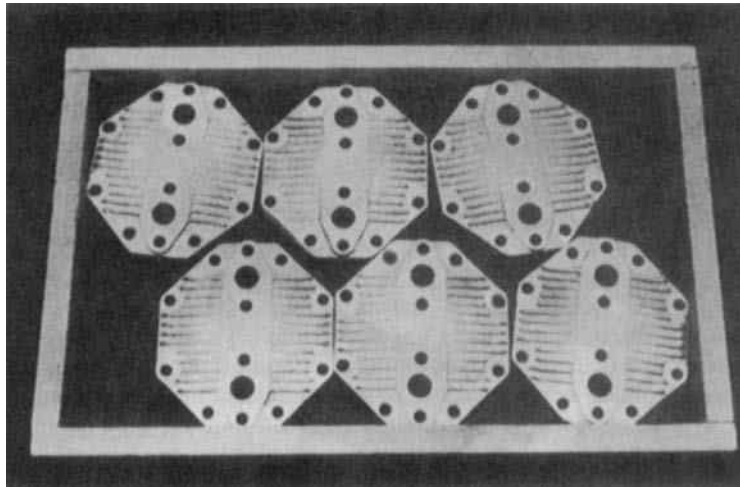


Figure 8.12 Objects touching each other.

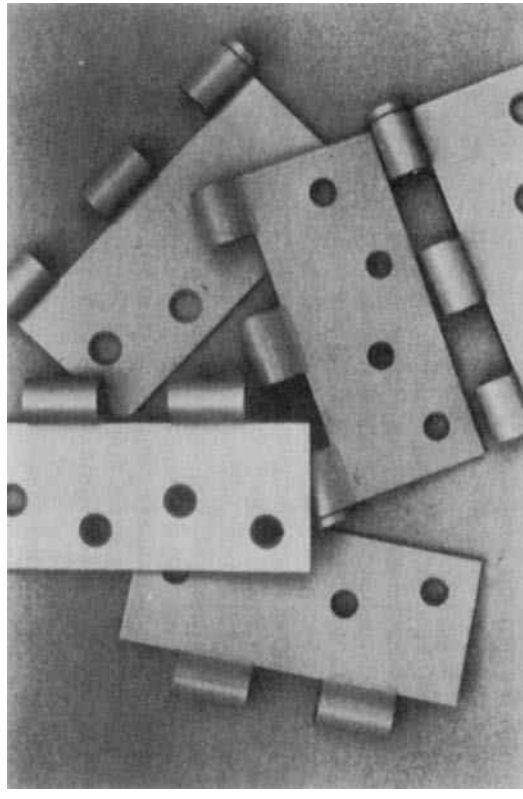


Figure 8.13 Objects lying on top of each other.

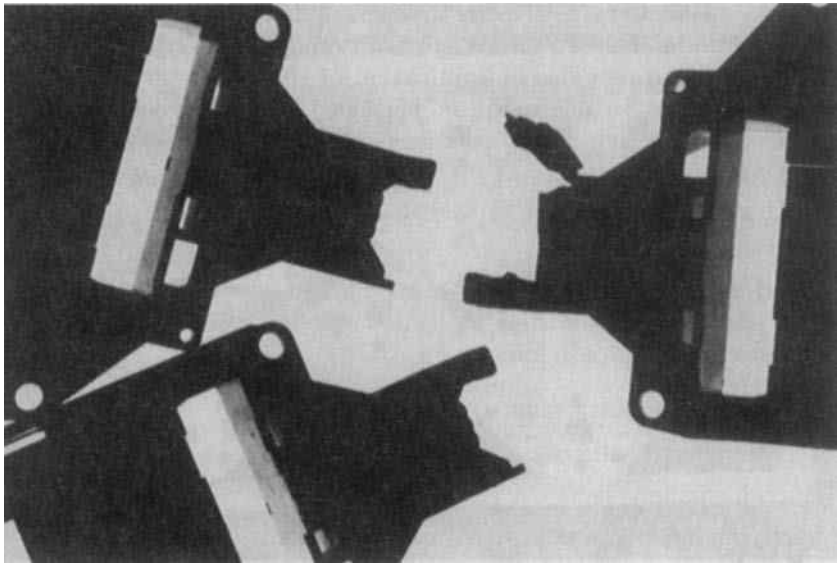


Figure 8.14 Trained image.

be unrecognizable because they are partially or completely occluded. And, finally, if the objects are defective (Fig. 8.14), the features are even less predictable and hence harder to find.

Since global features are not computable from a partial view of an object, recognition systems for these more complex tasks are forced to work with either local features, such as small holes and corners, or extended features like a large segment of an object's boundary. Both types of feature, when found, provide constraints on the position and the orientations of their objects. Extended features are in general computationally more expensive to find, but they provide more information because they tend to be less ambiguous and more precisely located.

Given a description of an object in terms of its features, the time required to match this description with a set of observed features appears to increase exponentially with the number of features. The multiplicity of features precludes the straightforward application of any simple matching technique. Large numbers of features have been identified by locating a few extended features instead of many local ones. Even though it costs more to locate extended features, the reduction in the combinatorial explosion is often worth it. The other approach is to start by locating just one feature and use it to restrict the search area for nearby features. Concentrating on one feature may be risky, but the reduction in the total number of features to be considered is often worth it. Another approach is to sidestep the

problem by hypothesizing massively parallel computers that can perform matching in linear time. Examples of these approaches include graph matching, relaxation, and histogram analysis. The advantage of these applications is that the decision is based on all the available information at hand.

The basic principle of the *local-feature-focus* (LFF) method is to find one feature of an image, referred to as the *focus feature*, and use it to predict a few nearby features to look for. After finding some nearby features, the program uses a graph-matching technique to identify the largest cluster of image features matching a cluster of object features. Since the list of possible object features has been reduced to those near the focus feature, the graph is relatively small and can be analyzed efficiently.

The key to the LFF method is an automatic feature-selection procedure that chooses the best focus features and the most useful sets of nearby features. This automatic-programming capability makes possible quick and inexpensive application of the LFF method to new objects. As illustrated in Fig. 8.15, the training process, which includes the selection of features, is performed once and the results are used repeatedly.

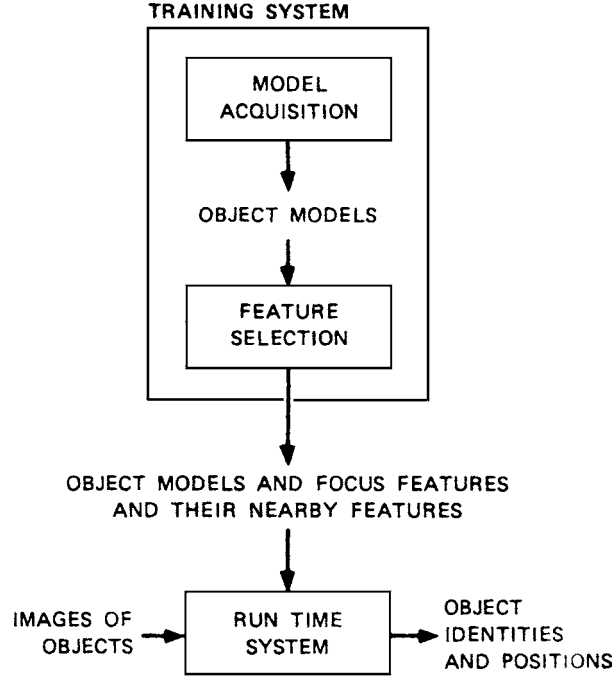


Figure 8.15 Runtime phase procedure.

8.7.1 Runtime Phase

The runtime phase of the LFF acquires images of partially visible objects and determines their identities, positions, and orientations. This processing occurs in four steps:

1. Reading task information
2. Locating local features
3. Hypothesizing objects
4. Verifying hypotheses

The procedure (Fig. 8.15) is to input the object model together with the list of focus features and their nearby cofeatures. Then, for each image, the system locates all potentially useful local features, forms clusters of them to hypothesize object occurrences, and finally performs template matching to verify these hypotheses.

8.8 Ultrasonic End Effector

An end effector on a welding robot (Fig. 8.16) contains an ultrasonic sensor for inspection of the weld. An ultrasonic sensor detects such flaws as tungsten inclusions and lack of penetration of weld. The end effector determines the quality of a weld immediately after the weld contact has been made, while the workpiece is still mounted on the weld apparatus; a weld can be reworked in place, if necessary. The delay caused by the paperwork and setup involved in returning the workpiece for rework is thereby avoided.

The ultrasonic end effector can be mounted on any standard gas tungsten arc welding torch. It may also be equipped with a through-the-torch vision system. The size of the ultrasonic end effector is the same as that of a gas cup with a cathode.

A set of extension springs stabilizes the sensor and ensures that its elastomeric dry-couplant pad fits squarely in the weldment surface. The sensor can be rotated 360° and locked into alignment with the weld lead. A small force-actuated switch halts downward travel of the robot arm toward the workpiece and sets the force of contact between the sensor and the workpiece.

8.9 End Effector Sound-Vision Recognition Sensors

The sound recognition sensor consists of a source that emits sound waves to an object and a sound receiver that receives the reflected sound waves from the same object (Fig. 8.17). The sound recognition sensor array consists of one sound source and one to as many as 16 receivers fitted intricately on an end effector of a robot.

The sound-vision recognition sensor array measures reflections from some surface of interest on the object, called the *measured surface*,

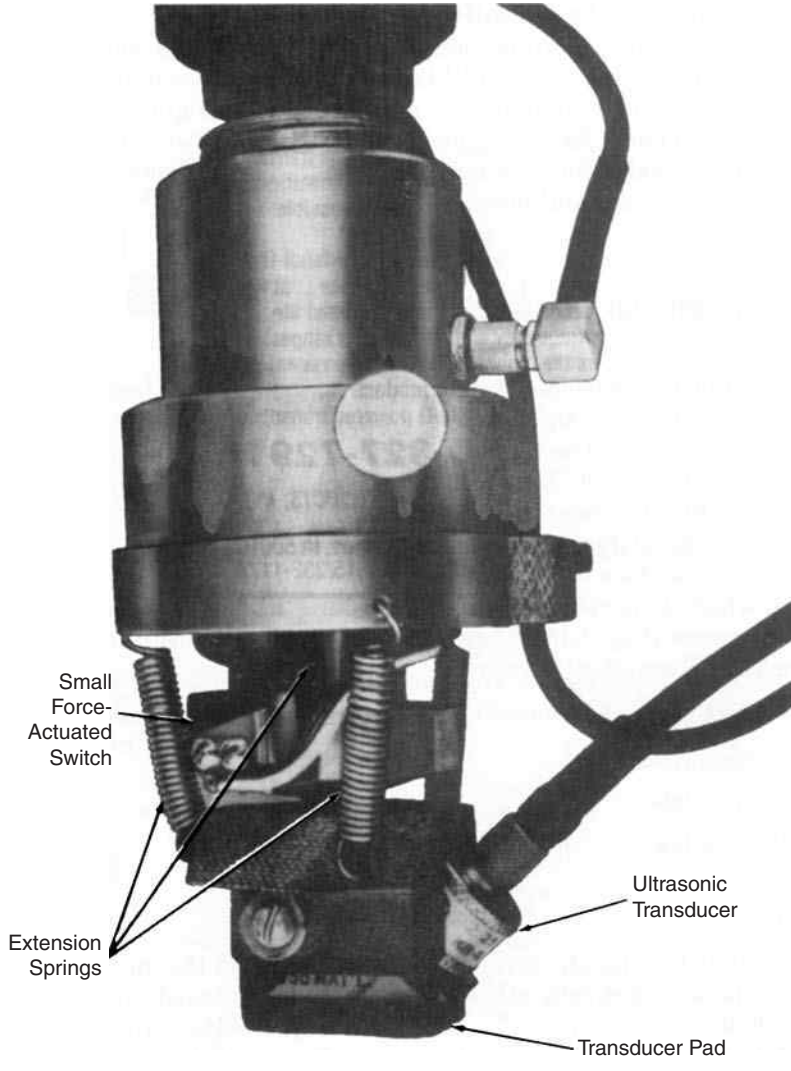


Figure 8.16 End effector on a welding robot.

which is perpendicular to the sound waves emitted from the sound source (Fig. 8.18). Four conditions govern the performance of sound-vision sensors:

- Standoff
- Large surfaces
- Small surfaces
- Positioning

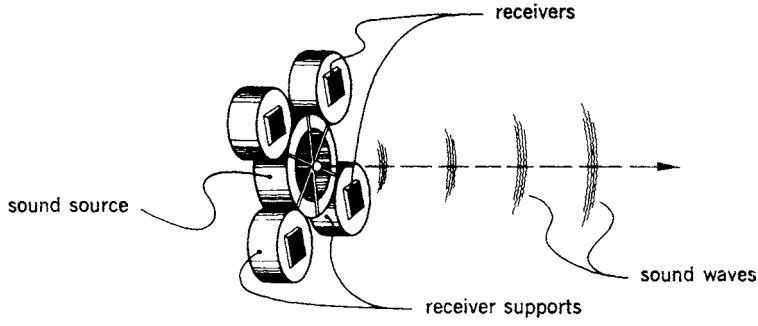


Figure 8.17 End effector sound-vision recognition system.

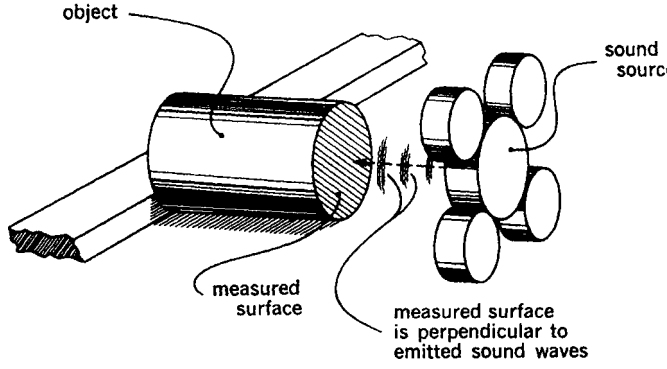


Figure 8.18 The measured surface should be perpendicular to the sound waves emitted from the sound source.

8.9.1 Standoff

Standoff is how far the array must be located from the measured surface. The standoff, like other measurements, is based on the wavelength of the sound used. Three different wavelengths are used in the sound-vision sensor recognition system. The array standoff d should be one or two wavelengths λ_s from the measured surface for the highest accuracy. The standoff can be as great as 12 wavelengths, albeit with reduced accuracy (Fig. 8.19).

$$1.5\lambda_s \le d \le 12\lambda_s \tag{8.1}$$

where λ_s is the sound wavelength and d is the standoff distance. The typical standoff distance in terms of frequency and wavelength is described in Table 8.1.

8.9.2 Large Surface Measurements

Large surface measurements achieve more accuracy than those made on small surfaces (Fig. 8.20). Whether a surface is large or small, the

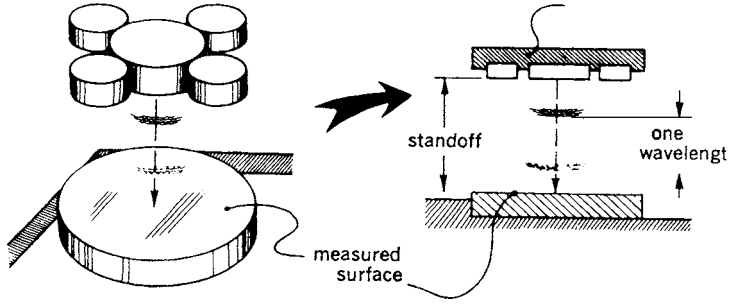


Figure 8.19 Standoff.

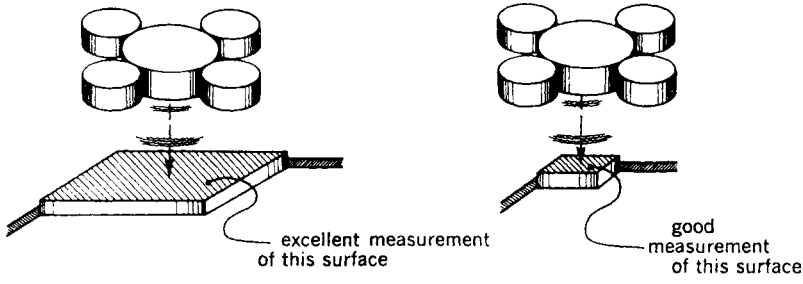


Figure 8.20 Large surface measurements.

Frequency, kHz	Wavelength, mm	Standoff Distance, mm
20	17	25
40	8	12
80	4	6

Table 8.1 Correlation Functions of Typical Standoff d, Wavelength λ_s, and Frequency f

Frequency f, kHz	20	40	80
Wavelength λ_s, mm	17	8	4
Minimum area of surface, mm^2	275	70	20

Table 8.2 Correlation of Minimum Size of Large Surface, Frequency, and Wavelength

accuracy depends on the wavelength of the sound waves selected. A "large" surface must be at least one wavelength distance on each side (Table 8.2).

The large surface being measured should change its dimension perpendicular to the surface, in the same direction as the sound wave emitted from the sound source (Fig. 8.21).

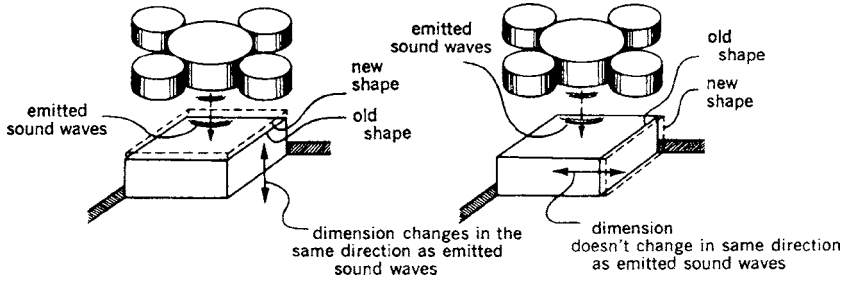


Figure 8.21 The large surface being measured should change its dimension perpendicular to the surface.

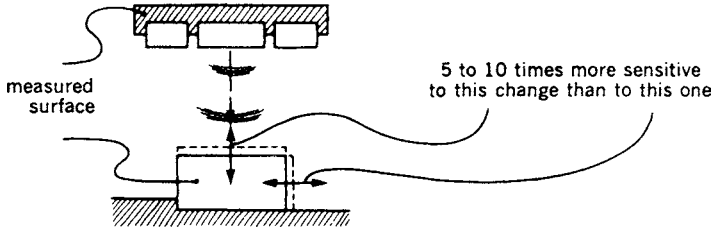


Figure 8.22 Sensitivity to dimension changes.

8.9.3 Sensitivity of Measurements

The sensitivity of the measurements to dimension changes is five or ten times greater when the change is in the same direction as the emitted sound wave (Fig. 8.22).

8.9.4 Small Surfaces

Small surfaces can be measured as long as the robot end effector carrying the sensor array directs the sound wave from the side of the object (Fig. 8.23). Small surfaces are either a small portion of a large object or simply the surface of a small object (Fig. 8.24).

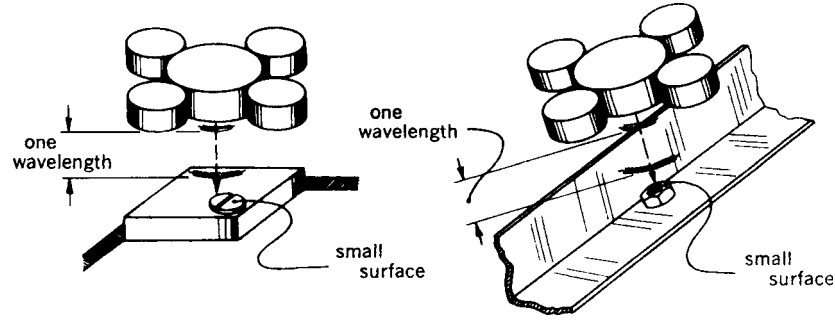


Figure 8.23 Small surfaces.

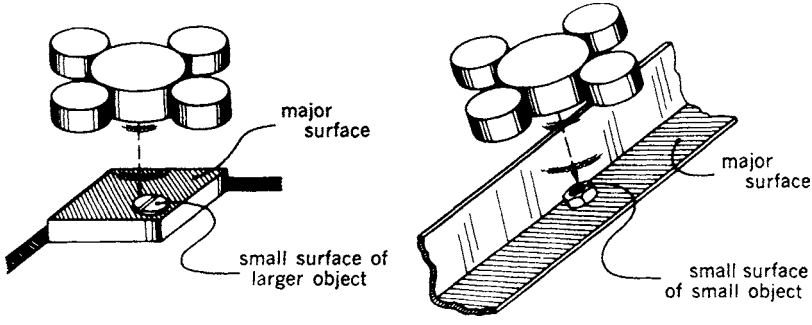


Figure 8.24 Small surfaces are either a small portion of a large object or simply the surface of a small object.

For small surfaces, the sound waves diffract or wrap around the surface rather than diverge from it as in a large surface. Similarly, ocean waves diffract around a rock that is smaller than the distance between crests (Fig. 8.25). Because of diffraction, the sound-vision recognition system is sensitive to the volume of the shape change on a small surface (Fig. 8.26).

Volume changes can be positive or negative. Small objects or protrusion from a surface represent positive volume changes, while holes or cavities represent negative volume changes (Fig. 8.27).

Measurement accuracy for small surfaces depends on the change in volume of the surface being measured. Listed in Table 8.3 are the smallest volume change that can be detected and the approximate size of a cube that has that volume for representative acoustic frequencies.

8.9.5 Positioning

The sound-vision sensor array system compares a particular par with a reference or standard part and detects the difference between the

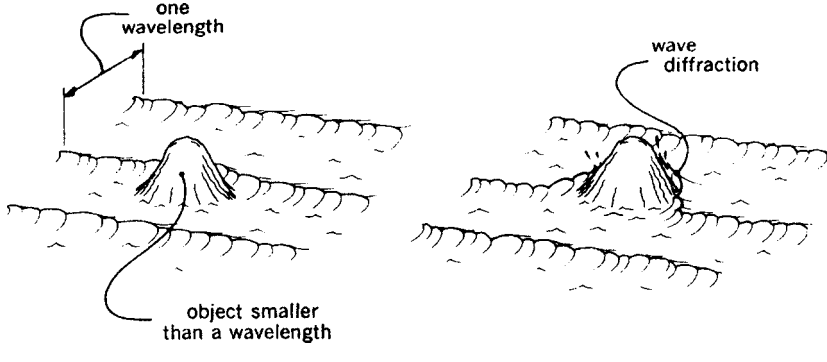


Figure 8.25 Ocean waves diffract around a rock that is smaller than the distance between crests.

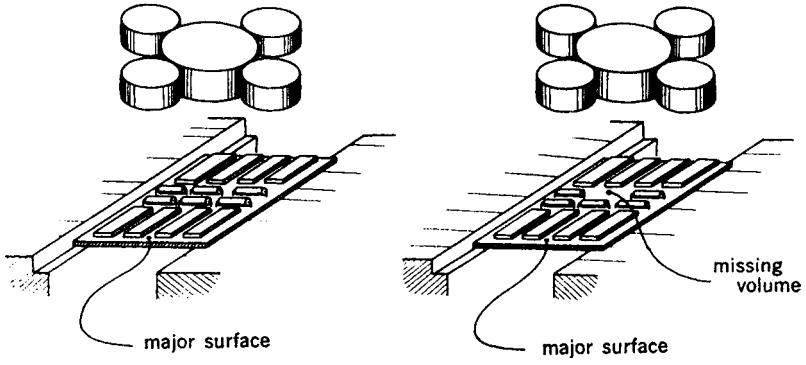


Figure 8.26 Sound-vision recognition system.

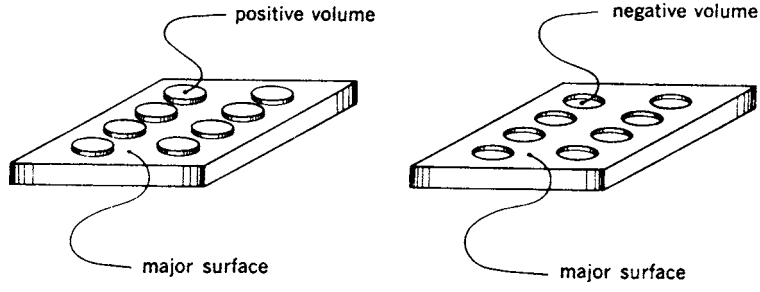


Figure 8.27 Volume changes can be positive or negative.

Frequency *f*, kHz	20 kHz	40 kHz	80 kHz
Smallest detectable volume change, μm^3	5×10^{-4}	6×10^{-5}	8×10^{-6}
Smallest detectable cube, mm^3	12×10^{2}	6×10^{2}	3×10^{2}

Table 8.3 The Least Measurable Volume

two. Objects being inspected—either a part or its standard—must be located relative to the array with at least the same accuracy as the expected measurement. Rectangular parts are usually located against stops; rotational parts are usually located in V-blocks on the face of the end effector (Fig. 8.28).

Rotationally asymmetric parts must be oriented the same way as the standard in order to be compared. The end effector sensor array system can be used to direct a stepper motor to rotate the part until a match between part and standard is found (Fig. 8.29).

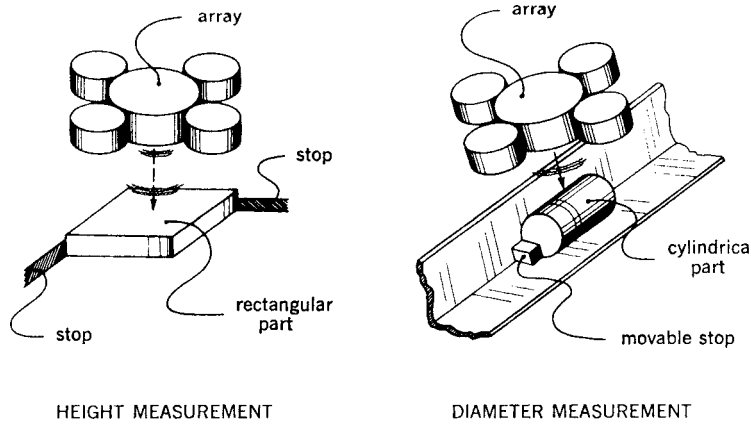


Figure 8.28 Diameter and height measurement.

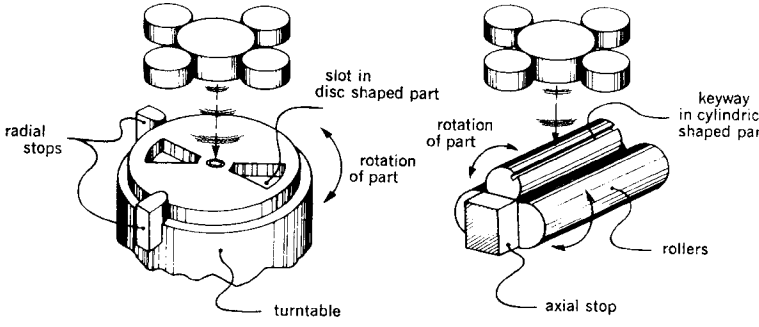


Figure 8.29 The end effector sensor can be used to direct a stepper motor to rotate the part until a match between part and standard is found.

8.10 End Effector Linear Variable-Displacement Transformer Sensor

Sensing capability in a robot can have widely ranging degrees of sophistication in addition to a variety of sensing media. For instance, sensing capability can vary from a simple photoelectric cell to a complex, three-dimensional sound-vision system as described in the previous section.

The *linear variable-displacement transformer* (LVDT) sensor is an electromechanical device that can be attached to a robotic manipulator or can be itself a drive control for a robotic gripper. The LVDT produces an electrical output proportional to the displacement of a separate movable core. It consists of a primary coil and two secondary coils, intricately spaced on a cylindrical form. A free-moving rod-shaped magnetic core inside the coil assembly provides a path for the magnetic flux linking the coils. A cross-section of the LVDT sensor and a plot of its operational characteristics are shown in Figs. 8.30 and 8.31.

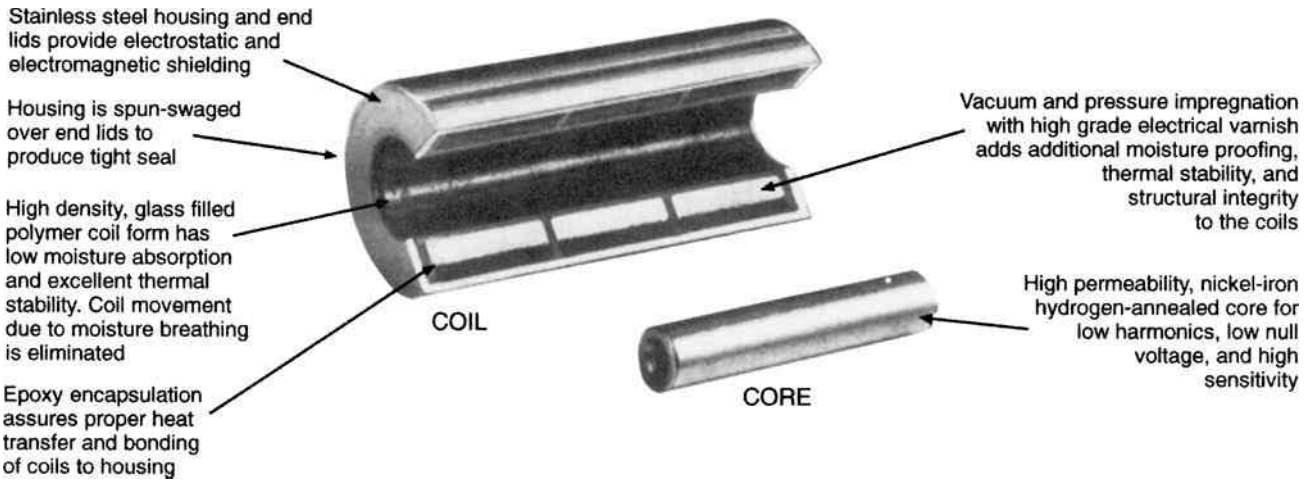


Figure 8.30 Cross-section of a linear variable displacement transducer sensor.

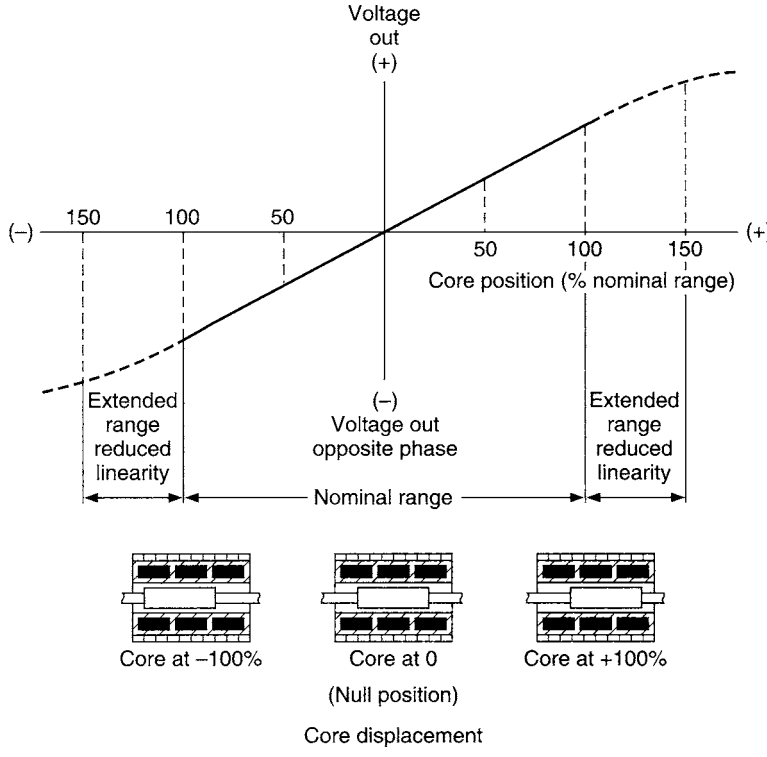


Figure 8.31 Plot of LVDT operational characteristics.

When the primary coil is energized by an external AC source, voltages are induced in the two secondary coils. These are connected series-opposing so that the two voltages are of opposite polarity. Therefore, the net output of the sensor is the difference between these voltages, which is zero when the core is at the center or null position. When the core is moved from the null position, the induced voltage in the coil toward which the core is moved increases, while the induced voltage in the opposite coil decreases. This action induces a differential voltage output that varies linearly with core position. The phase of this output voltage changes abruptly by 180° as the core is moved from one side of the null to the other.

8.10.1 Extreme Environments

With increasingly sophisticated technology, more and more instrumentation applications have arisen for sensors capable of operating in such hostile environments as extremely cold temperatures, very high temperatures, and/or intense nuclear radiation. The LVDT has been developed with materials that can tolerate extreme environmental

requirements. Although the operating environments vary greatly, these LVDT designs use similar materials of construction and share the same physical configurations.

Currently, these LVDT sensors are built entirely with inorganic materials. The coil form is made of dimensionally stable, fired ceramic wound with ceramic-insulated high-conductivity magnet wire specially formulated for the application. Joints between the windings and lead wires are brazed or welded for mechanical reliability and electrical continuity. Ceramic cements and fillers are chosen to optimize heat transfer and bonding between windings, coil form, and housing. The potted assembly is cured at elevated temperatures, fusing the components together into a solidified structure.

Most inorganic insulations tend to be hygroscopic by nature, so the cured coil assembly is encased in an evacuated stainless steel shell that is hermetically sealed by *electron beam* (EB) welding. This evacuation and sealing process prevents moisture accumulation and subsequent loss of the insulation's dielectric strength. It also seals out surrounding media from the windings, while permitting the core to move freely.

Electrical connections are made to the windings with nickel conductors mutually insulated from each other by magnesium oxide filler and sheathed in a length of stainless-steel tubing. This cable assembly can be terminated by a hermetically sealed header for a connector when the application requires it.

The preceding description gives a brief insight into the material and techniques currently used in constructing the sensor for extremely severe environments. However, the state of the art in materials technology is being continually advanced. As new materials and methods of construction are evaluated, tested, and proved to upgrade performance, they will be incorporated into these sensors.

8.10.2 Cryogenic Manufacturing Applications

An LVDT sensor connected to the gripper of a robot is designed to cover a wide range of cryogenic applications ranging from general scientific research to space vehicle analysis and cryogenic medicine. A significant feature of the LVDT sensor is its ability to withstand repeated temperature cycling from room ambient conditions to the liquefaction temperatures of atmospheric gases such as nitrogen and oxygen. In order to survive such rigorous temperature changes, the sensor is constructed of materials selected for compatible coefficients of expansion while maintaining good electrical and magnetic properties even at –450°F (–270°C). The evacuated and hermetically sealed stainless-steel case prevents damage that could otherwise result from repeated condensation, freezing, and revaporization. Internal magnetic and electrostatic shielding renders the sensor insensitive to external magnetic and electrical influences.

8.10.3 Measurement at High Temperatures in Manufacturing

The LVDT sensor has been developed for measurements involving very high temperatures. It is capable of operating continuously at 1100°F (600°C) and surviving temperatures as high as 1200°F (650°C) for several hours. Typical uses include position feedback from jet engine controls located close to exhaust gases and measurement of roller position and material thickness in hot strip or slabbing mills. In scientific research, it can be used to directly measure dimensional changes in heated test specimens without requiring thermal isolation, which could induce measurement errors. The sensor is the culmination of the development of sophisticated construction techniques coupled with careful selection of materials that can survive sustained operation at high temperatures. Because magnetic properties of a metal vanish above its magnetic transformation temperature (Curie point), the core material must be made from one of the few magnetic materials having Curie temperatures above 1100°F (600°C). Another problem is that, at high temperature, the resistance of windings made of common magnet wire materials increases so much that an LVDT sensor using ordinary conductor materials would become virtually useless. Thus, the winding uses a wire of specially formulated high-conductivity alloy. The sensors are made with careful attention to internal mechanical configuration and with materials having compatible coefficients of expansion to minimize null shifts due to unequal expansion or unsymmetrical construction. Hermetic sealing allows the sensor to be subjected to hostile environments such as fluid pressure up to 2500 psi (175 bars) at 650°F (350°C). Units can be factory calibrated in a special autoclave that permits operation at high temperature while they are hydrostatically pressurized.

8.11 Robot Control through Sensors

In order to pick up an object, a robot must be able to sense the strength of the object being gripped so as not to crush the object. Accordingly, the robot gripper is equipped with sensing devices to regulate the amount of pressure applied to the object being retrieved.

Several industrial sensing devices enable the robot to place objects at desired locations or perform various manufacturing processes:

- *Transducers.* Sensors that convert nonelectrical signals into electrical energy
- *Contact sensors (limit switches).* Switches designed to be turned ON or OFF by an object exerting pressure on a lever or roller that operates the switch
- *Noncontact sensors.* Devices that sense through changes in pressure, temperature, or electromagnetic field

- *Proximity sensors.* Devices that sense the presence of a nearby object by inductance, capacitance, light reflection, or eddy currents
- *Range sensors.* Devices such as laser-interferometric gauges that provide a precise distance measurement
- *Tactile sensors.* Devices that rely on touch to detect the presence of an object; strain gauges can be used as tactile sensors
- *Displacement sensors.* Provide the exact location of a gripper or manipulator. Resistive sensors are often used—usually wire-wound resistors with a slider contact; as force is applied to the slider arm, it changes the circuit resistance
- *Speed sensors:* devices such as tachometers that detect the motor shaft speed
- *Torque sensors.* Measure the turning effort required to rotate a mass through an angle
- *Vision sensors.* Enable a robot to see an object and generate adjustments suitable for object manipulation; include dissectors, flying-spot scanners, vidicons, orthicons, plumbicons, and charge-coupled devices

8.12 Multisensor-Controlled Robot Assembly

Most assembly tasks are based on experience and are achieved manually. Only when high volume permits are special-purpose machines used. Products manufactured in low-volume batches or with a short design life can be assembled profitably only by general-purpose flexible assembly systems that are adaptable and programmable. A computer-controlled multisensor assembly station responds to these demands for general-purpose assembly. The multisensor feedback provides the information by which a robot arm can adapt easily to different parts and accommodate relative position errors.

The assembly task may be viewed as an intended sequence of elementary operations able to accept an originally disordered and disorganized set of parts and to increase gradually their order and mating degrees to arrive finally at the organization level required by the definition of the assembly. In these terms, it may be considered that assembly tasks perform two main functions: (1) ordering of parts and (2) mating of parts in the final assembly.

The ordering function reduces uncertainty regarding assembly parts by supplying information about their parameters, type, and position/orientation. The performance criterion of this function may be expressed as an entropy. The final goal of ordering is to minimize the relative entropy sum of the components of the assembly. In an assembly station, the ordering is performed either in a passive way

by sensors or in an active way by mechanical means such as containerization, feeding, fixturing, or gripping.

The part mating function imposes successive modifications of the positions of the parts in such a way as to mate them finally in the required assembly pattern. The modification of the position of the parts is carried out by manipulating them by transport systems and robot arms. The part mating function requires *a priori* information about the parts to be manipulated and the present state of the assembly. During this manipulation, the part entropy may increase because of errors in the robot position or by accidental changes in the manipulated parts.

An example of a multisensor system is a test bed for research on sensor control of robotic assembly and inspection, particularly comparisons of passive entropy reduction to active mechanical means. Programs for the system must be easy to develop and must perform in real time. Commercially available systems and sensors had to be used as much as possible. The system had to be very flexible in interfacing with different subsystems and their protocols. The flexible assembly system (Fig. 8.32) consists of three main parts:

- Robot equipped with a transport system
- Vision sensors, force-torque sensors, and ultrasound sensors
- Control system

The system was built with a *variable transport system* (VTS) (Fig. 8.33). It is a modular system in which product carriers are transported to the system. Two docking stations that can clamp the product carriers are included in the system. Some of the product carriers have translucent windows to allow backlighting of the vision system. The VTS system has its own controller operated by parallel input/output lines.

The robot is equipped with six degrees of freedom and a payload capacity of 4 kg. The absolute accuracy is 0.2 mm. The robot controller consists of two 68000 microprocessors. The standard way of programming the robot is through a teach pendant.

Three different interfaces can connect the robot and control system: (1) computer link at 9600 baud (RS-232), (2) adaptive control, and (3) parallel interrupt lines.

The computer link lets the control computer give commands to the robot such as:

- Load programs from disk
- Start programs
- Set position registers
- Read out robot positions
- Take direct control of robot movement
- Up- and download robot programs

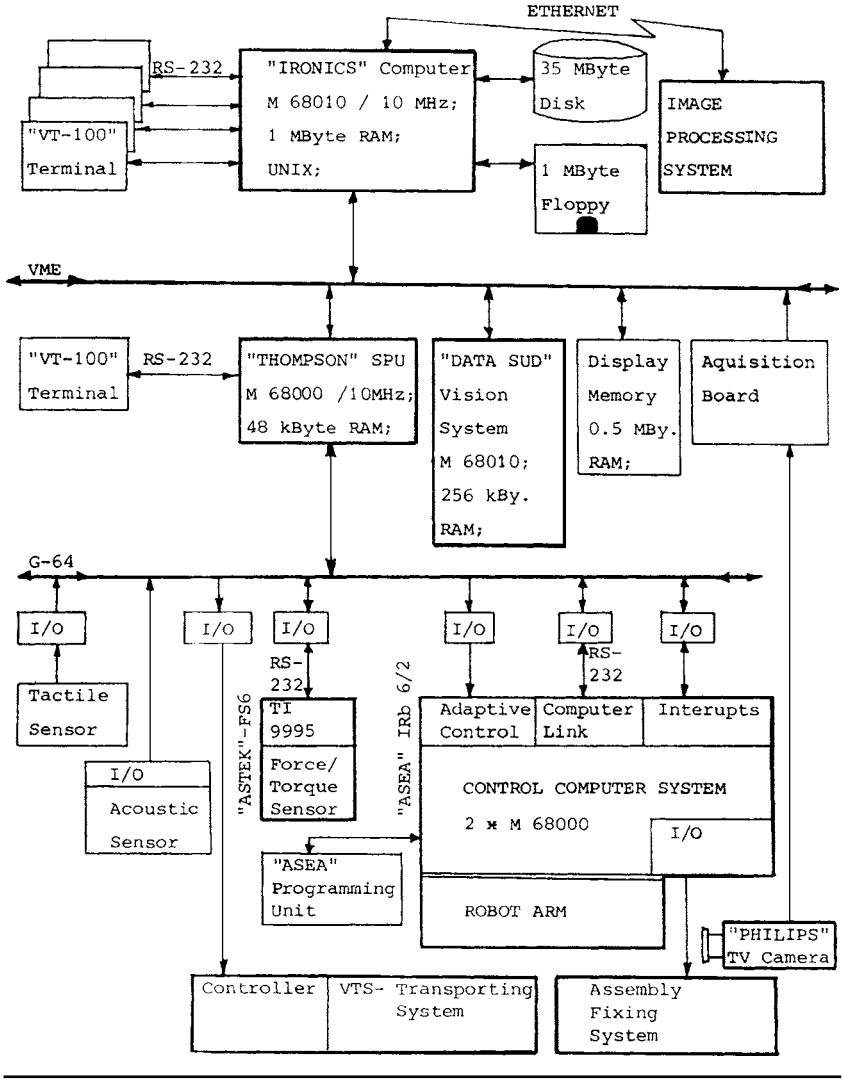


FIGURE 8.32 Structure of a flexible manufacturing system.

The adaptive control offers search and contour-following modes. Up to three sensor signals can be used to guide the movements of the robot. The direction in which the robot is to move must be given for each signal. The displacement or velocity is proportional to the signal value generated by the displacement sensor. Both digital and analog signals produced by sensors are accepted. Digital signals may consist of 1 to 8 bits.

The interrupt lines are used to start or stop a robot movement and to synchronize tasks.

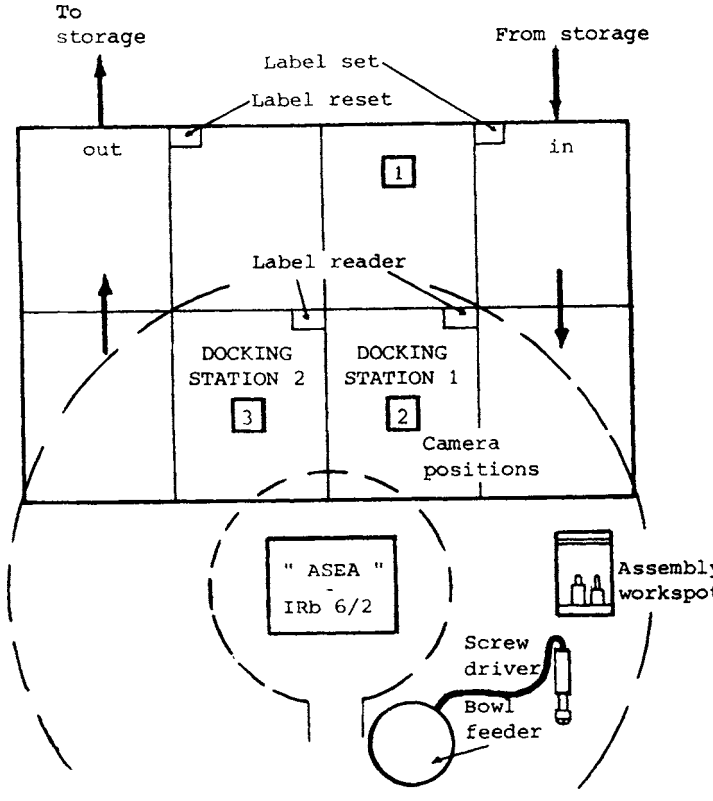


FIGURE 8.33 Layout of a robot assembly and variable transport system.

8.12.1 Control Computer

The computer used to control the system is microprocessor-based. The *main processor unit* (MPU) is a VME-bus-compatible single-board computer featuring a 68010 processor, 1-Mb dual-ported RAM, a 35-Mb Winchester disk, and a 1-Mb floppy disk. The system runs under the UNIX V operating system.

Real-time performance is obtained by two special processing units (SPUs). The first SPU realizes the control and interfacing of the robot and the interfacing of the sensors. It links the VME bus to the G64 bus to which the robot, the VTS, the torque-force sensor, and the ultrasound sensor are interfaced. This processor controls the computer link and the adaptive control of the robot. The processor board consists of a 68000 processor and 48-kb RAM. The second SPU is the vision processing board. This board communicates with the vision boards and the other SPU via the VME bus. It consists of a 68010 processor, 256-kb RAM, and an NS16081 floating-point processor.

8.12.2 Vision Sensor Modules

The vision module consists of two Primagraphics VME boards. One board contains the frame image of 768 × 576 pixels and 128 gray values. A frame-transfer solid-state camera is connected to the system.

Force sensing is available and can measure three force and three torque components. Forces up to 200 N are measured with a resolution of 0.1 N. Torques up to 4 N·m is measured with a resolution of 0.002 N·m. The data rate ranges from 1.8 to 240 Hz. The torque-force sensor is controlled by a Texas Instruments 9995 microprocessor. Besides a high-speed RS-232 link 38400 board, an analog output is available.

An ultrasonic range finder sensor has been interfaced to the G64 bus. A tactile array based on the piezoelectric material PVDF has been developed. A stereo vision system is incorporated to the system.

8.12.3 Software Structure

Because an important aim of the testbed is to investigate the use of sensors in an assembly process, it is essential that an optimal environment is present for program development and real-world testing. A user-friendly program environment is provided by the UNIX V operating system. Real-time operation is provided by the special processing units. Programs developed under the UNIX system can be downloaded to the SPUs via the VME bus.

The assembly process consists of a sequence of assembly stages. A succeeding assembly stage is entered when certain entrance conditions are met. This means that the assembly process can be controlled by a state machine. An assembly stage may consist of a number of states. With this approach, a modular and hierarchical control structure is obtained in which certain processes are executed at the lowest level.

The control software for the assembly task consists of three different levels: (1) state machine level, (2) executor level, and (3) driver level.

At the highest level, a command interpreter activates certain states of the machine. The state machine activates another (nested) state machine or a process (executor). Executors may run in parallel. For instance, the vision system may analyze an image of a product while the robot is assembling. The synchronization of the processes is obtained from the required entrance conditions for the next state of the state machine. In general, this level requires a moderately fast response and may run under the UNIX system.

At the second level, the executors are activated by the state machine. An executor performs a certain subtask and requires a real-time response. Examples are robot movements and image analysis routines. The executors reside in the SPUs and in the robot controller (robot programs). An executor is the only program that communicates directly with a driver.

The drivers for the different subsystems form the third level and also reside in the SPUs. The drivers take care of the protocol conversion and of the error checking in the communication with the subsystems.

During the assembly task, a number of processes (executors) are running, requiring synchronization of data transfer. The data may consist of, for instance, the type, position, and orientation of a detected part by the vision system. The communication between the processes is realized through a shared memory lock in the VME bus.

8.12.4 Vision Sensor Software

The first step consists of a segmentation of gray-value image into a binary image. Because backlighting is used in the system, contrast is excellent and simple thresholding is sufficient for segmentation. During initialization, a threshold is calculated from the image histogram and used until the next initialization takes place. Before the system starts an assembly operation, the vision system must be calibrated to the robot coordinate system. This is done by moving a ring in the vision field. From the known robot coordinates, the ring position, and the computed ring coordinates in the vision system, the necessary transformation is calculated.

A drawback of connectivity analysis is that objects may not touch. The vision package may be extended with graph-matching techniques to allow recognition of touching and overlapping parts.

8.13 History of Industrial Robotics

The word *robot* was introduced to the public by Czech writer Karel Capek in his play *R.U.R. (Rossum's Universal Robots)*, which premiered in 1921 (Fig. 8.34). The play begins in a factory that makes artificial people called *robots*, but they are closer to the modern ideas of

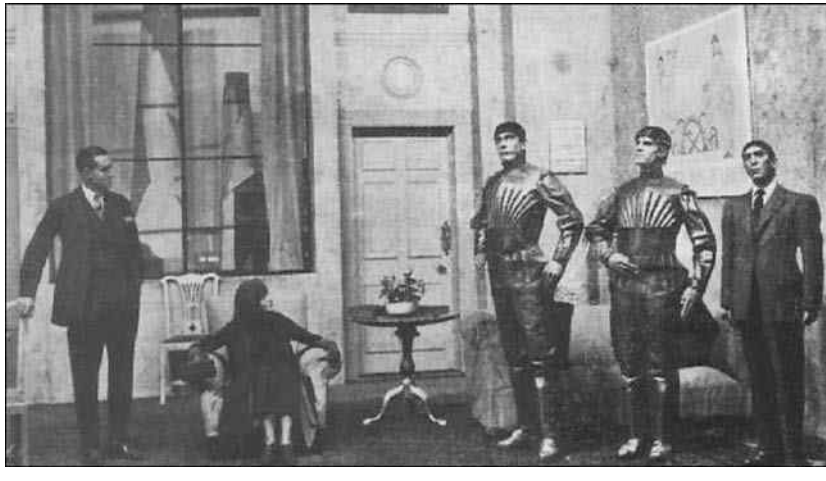


Figure 8.34 A scene from Karel Capek's 1920 play R.U.R. (Rossum's Universal Robots), showing three robots.

androids and clones, creatures who can be mistaken for humans. They can plainly think for themselves, though they seem happy to serve. At issue is whether the *robots* are being exploited and the consequences of their treatment.

However, Karel Capek himself did not coin the word; he wrote a short letter in reference to an etymology in the Oxford English Dictionary in which he named his brother, the painter and writer Josef Capek, as its actual originator. In an article in the Czech journal *Lidové noviny* in 1933, he explained that he had originally wanted to call the creatures *labori* (from Latin labor, work). However, he did not like the word, and sought advice from his brother Josef, who suggested "roboti." The word *robota* means literally work, labor or serf labor, and figuratively "drudgery" or "hard work" in Czech and many Slavic languages. Serfdom was outlawed in 1848 in Bohemia, so at the time Capek wrote R.U.R., usage of the term robota had broadened to include various types of work, but the obsolete sense of "serfdom" would still have been known.

8.13.1 Early Modern Developments

Leonardo da Vinci (1452–1519) sketched plans for a humanoid robot around 1495. Da Vinci's notebooks, rediscovered in the 1950s, contain detailed drawings of a mechanical knight now known as Leonardo's robot, able to sit up, wave its arms, and move its head and jaw. The design was probably based on anatomical research recorded in his Vitruvian Man. It is not known whether he attempted to build it.

In 1738 and 1739, Jacques de Vaucanson exhibited several life-sized automatons: a flute player, a pipe player, and a duck. The mechanical duck could flap its wings, crane its neck, and swallow food from the exhibitor's hand, and it gave the illusion of digesting its food by excreting matter stored in a hidden compartment. Complex mechanical toys and animals built in Japan in the 1700s were described in the *Karakuri zui* (*Illustrated Machinery*, 1796). (See Fig. 8.35.)

8.13.2 Modern Developments

The Japanese craftsman Hisashige Tanaka (1799–1881), known as "Japan's Edison," created an array of extremely complex mechanical toys, some of which served tea, fired arrows drawn from a quiver, and even painted a Japanese *kanji* character. In 1898, Nikola Tesla publicly demonstrated a radio-controlled torpedo. Based on patents for "teleautomation," Tesla hoped to develop it into a weapon system for the U.S. Navy.

In 1926, Westinghouse Electric Corporation created Televox, the first robot put to useful work. They followed Televox with a number of other simple robots, including one called Rastus, made in the crude image of a black man. In the 1930s, they created a humanoid robot known as Elektro for exhibition purposes, including the 1939 and 1940 World's Fairs. In 1928, Japan's first robot, Gakutensoku, was designed



Figure 8.35
Tea-serving karakuri, with mechanism, 19th century. *Tokyo National Science Museum.*

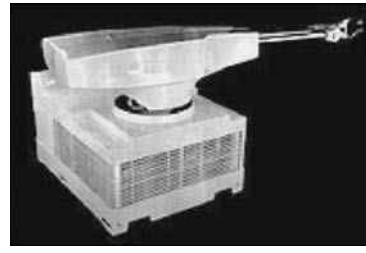


Figure 8.36
The first Unimate.

and constructed by biologist Makoto Nishimura.The first electronic autonomous robots were created by William Grey Walter of the Burden Neurological Institute at Bristol, England in 1948 and 1949. They were named *Elmer* and *Elsie*. These robots could sense light and contact with external objects, and use these stimuli to navigate.

The first truly modern robot, digitally operated and programmable, was invented by George Devol in 1954 and was ultimately called the Unimate (Fig. 8.36). Devol sold the first Unimate to General Motors in 1960, and it was installed in 1961 in a plant in Trenton, New Jersey to lift hot pieces of metal from a die casting machine and stack them.

8.14 The Invention of Industrial Robots

George Devol applied for the first robotics patents in 1954 (granted in 1961). The first company to produce a robot was Unimation, founded by George Devol and Joseph F. Engelberger in 1956, and was based on Devol's original patents. Unimation robots, Fig. 8.36, were also called programmable transfer machines since their main use at first was to transfer objects from one point to another, less than a dozen feet or so apart. They used hydraulic actuators and were programmed

in joint coordinates (i.e., the angles of the various joints were stored during a teaching phase and replayed in operation). They were accurate to within 1/10,000 of an inch. However, accuracy is not an appropriate measure for robots. Usually, they are evaluated in terms of repeatability. Unimation later licensed their technology to Kawasaki Heavy Industries and Guest-Nettlefolds, manufacturing Unimates in Japan and England, respectively. For some time, Unimation's only competitor was Cincinnati Milacron Inc. of Ohio. This changed radically in the late 1970s when several big Japanese conglomerates began producing similar industrial robots.

In 1969, Victor Scheinman at Stanford University invented the Stanford arm, an all-electric, six-axis articulated robot designed to permit an arm solution. This allowed it to accurately follow arbitrary paths in space and widened the potential use of the robot to more sophisticated applications such as assembly and welding. Scheinman then designed a second arm for the MIT AI Lab, called the "MIT arm." Scheinman, after receiving a fellowship from Unimation to develop his designs, sold those designs to Unimation who further developed them with support from General Motors and later marketed it as the *Programmable Universal Machine for Assembly* (PUMA). In 1973, KUKA Robotics built its first robot, known as FAMULUS, this is the first articulated robot to have six electromechanically driven axes.

Interest in robotics increased in the late 1970s, and many U.S. companies entered the field, including large firms like General Electric, and General Motors (which formed joint venture FANUC Robotics with FANUC LTD of Japan). U.S. startup companies included Automatix and Adept Technology, Inc. At the height of the robot boom in 1984, Unimation was acquired by Westinghouse Electric Corporation for 107 million U.S. dollars. Westinghouse sold Unimation to Stäubli Faverges SCA of France in 1988, which is still making articulated robots for general industrial and cleanroom applications, and even bought the robotic division of Bosch in late 2004. Recently, several research and development organizations are working on unmanned robots performing remote surgery operations in particularly at harsh and severe environmental conditions and events, Fig. 8.37.

The author of this book was the first entrepreneur to incorporate several robotic systems in synchronous rotary indexing assembly machines, and asynchronous assembly and fabrication systems. An electronic memory typewriter was fully assembled using the robotic systems fitted with traditional assembly machines for QYX, Exxon Corporation. Also, the author was the first to incorporate highly flexible robotic applications in the manufacture and assembly of electromagnetic media drives at a rate of 3 units per second for Control Data Corporation USA. In addition, the author successfully pioneered the implementation of several robotic and vision applications in the meat, pork, and poultry industry for ConAgra, Swift and Co., and Monfort Pork and Beef Corporation.



Figure 8.37 An unmanned laparoscopic robotic surgery machine.

Only a few non-Japanese companies ultimately managed to survive in this market, the major ones being Adept Technology, Stäubli-Unimation, the Swedish-Swiss company ABB Asea Brown Boveri, and the German company KUKA Robotics.

8.14.1 Technical Description

The following are several definitions of robotic terms.

- *Number of axes.* Two axes are required to reach any point in a plane; three axes are required to reach any point in space. To fully control the orientation of the end of the arm (i.e. the wrist) three more axes (yaw, pitch, and roll) are required. Some designs (e.g., the SCARA robot) trade limitations in motion possibilities for cost, speed, and accuracy.

- *Degrees of freedom.* Usually the same as the number of axes.
- *Working envelope.* The region of space a robot can reach.
- *Kinematics.* The actual arrangement of rigid members and joints in the robot, which determines the robot's possible motions. Classes of robot kinematics include articulated, Cartesian, parallel, and SCARA.
- *Carrying capacity or payload.* How much weight a robot can lift.
- *Speed.* How fast the robot can position the end of its arm. This may be defined in terms of the angular or linear speed of each axis or as a compound speed (i.e., the speed of the end of the arm when all axes are moving).
- *Acceleration.* How quickly an axis can accelerate. Since this is a limiting factor, a robot may not be able to reach it's specified maximum speed for movements over a short distance or a complex path requiring frequent changes of direction.
- *Accuracy.* How closely a robot can reach a commanded position. Accuracy can vary with speed and position within the working envelope and with payload (*see* Compliance). It can be improved by robot calibration.
- *Repeatability.* How well the robot will return to a programmed position. This is not the same as accuracy. It may be that when told to go to a certain X-Y-Z position, it gets to within only 1 mm of that position. This would be its accuracy, which may be improved by calibration. But if that position is taught into controller memory and each time it is sent there it returns to within 0.1 mm of the taught position, then the repeatability will be within 0.1 mm.
- *Motion control.* For some applications, such as simple pick-and-place assembly, the robot need merely return repeatably to a limited number of pre-taught positions. For more sophisticated applications, such as welding and finishing (spray painting), motion must be continuously controlled to follow a path in space, with controlled orientation and velocity.
- *Power source.* Some robots use electric motors, others use hydraulic actuators. The former are faster, the latter are stronger and advantageous in applications such as spray painting where a spark could set off an explosion; however, low internal air-pressurization of the arm can prevent ingress of flammable vapors as well as other contaminants.
- *Drive.* Some robots connect electric motors to the joints via gears; others connect the motor to the joint directly (direct drive). Using gears results in measurable "backlash," which is free movement in an axis. In smaller robot arms with DC electric motors, because DC motors are high-speed low-torque motors,

they frequently require high ratios causing backlash to be a problem. In such cases, the harmonic drive is often used.

- *Compliance.* This is a measure of the amount in angle or distance that a robot axis will move when a force is applied to it. Because of compliance when a robot goes to a position carrying it's maximum payload, it will be at a position slightly lower than when it is carrying no payload. Compliance can also be responsible for overshoot when carrying high payloads, in which case acceleration would need to be reduced.

8.14.2 Robot Programming and Interfaces

- *Main article–robotics suite.* The setup or programming of motions and sequences for an industrial robot is typically taught by linking the robot controller to a laptop, desktop computer, or (internal or Internet) network, Fig. 8.38.
- *Software.* The computer is installed with corresponding interface software. The use of a computer greatly simplifies the programming process. Specialized robot software is run either in the robot controller or in the computer or both depending on the system design.
- *Teach pendant.* Robots can also be taught via a teach pendant, Fig. 8.39; a handheld control and programming unit.

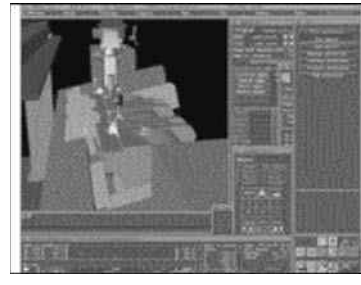


Figure 8.38
Offline programming by ROBCAD.



Figure 8.39
A typical well-used teach pendant with optional mouse.

The common features of such units are the ability to manually send the robot to a desired position, or "inch" or "jog" to adjust a position. They also have a means to change the speed since a low speed is usually required for careful positioning or while test-running through a new or modified routine. A large emergency stop button is usually included as well. Typically, once the robot has been programmed there is no more use for the teach pendant.

- *Lead-by-the-nose.* Lead-by-the-nose is a technique offered by most robot manufacturers. While the user holds the robot end effector, another person enters a command that de-energizes the robot and it goes limp. The user then moves the robot by hand to the required positions or along a required path, while the software logs these positions into memory. The program can later run the robot to these positions or along the taught path. This technique is popular for tasks such as paint spraying.

In addition, machine operators often use user interface devices, typically touch screen units, which serve as the operator control panel. The operator can switch from program to program, make adjustments within a program and also operate a host of peripheral devices that may be integrated within the same robotic system. These include end effectors, feeders that supply components to the robot, conveyer belts, emergency stop controls, machine vision systems, safety interlock systems, bar code printers, and an almost infinite array of other industrial devices that are accessed and controlled via the operator control panel.

The teach pendant or PC is usually disconnected after programming and the robot then runs on the program that has been installed in its controller. However, a computer is often used to "supervise" the robot and any peripherals, or to provide additional storage for access to numerous complex paths and routines.

A robot and a collection of machines or peripherals is referred to as a work cell, or cell. A typical cell might contain a parts feeder, a molding machine and a robot. The various machines are "integrated" and controlled by a single computer or PLC.

8.14.3 End Effectors

The most essential robot peripheral is the end effector, or end-of-arm-tooling. Common examples of end effectors include welding devices, such as MIG-welding guns, spot-welders, and others; spray guns, grinding and deburring devices (such as pneumatic disk or belt grinders, burrs, etc.); and grippers, devices that can grasp an object, and that are electromechanical or pneumatic. Another common means of picking up an object is by vacuum. End effectors are frequently highly complex, made to match the handled product, and are

often capable of picking up an array of products at one time. They may utilize various sensors to aid the robot system in locating, handling, and positioning products.

8.14.4 Movement and Singularities

Most articulated robots perform by storing a series of positions in memory, and moving to them at various times in their programming sequence. For example, a robot that is moving items from one place to another might have a simple "pick and place" program similar to the following:

Define points P1–P5:

- Safely above workpiece (defined as P1)
- 10 cm Above bin A (defined as P2)
- At position to take part from bin A (defined as P3)
- 10 cm Above bin B (defined as P4)
- At position to take part from bin B (defined as P5)
- Define program:
- Move to P1
- Move to P2
- Move to P3
- Close gripper
- Move to P2
- Move to P4
- Move to P5
- Open gripper
- Move to P4
- Move to P1 and finish

For a given robot, the only parameters necessary to completely locate the end effector (gripper, welding torch, etc.) of the robot are the angles of each of the joints or displacements of the linear axes (or combinations of the two for robot formats such as SCARA). However, the points can be defined in many different ways. The most common and convenient way of defining a point is to specify a Cartesian coordinate for it (i.e., the position of the "end effector" in mm in the X, Y, and Z directions relative to the robot's origin). In addition, depending on the types of joints a particular robot may have, the orientation of the end effector in yaw, pitch, and roll and the location of the tool point relative to the robot's faceplate must also be specified. For a jointed arm, these coordinates must be converted to joint angles by the robot controller, and such conversions are known as Cartesian

Transformations, which may need to be performed iteratively or recursively for a multiple axis robot. The mathematics of the relationship between joint angles and actual spatial coordinates is called kinematics.

Positioning by Cartesian coordinates may be done by entering the coordinates into the system or by using a teach pendant, which moves the robot in X-Y-Z directions. It is much easier for a human operator to visualize motions up/down, left/right, and so forth than to move each joint one at a time. When the desired position is reached, it is then defined in some way particular to the robot software in use (e.g., P1–P5 cited earlier).

The American National Standard for Industrial Robots and Robot Systems—Safety Requirements (ANSI/RIA R15.06-1999) defines a singularity as "a condition caused by the collinear alignment of two or more robot axes resulting in unpredictable robot motion and velocities." It is most common in robot arms that utilize a "triple-roll wrist." This is a wrist about which the three axes of the wrist, controlling yaw, pitch, and roll, all pass through a common point. An example of a wrist singularity is when the path through which the robot is traveling causes the first and third axes of the robot's wrist to line up. The second wrist axis then attempts to spin 360° in zero time to maintain the orientation of the end effector. Another common term for this singularity is a "wrist flip." The result of a singularity can be quite dramatic and can have adverse effects on the robot arm, the end effector, and the process. Some industrial robot manufacturers have attempted to sidestep the situation by slightly altering the robot's path to prevent this condition. Another method is to slow the robot's travel speed, thus reducing the speed required for the wrist to make the transition. The ANSI/RIA has mandated that robot manufacturers shall make the user aware of singularities if they occur while the system is being manually manipulated.

8.15 Robot Programming

The robot is programmed in the teach-in mode with a standard pendant. Different subprograms, which are called from the control computer, are thus realized. The robot has 100 registers in which positions and orientations can be stored by a control computer. In this way, position and orientation determined by the vision module can be transferred to the robot controller. The search and contour tracing options of the robot controller are used for adaptive control.

The products arrive in a varying order and in an undefined position. Because the parts (except for bolts) are supplied on a flat surface in their most stable position (i.e., resting on their flat side), they have three degrees of freedom, namely the x and y coordinates and the orientation angle. The different parts are recognized by the vision



Figure 8.40 Vehicle assembly and inspection robots.

system on the basis of area, perimeter, and position of the holes in the objects. The vision system is often placed adjacent to the robot manipulator, Fig. 8.40. The robot controller is fully integrated with the vision controller to synchronize part image with robot handling of such parts from their identified coordinates and designated gripping locations. The position and orientation of parts are located from the center of gravity and the part hole positions.

Often the vision system is deployed with a robot wheeled system vital to detonate an explosive device, Fig. 8.41. An automated guided vehicle system is another example of a robotic system equipped



Figure 8.41 A U.S. Marine Corps technician prepares to deploy a device that will detonate a buried improvised explosive device near Camp Fallujah, Iraq.

with a lot of sensors to guide the system to deliver documents, mails, and light objects to various work stations identified by certain coding techniques, Fig. 8.42. Employing robust proximity sensors, together with servo motors and close loop servo techniques have contributed to the success of robotic vacuum cleaners, Fig. 8.43.

8.16 The 2006 World Robot Market—Total Worldwide Sales

The world market decreased by 11 percent in 2006. After the peak in 2005, the world market was down by 11 percent in 2006, at 112,203 newly supplied industrial robots. Nevertheless, developments were quite dissimilar in the three large industrial regions of Europe, America, and Asia, as shown in Fig. 8.44. After huge investments the previous year, robot sales in Asia and America plummeted in 2006. Europe recovered after a weak year in 2005. In 2006, world-wide shipments to the automotive industry decreased by 17 percent compared to 2005. Across the total electrical/electronics industry (including office and computing machinery and equipment; radio, TV, and communication devices and equipment; and medical, precision, and optical instruments), installations surged by 88 percent in 2005, to 28,600 units. This was the second year in a row that has seen huge increases in robot installations in these industries. Thus, it is not surprising that



Figure 8.42 Automated guided vehicle carrying various medical supplies and records.

in 2006, investment slowed down. Robot sales to the fabricated metal products industry, the chemical industry, and the food industry increased substantially.

In 2006, more than 61,700 robots were supplied to Asian countries (including Australia and New Zealand), about 19 percent fewer than in 2005. The electrical/electronics industry in Asia, which invested very heavily in 2005, cut robot purchases by half. Supplies to the automotive industry also decreased slightly.

Supplies to all other industries—in particular the metal products and machinery industry, the chemical industry, and the food industry—surged.

In Japan, supplies fell by 26 percent, to about 37,400 units. After the substantial investments within the automotive and electrical/electronics industry in 2007, purchases in both sectors were down in 2008.

The sharp fall was compensated by increased sales to the machinery and metal industry as well as to the chemical industry.



Figure 8.43 The Roomba domestic vacuum cleaner robot does a menial job.

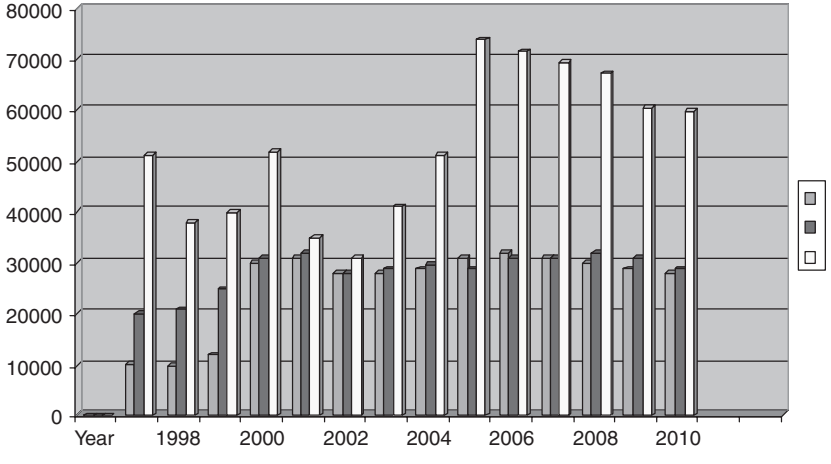


Figure 8.44 Estimated yearly shipments of industrial robots.

In 2009, about 11,800 industrial robots were ordered for destinations in the Republic of Korea, down 17 percent from 2008. In 2008, the supply was about 12,800 units, which was largely the result of strong demand from the electronic components industry and more complete reporting on robots in this particular industry. In 2008, this sector cut investments by almost 70 percent. This sharp fall has been somewhat offset by a huge increase in robot supplies to the automotive industry and the communications industry.

China was the third largest robot market in Asia, with 5,800 newly installed robots, about 29 percent more than in 2008. The automotive industry is still the predominant user of industrial robots, but the rubber and plastics industry and the electrical/electronics industry are gaining importance. In 2008, the number of industrial robots supplied in India almost doubled, to about 950 robots. At the moment, the actual numbers are quite small, but this escalation in the supply of robots testifies to the dynamism of the Indian market. Table 8.5 gives the installation and operational stock for 2005-2010.

Total supplies in all other Asian markets, including, Indonesia, Malaysia, the Philippines, Singapore, Taiwan, Thailand, and Vietnam, decreased by 8.5 percent.

In America, robotics investments were down by 19 percent, to about 17,990 units.

Although robot sales to the chemical industry, including the rubber and plastic products industry, increased remarkably in 2008, these increases were not sufficient to offset the sharp cyclical decline in

	Yearly installations				Operational stock at year-end			
Country	**2005**	**2006**	**2007**	**2010**	**2005**	**2006**	**2007**	**2010**
America	**21,986**	**17,910**	**21,400**	**24,400**	**143,634**	**154,680**	**167,100**	**209,000**
North America (Canada, Mexico, USA)	21,567	17,417	20,500	23,000	139,984	150,725	162,400	200,900
Central and South America	419	493	900	1,400	3,650	3,955	4,700	8,100
Asia/Australia	**76,047**	**61,748**	**66,000**	**75,000**	**481,652**	**479,027**	**500,500**	**579,900**
China	4,461	5,770	6,600	7,900	11,557	17,327	23,900	47,000
India	450	836	1,600	4,500	1,069	1,905	3,500	14,100
Japan	50,501	37,393	39,900	42,300	373,481	351,658	356,000	362,900
Republic of Korea	13,005	10,756	10,700	11,800	61,576	68,420	73,600	94,000
Taiwan, Province of China	4,096	4,307			15,464	19,204		
Thailand	1,458	1,102			2,472	3,574		
Other Asia	1,163	812			11,095	11,385		
Australia/New Zealand	913	772			4,938	5,554		
Europe	**28,432**	**31,536**	**35,000**	**39,000**	**296,918**	**315,624**	**329,800**	**380,000**
Austria	485	498			4,148	4,382		
Benelux	1,097	1,459			9,362	10,128		
Denmark	354	417			2,661	3,013		
Finland	556	321			4,159	4,349		
France	3,077	3,071	3,300	3,200	30,236	32,110	34,000	38,800
Germany	10,075	11,425	12,700	13,000	126,294	132,594	137,900	147,400
Italy	5,425	6,259	6,900	8,400	56,198	60,049	63,800	72,000
Norway	115	181			811	960		
Portugal	144	268			1,542	1,710		
Spain	2,709	2,409			24,141	26,008		
Sweden	939	865			8,028	8,245		
Switzerland	442	456			3,732	3,940		
Turkey	207	368			403	771		
United Kingdom	1,363	1,220	1,000	800	14,948	15,082	15,300	13,800
Central/Eastern European countries	1,287	1,322			9,446	10,781		
other Europe	157	995			809	1,502		
Africa	**204**	**426**	**700**	**900**	**634**	**1,060**	**1,700**	**4,400**
Total	**126,669**	**112,203**	**123,100**	**139,300**	**922,838**	**950,974**	**999,100**	**1,173,300**

Source: IFR, national robot associations and UNECE (up to 2004)

Table 8.4 Yearly Robot Installations and Operational Stocks 2005–2010

sales to the automotive sector. After the extraordinarily high investments seen in 2006, robotic purchases by automotive companies and their suppliers slowed in 2008 in the highly competitive U.S. and Canadian markets. Sales and production volumes of cars have stagnated or even declined over the last few years, which have also witnessed changes in the respective market shares of the individual manufacturers.

American suppliers have reduced capacity and will continue to do so. Nevertheless, the decrease of robot supplies in 2009 is likely to be only temporary, since their Asian and European competitors have announced additional investments in production facilities. In 2009, increasing demand for robots has already been registered. In the first half of 2008, North American orders rose by 39 percent according to the RIA (Robotics Industry Association), USA. The intense competition in the biggest and most saturated market is set to continue.

An increasing demand for industrial robots has been registered in Mexico, Argentina, and Brazil. The deliveries to these countries increased by 9.7 percent.

Sales of industrial robots in Europe were up by 11 percent, to about 31,900 units. This was the result of substantially increasing industrial robot investment by the metal products industry, the rubber and plastics industry, and the food and beverages industry. The motor vehicle industry and the automotive parts industry again reduced their purchases in Europe. In Germany—the biggest market for industrial robots in Europe—supplies increased by 12.7 percent to about 11,900 units. This was the result of surging investments in general industry—all industries except the automotive. Here as well, supply to the automotive industry slightly decreased. It seems that the peak of yearly robot investment within the automotive industry was reached in 2004: Germany as a production site for motor vehicles serves an almost saturated market in Western Europe.

Italy, the second largest market, recovered by 14 percent to almost 6,900 units. The rubber and plastics industry and the fabricated metal products industry invested very heavily in industrial robots. Sales to all other industries, including the automotive industry, decreased.

The yearly supply of industrial robots in France has been fairly stable over the last five years, at around 3,200 units. The reason for this similar figure has been the largely stagnant investment climate in the automotive industry and the machinery industry, which has been somewhat compensated by a significant increase of robot supplies to the metal industry and a remarkable increase of robot supplies to the chemical and food industries.

Sales to Spain, the United Kingdom, and Sweden were down. Of the Eastern European countries, only the Czech Republic and Poland saw a significant increase in robot supplies. Compared to the established robot markets in Western Europe, the Turkish one is still quite

minor, with 380 robots supplied in 2008, but it is growing in importance. Turkey is becoming an attractive production site for the international automotive industry, along with China, India, Malaysia, the Czech Republic, Slovakia, and Iran. Exports of motor vehicles and automotive parts are increasing rapidly. More motor vehicles are now assembled in Turkey than in the United Kingdom or Sweden.

Total accumulated yearly sales, measured since industrial robots started to be introduced in industry at the end of the 1960s, amounted to more than 1,980,000 units at the end of 2008. Many of the early robots, however, have by now been taken out of service.

The stock of industrial robots in actual operation is therefore lower. The total worldwide stock of operational industrial robots at the end of 2008 is estimated to be between 1,151,000 units, with a possible maximum of 1,350,000 units.

The average length of service life is 12 years. A pilot study has indicated that the average service life might in fact be as long as 15 years, which would result in a worldwide stock of 1,350,000 units.

When the minimum 2008 stock of 1,151,000 units is compared with the 1,116,400 units at the end of 2007, it represents an increase of 3 percent.

The world market for industrial robots is projected to increase from 112,200 units in 2006 to 123,100 in 2008. From 2008, it will rise by a yearly average of 4.2 percent to 139,300 in 2010, as shown Fig. 8.45.

In 2008, the worldwide supply of industrial robots has increase by 10 percent. Investments by the automotive industry will be more buoyant than in 2007: In Europe, increasing demand has already been registered, especially in Central/Eastern Europe, Italy, Sweden, and Germany. In North America, competition within the automotive industry will cause robot purchases to rise again. Asian car suppliers

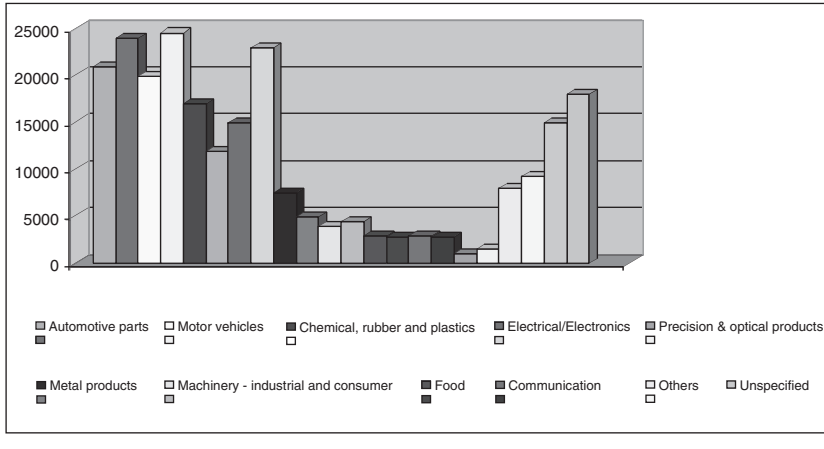


Figure 8.45 Forecasts for 2007–2010.

boosted their investment in the home market as well as overseas. Demand in China will grow more moderately. In India, Asia, Central/Eastern Europe, and South America, the robot supply will again increase at an above-trend rate. Robot supplies will stagnate in the Republic of Korea, while in Japan growth of about 7 percent can be expected. The optimistic view for 2010 in all three regions is partly a result of encouraging robot supplies to "general industry"—meaning, all industries except the automotive. After the decrease in 2006, demand from the electrical/electronics industry also should be higher in 2010.

Robust growth in robot shipments worldwide can be expected between 2009 and 2010.

Stagnating or only slowly growing investments by motor vehicle suppliers worldwide will be compensated—to a certain degree—by more vigorous demand across all other industries.

In terms of units, it is estimated that the worldwide stock of operational industrial robots will increase from about 1,151,000 units at the end of 2008 to 1,350,000 at the end of 2010, representing an average annual growth rate of 5.4 percent.

8.17 Measurements of Robot Density Based on the Total Number of Persons Employed

In 2001–2008, employment slightly decreased in most of the countries surveyed, while the stock of robots continued to increase. In Japan, the stock of operating robots as well as employment within the sector fell in the period 2001 to 2008.

In Japan and in the Republic of Korea, which collect data on all types of industrial robots and are therefore not comparable with other countries, they have a high density of robot installations. In 2008, 410 robots in Japan and 210 robots in the Rep. of Korea were in operation per 10,000 persons employed in the manufacturing industry.

With 186 robots per 10,000 employed in the manufacturing industry, Germany is the country with the highest robot density in Europe, followed by Italy with 138, and Sweden with 123. In Finland the density amounted to 101, followed by the United States with 99, France with 92, and Spain with 79 robots per 10,000 employed in the manufacturing industry. The densities in Denmark, Austria, the Benelux, Switzerland, Australia, and the United Kingdom ranged between 50 and 75. In Norway, the density was 33 and in Portugal 19. Countries in central and eastern Europe, with the exception of the Czech Republic, have even lower densities.

Despite this large range in the robot densities of the European countries mentioned, it is interesting to note that the robot density in Germany is about 90 percent higher than that of the United States.

8.17.1 Robot Densities—One Robot per Ten Workers in the Motor Vehicle Industry

Japan and Italy are in the lead with 1,820 robots and 1,630, respectively, per 10,000 workers. Thereafter follows Germany with a density of 1,220, France 1,160, Spain 970, United States 830, United Kingdom 600, and Sweden 590. The technological level with respect to robotics is thus rather homogeneous in the motor vehicle industry in most of the aforementioned countries.

8.17.2 Installations of Advanced Multipurpose Industrial Robots by Type

In 2008, 60 percent of the robots installed were articulated robots, up from 59 percent in 2007; 22 percent were linear/Cartesian/gantry robots, up from 20 percent in 2007; 4 percent were cylindrical robots, down from 12 percent in 2007; and 13 percent were SCARA robots, up from 8 percent in 2007. In terms of absolute numbers, the following figures for installations by types of robots were recorded: 67,478 articulated robots installed in 2008, 9 percent fewer than 2007; 24,148 linear/Cartesian/gantry robots, 4 percent fewer than 2007; 15,509 cylindrical robots, down 71 percent from 2007; and 1,867 SCARA robots, up 48 percent from 2007.

8.17.3 Distribution of Service Robots for Professional Use

Up to the end of 2008, service robots installed for professional use 41,000 units. With 9,895 units, service robots in defense, rescue, and security applications accounted for 23 percent of the total number of service robots for professional use installed up to the end of 2008. Thereafter, followed data of field robots (mainly milking robots) with 16 percent, and cleaning robots and underwater systems with 14 percent, each. Construction and demolition robots (10 percent), medical robots (9 percent), and mobile robot platforms for general use (8.4 percent) take up a secondary position. Minor installation numbers were counted for logistic systems (1,550 units), inspection systems (nearly 600 units), and public relation robots (about 100 units).

8.17.4 Distribution of Service Robots for Personal and Private Use

Approximately 2.47 million units for domestic use and about 1.15 million units for entertainment and leisure have been sold up to the end of 2008. Service robots for personal and domestic use are recorded separately since their unit value is only a fraction of that of many types of service robots for professional use. They are also produced for a mass market with completely different marketing

channels. Service robots for personal and domestic use are mainly in the areas of domestic (household) robots, which include vacuum cleaning and lawn-mowing robots, and entertainment and leisure robots, such as toy robots, hobby systems, and education and training robots.

The market for robots for handicap assistance is still small, but is expected to double in 2010. Robots for personal transportation and home security and surveillance robots will also increase in importance in the future.

Up to the end of 2008, accumulated sales of vacuum cleaning robots resulted in 2.37 million units. At the end of 2008, the stock of lawn-mowing robots amounted to 91,950 units.

8.18 Projections for the Period 2007–2010

Turning to the projections for the period 2007–2010, the stock of service robots for professional use is forecast to increase by some 35,900 units (Fig. 8.46). Application areas with strong growth are defense, rescue and security applications, field robots, cleaning robots, medical robots, and mobile robot platforms for multiple uses.

Around 3.7 million units of service robots for personal use are expected to be sold. It is projected that sales of all types of domestic robots (vacuum cleaning, lawn-mowing, window cleaning, and other types) in the period 2007–2010 could reach some 1.39 million units (Fig. 8.47). The market for entertainment and leisure robots, which includes toy robots, is forecast at about 2.25 million units, most of which, of course, are very low cost.

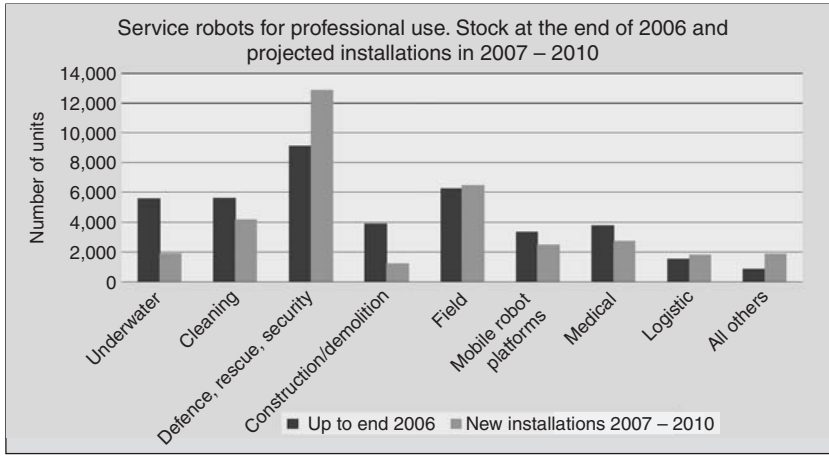


Figure 8.46 Robot installation projection 2006–2010 for industrial use.

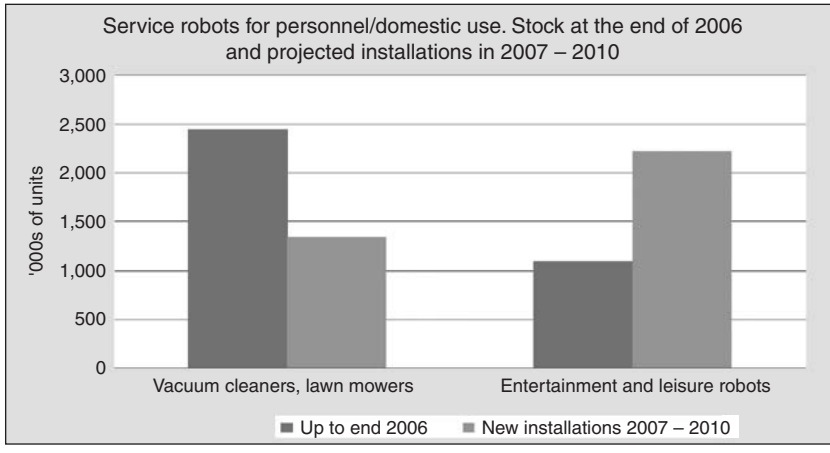


Figure 8.47 Robot installation projection 2006–2010 for domestic and personal use.

Further Reading

Ambler, A. P., et al., "A Versatile Computer-Controlled Assembly System," *Proc. of IJCAI-73*, Stanford, California, 298–307, August 1973.

Ballard, D. H., "Generalizing the Hough Transform to Detect Arbitrary Shapes," *Pattern Recognition*, 13(2): 111–112 (1981).

Barrow, H. G., and R. J. Popplestone, "Relational Descriptions in Picture Processing," *Machine Intelligence* 6, Edinburgh University Press, 1971.

"Decade of Robotics," 10th anniversary issue of *International Robot*, IFS Publications, 1983.

Dell'Acque, R., "A Bright Future for Thick Film Sensor," *SAE* 890480.

Duba, R. O., and P. E. Hart, "Use of Hough Transform to Detect Lines and Curves in Pictures," *Communications of AQCM*, 15(1): 11–15, January 1972.

DuPoint-Gateland, C. "Flexible Manufacturing Systems for Gearboxes," *Proc. 1st Int. Conf. on FMS*, Brighton, UK, 453–563, October 20–22, 1982.

Heintz, F., and E. Zabler, "Motor Vehicle Sensors Based on Film-Technology: An Interesting Alternative to Semiconductor Sensors," *SAE* 870477.

Freeman, H., "Computer Processing of Line-Drawing Images," *Computing Survey* 6(1), March 1974.

Gilbert, J. L., and V. Y. Paternoster, "Ultrasonic End Effector for Weld Inspection," *NASA Tech Briefs*, May 1992.

Hench, L. L., and J. K. West, *Principles of Electronic Ceramics*, John Wiley & Sons, New York, 1990.

Ingersol Engineers, *The FMS Report*, IFS Publications, 1982.

Karp, R. M., "Reducibility Among Combinatorial Problems," *Complexity of Computer Computations*, Plenum Press, 85–103, 1972.

Kingery, W. D., H. K. Bowen, and D. R. Uhlmann, *Introduction to Ceramics*, John Wiley & Sons, New York, 1976.

Morgan, T. K., "Planning for the Introduction of FMS," *Proc. 2d Int. Conf. on FMS*, 349–357, 1983.

Novak, A., "The Concept of Artificial Intelligence in Unmanned Production—Application of Adaptive Control," *Proc. 2d Int. Conf. on FMS*, 669–680, 1983.

Perkins, W. A., "A Model-Based Vision System for Industrial Parts," *IEEE Trans. Computers*, C-27, 126–143, February 1978.

Proc. 1st Int. Conf. on Flexible Manufacturing Systems, Brighton, 20–22 October 1982. IFS Publications and North-Holland Publishing Company, 1982.

Proc. 2d Int. Conf. on Flexible Manufacturing Systems, London, 26–28 October 1983, IFS Publications and North-Holland Publishing Company, 1983.

Proc. 4th Int. Conf. on Assembly Automation, Tokyo, 11–13 October 1983.

Ranky, P. G., *The Design and Operation of FMS*, IFS Publications and North-Holland Publishing Company, 1983.

Roe, J., "Touch Trigger Probes on Machine Tools," *Proc. 2d Int. Conf. on FMS*, 411–424, 1983.

Rosenfeld, A., *Picture Processing Computer*, Academic Press, New York, 1969.

Suzuki, T., et al., "Present Status of the Japanese National Project: Flexible Manufacturing System Complex," *Proc. 2d Int. Conf. on FMS*, 19–30, 1983.

Tsuji, S., and F. Matsumoto, "Detection of Ellipse by a Modified Hough Transformation," *IEEE Trans. Computers*, C-27(8): 777–781, August 1978.

Tsuji, S., and Nakamura, "Recognition of an Object in Stack of Industrial Parts," *Proc. IJCAI-75*, Tbilisi, Georgia, 811–818, August 1975.

Ullman, J. R., *Pattern Recognition Techniques*, Butterworths, London, 1972.

Woodwark, J. R., and D. Graham, "Automated Assembly and Inspection of Versatile Fixtures," *Proc. 2d Int. Conf. on FMS*, 425–430, 1983.

Zucker, S. W., and R. A. Hummel, "Toward a Low-Level Description of Dot Cluster: Labeling Edge, Interior, and Noise Points," *Computer Graphics and Image Proc.* 9(3), 213–233, March 1979.

CHAPTER 9
Communications

9.1 Introduction

The signals supplied by the sensors enable the processing subsystems to construct a real-time data of the process at a specific point in time. In this context, reference is often made to a *process control diagram*. The process control diagram is necessary to logically connect individual signals, to delay them, or to store them. The process of data processing and control can be used in the broadest sense for the processing of a conventional control system, even a simple one. The task of the processing and control function device is to connect the data supplied by the sensors as input signals, in accordance with the task description, and pass data as output signals on to the actuators. The signals generated are mainly binary signals.

9.2 Single-Board Computer

A single-board computer system is illustrated in Fig. 9.1, with input and output units suitable for the solution of control problems. The central processing unit is a microprocessor that coordinates the control of all sequences within the system in accordance with the program stored in the program memory.

A *read-only memory* (ROM) does not lose its contents even if the line voltage is removed. In the system in Fig. 9.1, the ROM contains a program that allows data to be entered via the keyboard and displayed on the screen. This program is a kind of mini-operating system; it is called a *monitor program* or simply a *monitor*.

A RAM loses its contents when the voltage is removed. It is provided to receive programs that are newly developed and intended to solve a specific control task.

The input/output devices enable the user to enter data, particularly programs, through the keyboard into the RAM and to monitor the inputs.

Binary signals generated by the sensing process can be read in through the input/output devices, generating an actuating process as

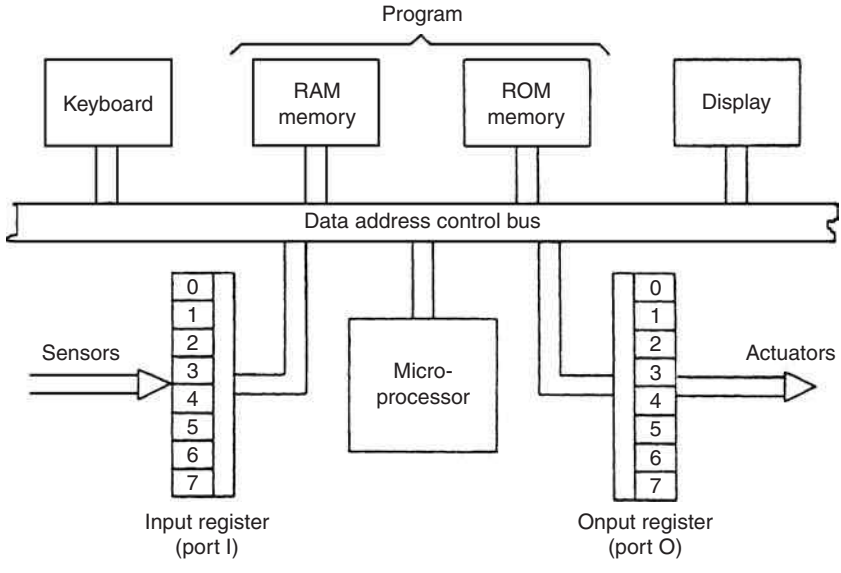


Figure 9.1 Single-board computer.

output response. Registers of this type, which are especially suitable for signal input and output, are also known as *ports*.

The system input/output units are connected by a bus system, through which they can communicate with one another. *Bus* is the designation for a system of lines to which several system components are connected in parallel. However, only two connected devices are able to communicate with one another at any time (Fig. 9.1).

9.3 Sensors for Input Control

A microcomputer is used for the dual-level input control of a reservoir (Fig. 9.2). The level of the liquid is kept constant within the specified limits. For this purpose, two sensors are placed at the desired maximum and minimum locations:

Maximum sensing level = maximum limit

Minimum sensing level = minimum limit

Maximum and minimum sensors are installed to recognize the upper and lower limits. The motor is switched ON when the minimum level has been reached:

Max = 0

Min = 0

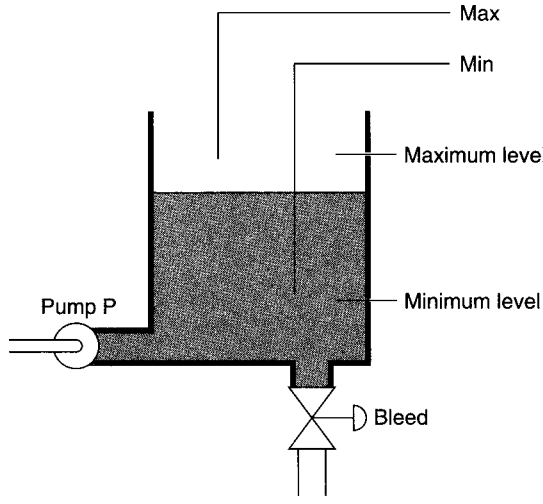


FIGURE 9.2 Dual-level control.

and OFF when the maximum level has been reached:

$$\text{Max} = 1$$

$$\text{Min} = 1$$

If the level is between the two levels

$$\text{Max} = 0$$

$$\text{Min} = 1$$

the motor maintains the current status.

The constant level is disturbed by varying the bleed-off valve. These requirements are shown in a *fact table*, where *P* represents a current status and PN represents the required pump status (Table 9.1).

Max	Min	*P*	Port	PN
0	0	0	I	1
0	0	1	I	1
0	1	0	I	0
0	1	1	I	1
1	1	0	I	0
1	1	1	I	0

TABLE 9.1 Fact Table

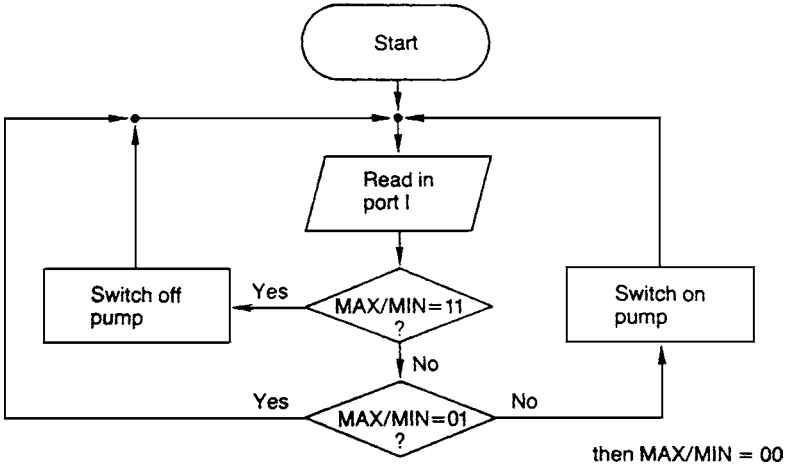


Figure 9.3 Program flowchart.

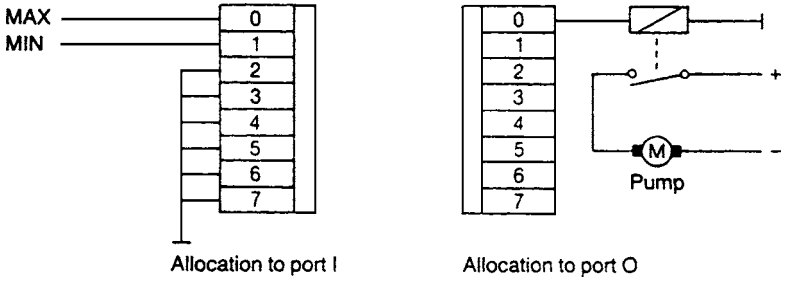


Figure 9.4 Ports to connect sensors and actuators.

The program flowchart in Fig. 9.3 can be used as a programming model. The ports to connect the sensors and actuators are specified in Fig. 9.4.

The programming model is converted into computer program statements. For this, it is necessary for the statement commands to be formulated in machine code. In turn, this makes it necessary to understand the type of microprocessor being used. The well-known Z80 type is used in the example in Table 9.2. The program appears in the abbreviated hexadecimal form.

The hexadecimal counting system is based on the number 16. Like the decimal system, the hexadecimal system uses the numerics 0 to 9 and, in addition, the letters A to F, which represent the numbers 10 to 15. Addresses and commands are entered on a keyboard that is comprised of the keys for the hexadecimal alphanumerics 0 to F and various special keys. An address that has been set and the data it contains can be read from a six-digit numerical display.

Address	Command	Description
0001	DB	Port I (address 10) to the actuator
0002	10	
0003	FE	Compare accu = 0000 0000
0004	00	
0005	CA	If accu = 0000 0000
0006	14	jump to address 14
0007	00	
0008	FE	If accu does not = 0000 0000
0009	01	jump to address 00
000A	C2	If accu does not = 0000 0001
000B	00	jump to address 00
000C	00	
000D	3E	Load 000 000 into accu
000E	00	
000F	D3	Bring accu to port O (address 11)
0010	11	
0011	C3	Jump to address 00
0012	00	
0013	00	
0014	3E	Load 0000 0001 into accu
0015	01	
0016	D3	Bring accu to port O (address 11)
0017	11	
0018	C3	Jump to address 00
0019	00	
001A	00	

TABLE 9.2 Program Statements

Once the program has been developed and tested, there is no further need to use a keyboard and display. Thus, these components can now be disconnected from the system. They do not need to be reconnected until program editing becomes necessary. The editing task and program statement change is usually designated to a qualified administrator or by the "Information Technology" personnel.

It always is a good practice to transfer a completed program from a RAM to a memory whose contents are retained even when the supply

voltage is switched off. An *electrically programmable ROM* (EPROM) could be used. Therefore, the RAM is replaced by an EPROM chip on the PC board.

Sometimes it is necessary to develop extensive programs, thus, comprehensive knowledge and considerable experience are needed. Programs in machine code, as a hexadecimal sequence of alphanumerics, are very complex, difficult to read and document, and hard to test.

9.4 Microcomputer Interactive Development System

A microcomputer interactive development system consists of the following key elements as illustrated in Fig. 9.5:

- Software
- Bulk storage (floppy disk, permanent disk, or other storage)
- Monitor
- Keyboard for interaction and data entry
- Printer

In principle, there are two methods of developing the communication software for an interactive control system:

- The interactive system is located near and directly connected to the installation workstation so that the development may take place directly at the machine under real conditions.
- The interactive system is located away from the installation workstation. In this case, the installation must be totally or partially simulated, otherwise the system must be connected to the installation from time to time.

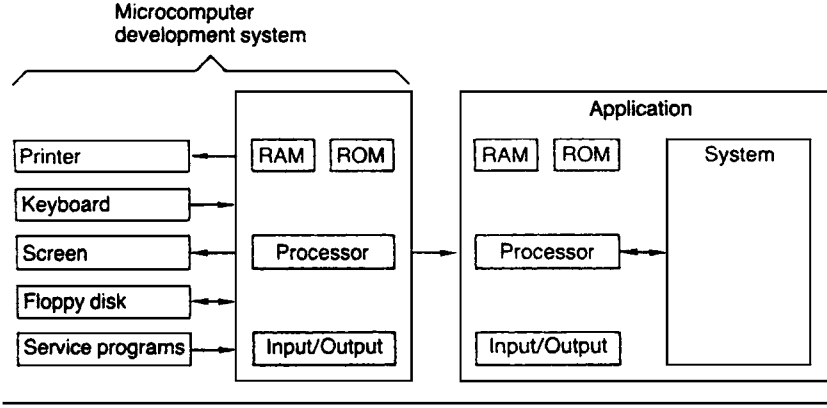


Figure 9.5 Microcomputer interactive development system.

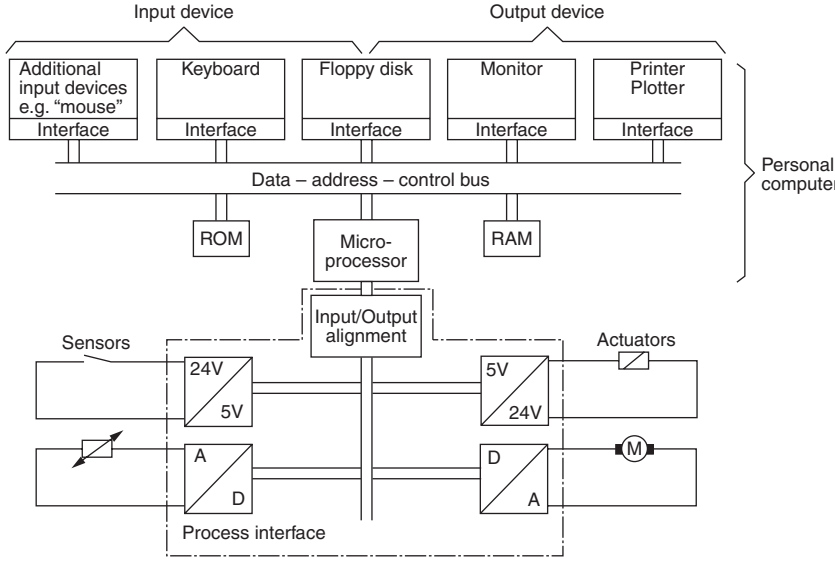


Figure 9.6 Microcomputer standard modules.

When the development of the interactive system has been completed, the system will contain a precise image of the hardware and the software, such as programming, memory size, ports, and number of inputs and outputs. The microcomputer system can then be either built as a printed circuit according to customer specifications or assembled from standard modules (Fig. 9.6).

9.5 The Personal Computer as a Single-Board Computer

The personal computer is not a completely new development in computer evolution. It simply represents a new stage in the development of single-board computers. A single-board computer can play a key role in developing active communication and interfaces for workstations since it can be effectively integrated with process computer peripherals using the following steps:

1. Inclusion of a single-board computer and power supply into a housing
2. Replacement of the simple switch or hexadecimal input by a keyboard similar to that of a typewriter
3. Replacement of the simple LED or seven-segment display by a screen with 24 lines and 40 or 80 characters per line

4. Addition of external mass storage for programs and data (cassette recorder or floppy disk)
5. Connection to a printer or other peripheral device
6. Integration of an interpreter for high-level programming language

A microcomputer system with this configuration is often referred to as a *personal computer*, which is misleading terminology. Although the microprocessor and single-board computer were initially developed for control purposes, the nomenclature directed interest toward purely commercial applications.

The personal computer (or single-board computer) is similar to any peripheral device that requires interfacing to communicate with a central control unit to perform control functions. The personal computer's input and output signals must also be adapted and made available to peripheral devices. Control technology employs a variety of signal forms and levels (e.g., with both analog and binary signals and voltages that may be as high as 24 V). The microcomputer operates with internal voltage levels of 0 and 5 V. Thus, interfacing must provide voltage-level conversion and D/A and A/D conversion. Figure 9.6 is a block diagram of a personal computer and its interfacing to process computer peripherals.

9.5.1 Role of Sensors in Programmable Logic Controllers—PLC

The *programmable logic controller* (PLC) is used in processes where mainly binary signals are processed. The PLC first came into the market at the beginning of the 1970s as a replacement for relay circuits, pioneered by Allen Bradley Company, currently Rockwell International Corporation, USA. Since then, these systems have been continually developed, supported by the progress in microelectronics. For that reason, the present systems are much more powerful than those originally named PLC. Therefore, the terms PLC, process computer, microcomputer, and so on, often overlaps one another. Technically, these systems are often indistinguishable. In addition, there are often many similarities with regard to their functions. Figure 9.7 shows the functional relations between the components of a PLC.

The input signals supplied by the sensors are passed onto the central control unit via an input module. The signals generated by the central control unit are prepared by the output modules and passed on to the actuators. The program is drawn up by using an external programming device and transferred into the program memory (input of the program varies, however, according to the PLC).

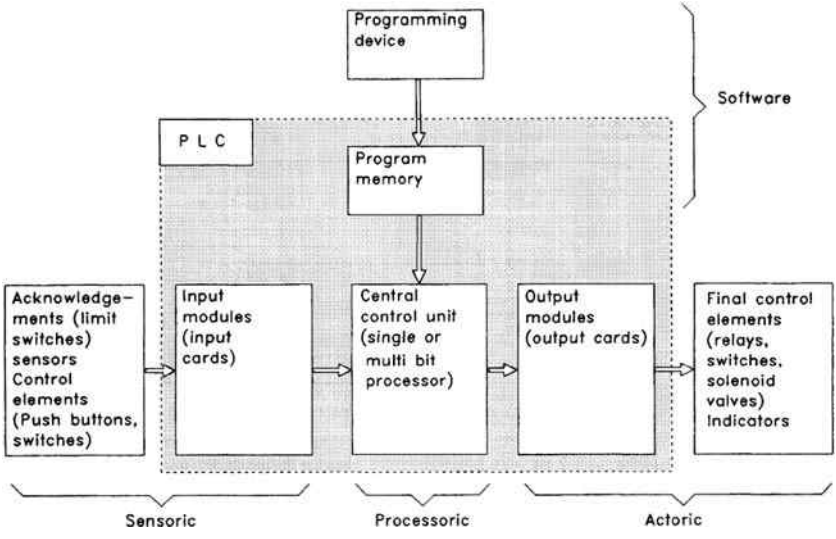


FIGURE 9.7 Relationships between sensor, processor, and actuator activities in the PLC.

The modules are designed to meet the following requirements on the input side:

- Protection against loss of the signal as a result of overvoltage or incorrect voltage
- Elimination of momentary interface impulses
- Possibility of connecting passive sensors (switches, push buttons) and initiators
- Recognition of error status

It depends on the manufacturer whether all requirements are met or only some of them. Figure 9.8 shows the basic elements of an input module.

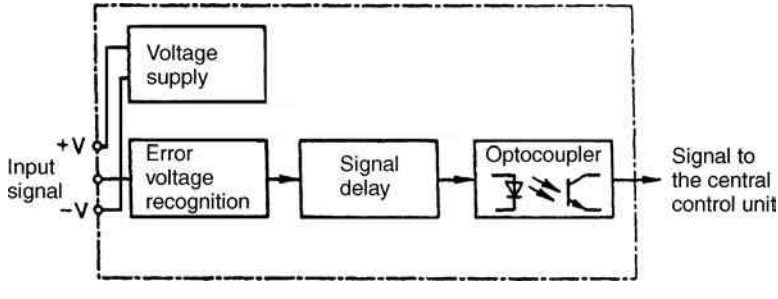


FIGURE 9.8 Elements of an input module.

The module contains an error voltage recognition facility, which is activated when the input voltage exceeds specified tolerance limits. A trigger circuit with signal delay (added in sequence) ensures that momentary interference peaks are suppressed. Positive and negative supply voltages (+V, –V) are provided at the input terminals for connection of various sensors. Optocouplers separate internal and external circuits so that interference is not able to penetrate the PLC via conductive lines. The input module contains visual indicators (mainly LEDs), which indicate whether a 1 or a 0 is present at the input.

The requirements at the output side are:

- Amplification of the output signals
- Protection against short circuiting

Figure 9.9 shows the basic elements of an output module. The optocoupler is used to separate the internal from the external circuit to prevent interference via conductive connections. The output signals need to be amplified so that final control elements and actuators that require additional current can be directly connected to the PLC. The three methods of safeguarding the outputs against short circuiting are:

- Protection by a safety fuse
- Conditional short-circuit-proof outputs (momentary protection)
- Sustained short-circuit-proof outputs

The signal status of the outputs can be read from indicators.

9.5.2 Central Control Units

The first PLCs consisted of a simple single-bit processor. Nowadays, a modern PLC may have at its disposal one or several multibit processors.

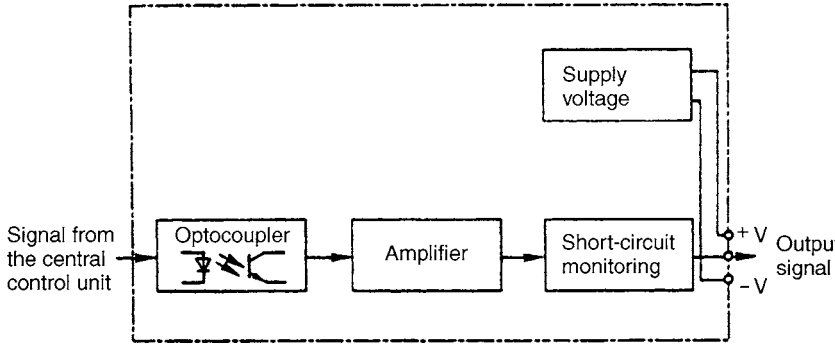


Figure 9.9 Elements of an output module.

Thus, in addition to logic comparison operations, a PLC may execute arithmetic, counting, and time functions. By using powerful microprocessors, it is evident PLCs provide capabilities previously found only in process computers. An example is multitasking, which makes it possible to process a number of unrelated programs simultaneously in the PLC. Obviously, the processor itself is still only able to process one command after another. However, because of high processing speeds and complex operating software, it is possible to give the user the illusion of a number of processors working simultaneously.

Almost all PLCs offer important functions like flags, counters, and timers. *Flags* are single-bit memories in which signal status and program status can be stored temporarily if they are not required until later in the program cycle. Most control tasks require counting and timing functions. *Counters and timers* help the user in programming such tasks.

With most PLCs, it is possible to simultaneously process a number of bits—for example, 8 bits (rather than just 1 bit, as with a switch signal). This is known as *word* processing. As a result, it is possible to connect an A/D converter to the PLC and thus evaluate analog signals supplied by the process. PLCs possess convenient interfaces, through which they communicate with overriding systems (e.g., control computers) or link up with systems of equal status (e.g., other PLCs).

PLC systems are available for small, medium, and large control tasks. The required number of inputs and outputs (i.e., the sum of all sensors and actuators) and the necessary program memory capacity generally determine whether a control task is to be designated as large or small. To facilitate matching PLC systems to requirements, they are generally offered as modular systems. The user is then in a position to decide the size of a system to suit the requirements.

Input and output modules are obtainable as compact external modules or as plug-in electronic cards, generally with 8 or 16 inputs/outputs. The number of such modules to be coupled depends on the size of the control task.

A Representative PLC would be composed of plug-in electronic cards in a standard format. The basic housing provides space for system modules and a power supply. Eight card locations might be provided for input/output modules, each containing 16 input/outputs, giving an I/O-connection capacity of 128. If this number were insufficient, a maximum of four housings could be united to form a system, increasing the number of input/outputs to a maximum of 512. With this system, a program memory capacity of 16 kb can be installed; i.e., up to 16,000 commands can be programmed. Thus, this controller would be the type used for medium to large control tasks.

9.5.3 Process Computer

The process computer is constructed from components similar to those used in a microcomputer system, with the difference that these components are considerably more powerful because of their processing speed and memory capacity. The supporting operating systems are also more powerful, meaning that many control tasks can be processed in parallel. For this reason, the process computer is used in places where it is necessary to solve extensive and complicated automatic control tasks. In such cases, it has the responsibility of controlling a complete process (such as a power station), which is made of a large number of subprocesses. Many smaller control systems, which operate peripherally, are often subordinated to a process computer. In such cases, the process computer assumes the role of a coordinating master computer.

9.6 The NC Controller

Numerical control systems are the most widely used control in machine tools. A blank or preworked workpiece is machined by NC machine tools in accordance with specifications given in a technical drawing and other processing data. Two-, three-, or multiple-axis machines may be used, depending on the shape of the workpiece and the processing technology. A two-axis machine tool might be an NC lathe, for example. A milling machine is an example of a three-axis machine tool.

In NC machines, sensors and actuators are connected to the processor by the manufacturer. The operator or programmer needs to know the operational principle and the interplay between the controller and machine.

Figure 9.10 illustrates the location of holes to be drilled in a workpiece. The skilled machinist needs to know the material from which the workpiece has been made so that suitable drills and feed values can be selected for machining. A drilling machine is equipped with a work table which is adjustable in two directions. The adjustment is

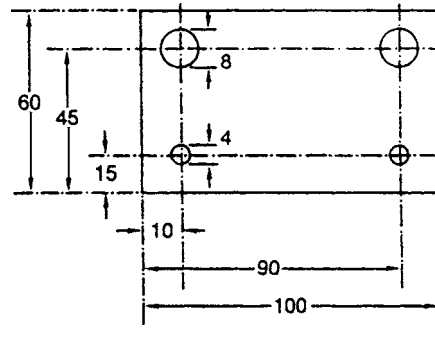


Figure 9.10 Location of holes to be drilled in a workpiece.

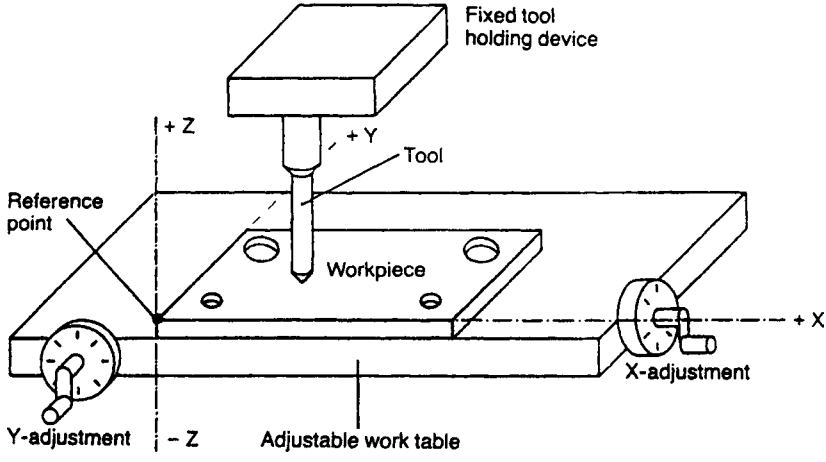


Figure 9.11 Drilling machine equipped with a work table adjustable in two directions.

carried out by using a lead screw, and the set value set can be read from a scale. The machine is illustrated in Fig. 9.11.

9.6.1 Manufacturing Procedure and Control

The manufacturing procedure for a machine tool application is as follows:

1. The table is moved up to the reference point. The reference point values are imparted to the computer. The value zero is read from both scales. The drill is located above the reference point.
2. The workpiece is placed on the work table in the precise position required and clamped.
3. The workpiece zero point is determined; a reference edge from which all dimensions are measured is also determined. The drill is positioned at this point. The workpiece zero point can be imagined as the origin of a rectangular system of coordinates: the y axis is in one direction of motion of the table, the x axis is in the other. Similarly, the dimensions in the drawing are referred to this point. Thus, it is easy to move the work table to each point in turn where holes need to be drilled. Modifications can be carried out if required.
4. Machining is commenced; the working sequence should be described in detail.

This description can completely replace the technical drawing. Any operator must be able to carry out the working cycle using this exact description.

9.6.2 The Machining Program

The machining program is described in terms of sequences:

1. Clamp drill 1.
2. Set *X* to 10 and *Y* to 15.
3. Drill at cutting speed 2.
4. Set *X* to 95.
5. Drill at cutting speed 2.
6. Unclamp drill 1, clamp drill 2.
7. Set *Y* to 45.
8. Drill at cutting speed 1.
9. Set *X* to 10.
10. Drill at cutting speed 1.
11. Remove workpiece.

The positioning data, cutting tool information, and the required cutting speed must be identified to accomplish any machining operation. Usually, the technological data such as speed and feed are included in the tool data sheet. Tool number, speed, feed, and direction of rotation together form a unit. The feed speed has been omitted in the preceding program; this decision has been left to the operator. If the process is to proceed automatically at a later date, this information will have to be added.

The positioning data is always referenced from the origin of the coordinates. In this case, this is always the point of intersection of the reference lines for the dimensioning of the drawing. The positional information was therefore given in absolute form and can be programmed in absolute dimensions.

The information can also be provided relative to the current position, which is entered as the *x* value, and a new value added for the next position. This is called *incremental positioning data*. The incremental positioning data may be expressed as follows:

1. Set *X* + 10, *Y* + 15.
2. Set *X* + 80.
3. Set *Y* + 30.
4. Set *X* –80.

Choice of the most suitable method often depends on the dimensioning of an engineering drawing. Sometimes incremental and absolute specifications of positional data are mixed. Then, obviously, it is necessary to describe the type of data in question.

In the example just described, the machining of the workpiece takes place only at four different points; only one machining position is approached. No rule exists as to which path should be taken to reach this position. Thus, no changes have been made at the workpiece when the drilling operation occurs, and the table is moved by a zigzag course to the position for the second drilling operation. This is referred to as a *point-to-point* control.

The situation is different if, as shown in Fig. 9.12, a groove is to be milled in a workpiece. The tool to be used is a milling cutter. The worktable might be the same as that used for the drill. The workpiece is to be contacted on the path from point 1 to point 2, and a zigzag course is, of course, impermissible. However, the problem can be easily solved since the machining parts are always parallel to one of the two axes. The tool is set to start at position P1. From this point on, only the y value is set to Y2. When the tool arrives there, the x value is set to the new value X2. This is a simple straight-line control.

Figure 9.13 represents a diagonal groove to be milled from point 1 to point 2. A simple solution can be realized using the zigzag course, but it must be calculated with precision, as illustrated in Fig. 9.14.

Without making tedious calculations, it is clear that the machine arrives at point 2 when the tool is moved by an increment of Δx, then

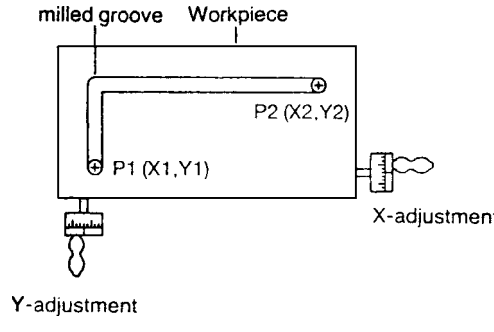


Figure 9.12 Groove to be drilled in workpiece.

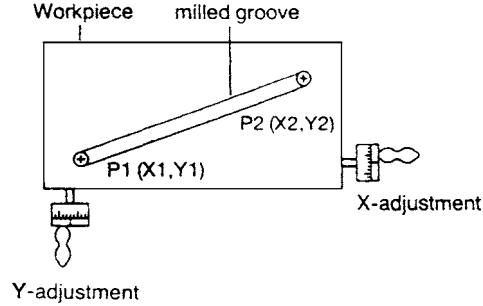


Figure 9.13 Diagonal groove to be milled from point 1 to point 2.

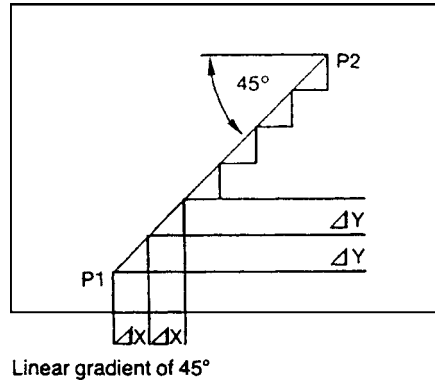


Figure 9.14
Zigzag course to mill diagonal groove.

by the same amount Δy, and then once more by Δx, and so on. If the increments Δx and Δy are small, the zigzag line is no longer noticeable and the deviation from the ideal course remains within the tolerance limits. If the required path is not under 45°, the values for Δx and Δy differ from each other, according to the linear gradient. These values must then be calculated. In an automatic control system, the computer solves this problem. The straight-line control in this example is referred to as an *extended straight-line control*.

A random course from point 1 to point 2 is illustrated in Fig. 9.15. The increments for the x direction and the y direction (Δx, Δy) must be formed. However, calculations of this path may be complicated if the curve to be traveled cannot be described by simple geometric equations. Accordingly, a contour control method is used in this manufacturing application. Contour controls can generally travel in arcs, even elliptical arcs. Complicated shapes must be transformed into a straight-line path by coordinating points. A programming workstation is necessary for this application.

Figure 9.16 illustrates in diagrammatic form the method of automating a drilling machine. The scales are replaced by incremental or absolute displacement measuring systems. Suitable actuators are used in place of the handwheels. Sensors are precisely placed at the extreme

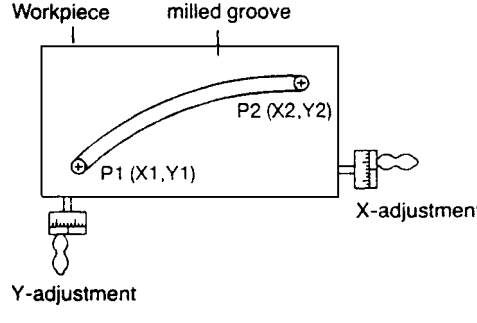


Figure 9.15
Random course from point 1 to point 2.

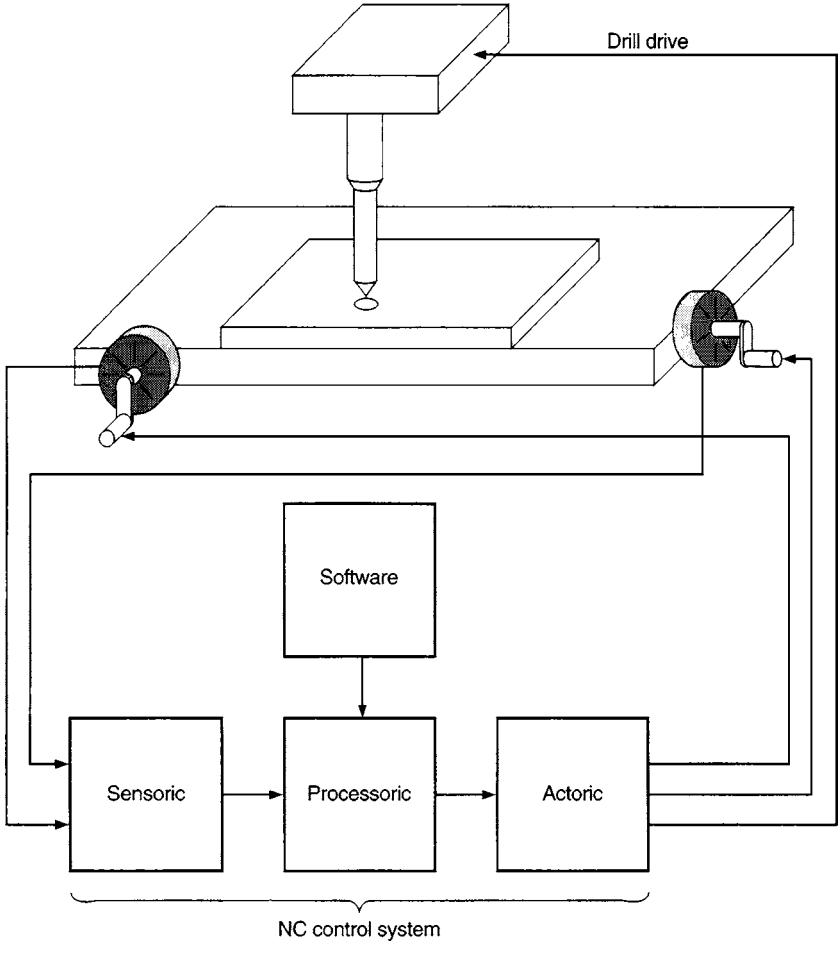


Figure 9.16 Method of automating a drilling machine.

ends of the traveled distance. Today, DC motors, which provide infinite and simple adjustments for all axes and for the main drive, are widely used.

Two different methods are used to describe the course of a tool or machine slide: incrementally and absolutely. Accordingly, the subcontrol task, movement of a machine slide, is solved differently for each.

The two types of incremental control systems are illustrated in Fig. 9.17. In an incremental control system with an incremental displacement encoder, the processor counts the pulses supplied by the linear displacement sensor. A pulse corresponds to the smallest measurable unit of displacement. The distance to be traveled and the direction of travel are passed on to the processor as reference values. From these data, the processor forms a control signal for the final

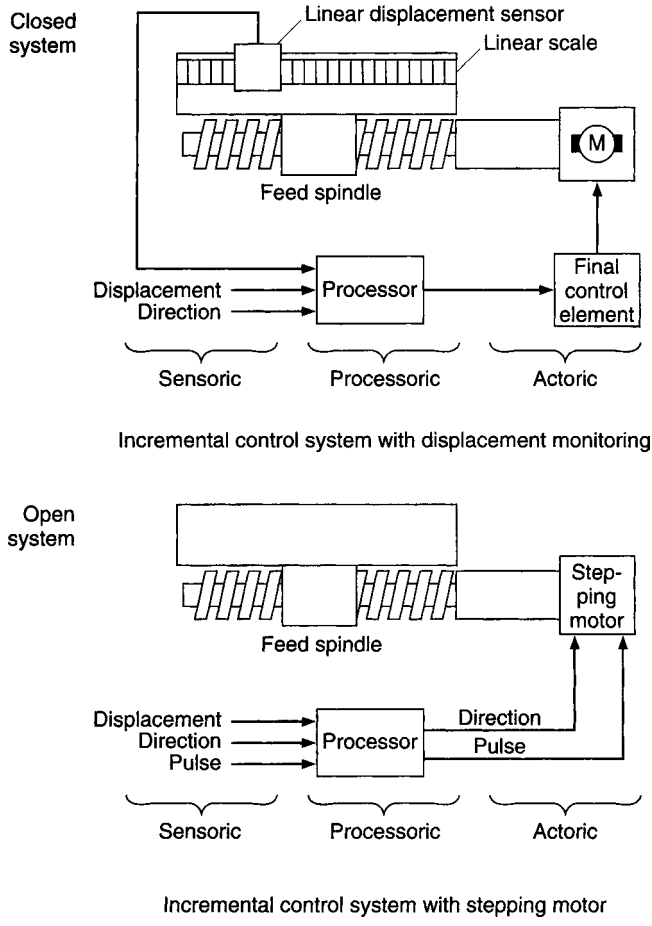


Figure 9.17 Two types of incremental control system.

control element of the drive motor. If the values do not agree, the motor must continue running. Once they agree, the processor transmits a stop signal to the final control element, indicating that the correct slide position has been reached. When a new reference value is supplied, this process is started up once again. It is important that the slide stop quickly enough. Braking time must be taken into consideration—in other words, before the end position is reached, the speed must be reduced as the remaining gap decreases.

In incremental control with a stepping motor as actuator, there is no acknowledgment of the slide position. In place of the impulse from the linear displacement sensor, a pulse is given straight to the processor. The processor now passes on pulses to the stepping motor until a pulse count adds up to the displacement reference value. Because the stepping motor covers an accurately defined angle of

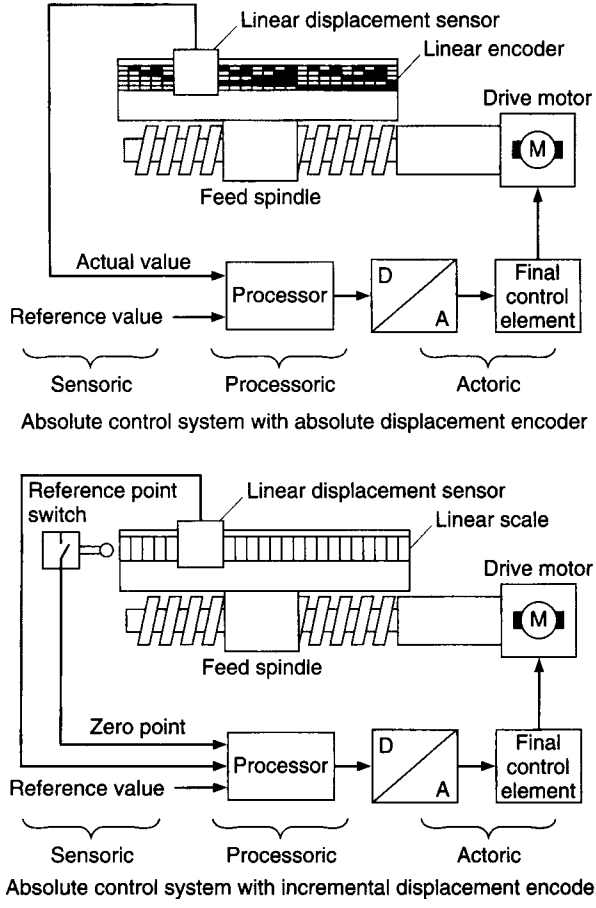


Figure 9.18 Two types of absolute control system, with absolute displacement encoder and incremental displacement encoder.

rotation for each step, each position can be easily approached by the slide. With stepping motors, only very limited speeds are possible, depending on the resolution.

Figure 9.18 illustrates the two types of absolute control system: one with an absolute displacement encoder and the other with an incremental displacement encoder.

9.6.3 Absolute Control

In an absolute control system with an absolute displacement encoder, such as an angular encoder or linear encoder, the actual value of absolute position of the slide is known at any given time. The processor receives this value and a reference value. The processor forms the arithmetic difference between the two values. The greater the difference,

the greater the deviation of the slide from its reference position. The difference is available as a digital value at the output of the processor. After D/A conversion, the difference is presented as an analog value to a final control element such as a transistor that triggers the drive motor. The greater the deviation from the reference value, the greater the motor speed and the adjustment speed of the slide. When there is no further deviation, the motor stops. The motor, of course, has both upper and lower speed limits. At lower speeds, the torque may become so low that the tool may falter.

In an absolute control system with an incremental displacement encoder, a sensor acts as a limit switch at the end position of the slide. Once the machine is started, the slide is moved to the end position, where an incremental displacement counter in the processor is set to the reference point value. Each additional movement of the slide is recorded by this counter so that, at any given time, it knows the absolute position of the counter, like the absolute system with a displacement encoder. The absolute control system with an incremental displacement encoder is the most widely used today.

9.6.4 NC Software

As with a PLC system, it is necessary to draw up an accurate model for an NC system before a program is written.

The machine must be told which movements it is to execute. For this purpose, displacement data must be prepared that can be taken from a drawing of the workpiece. In addition, the machine needs to know what tools to use and when, the tool feed speed, and the cutting speed. Certain additional functions such as the addition of lubricants must also be programmed.

All these data are clearly of great importance for conventional machining of workpieces. The data are compiled during the work preparation in a factory to create a work plan for a skilled operator. A conventional work plan is an excellent programming model. However, work plans for manual operation cannot, as a rule, be converted to NC operations.

To ensure there are no ambiguities in the description of displacement data, there is general agreement about the use of Cartesian coordinates for NC (Fig. 9.19). Different layers of the coordinate system apply to different processing machines, as shown in Fig. 9.20.

The commands in an NC instruction set are made up of code letters and figures. The program structure is generally such that sentences containing a number of commands can be constructed. Each sentence contains a sentence number. Table 9.3 shows an NC program for an automatic drilling machine.

Displacement conditions are given the code letter G. The displacement condition G00 is programmed in sentence 1, signifying point-to-point control behavior. This condition thus also applies to all

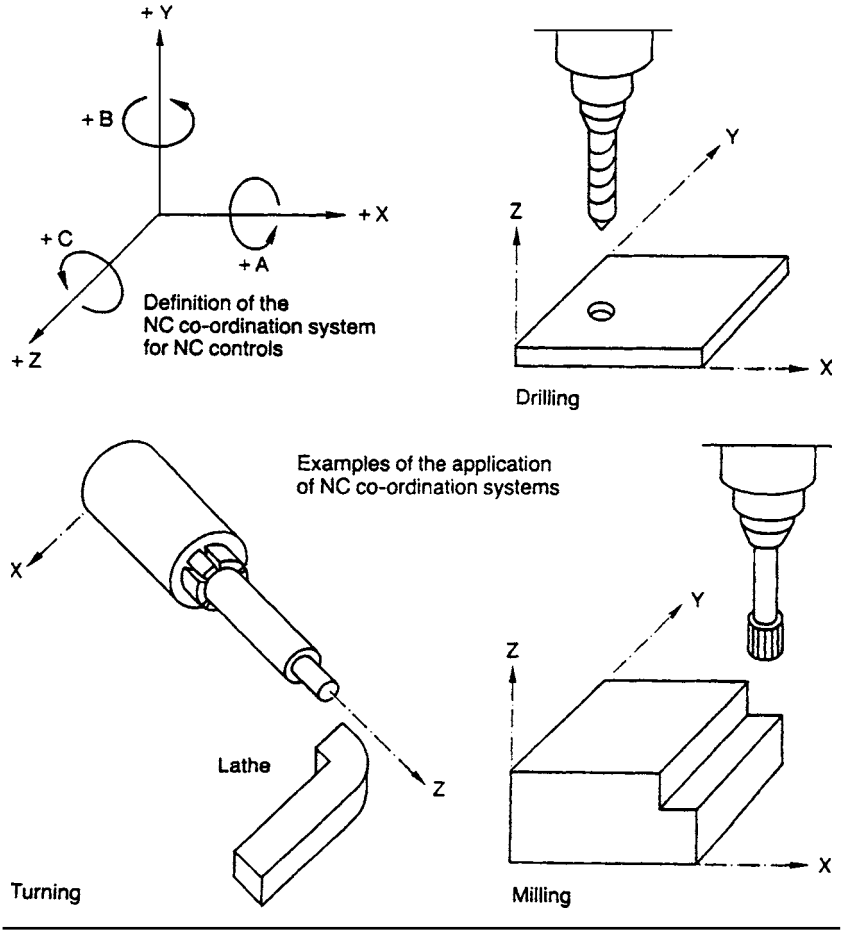


Figure 9.19 Cartesian coordinate system for NC and a coordinate system applied to different processing machines.

other sentences; it does not have to be repeated. The feed is specified with code letter F, followed by a code number or address where the actual feed speed is located. The cutting speed is given with a code letter S and a code number. The code for the tool is T; the following code number 01 in the program indicates that a drill should be taken from magazine 1. The additional function M09 indicates that a lubricant is to be used. M30 defines the program M.

An NC program for a lathe is described in Table 9.4. A part to be turned is illustrated in Fig. 9.20. It is assumed that the workpiece has already been rough-machined. Therefore, it is only necessary to write a finished program.

Path control behavior is programmed in the beginning sentences with displacement condition G01. The lathe is set in absolute values to

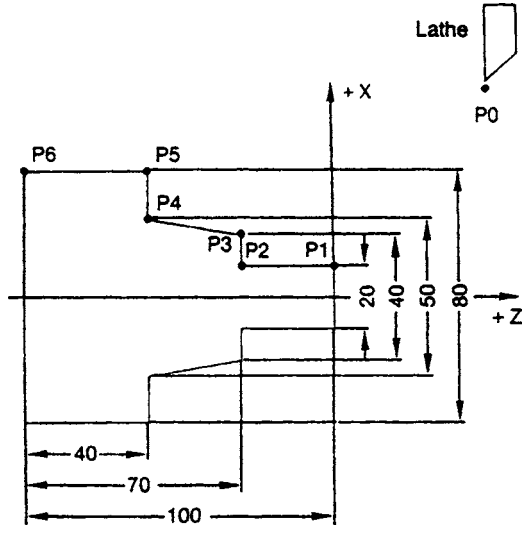


Figure 9.20 Parts to be turned.

Sentence Number	Conditions	Displacement Data			Switching Data			
					Tool			
		x	*y*	*z*	Feed	Speed	Tool	Addition
N001	G00	X10	Y15	Z2	F1000	S55	T01	M09
N002				Z0				
N003		X909		Z2				
N004				Z0				
N005			Y45	Z2	F500	S30	T02	
N006				Z0				
N007		X10		Z2				
N008				Z0				
N009		X0	Y0					M30

Table 9.3 NC Program—Automatic Drilling Machine

coordinates X20 and Z100. The displacement condition G91 of the second sentence determines relative addressing (i.e., the x and z values should be added to the previous values). Where the x values are concerned, it should be noted that diameter rather than radius is specified.

The NC programs shown in these examples must, of course, also be put on a suitable data carrier in machine-readable form. The classic data carrier for NC is punched tape. Each character in the program text is converted into a 7-bit combination of 1s and 0s and stamped

Sentence Number	Conditions	Displacement Data			Switching Data			
					Tool			
		x	*y*	*z*	Feed	Speed	Tool	Addition
N001	G01	X20		Z100	F1500	S60	T03	M09
N002	G91			Z–30				
N003		X + 20						
N004		X + 10		Z–30				
N005		X + 30						
N006				Z–40				M30

Table 9.4 NC Program—Lathe Machining Operation

into the tape as an appropriate combination of holes and spaces (1 = hole, 0 = no hole).

This process of conversion is called *coding*. The totality of all combinations of holes representing a character set is called a *code*.

Recommendations have been laid down as to which combination of holes represents a specific character. Figure 9.21 shows the N005 sentence statement in Table 9.3 as a punched tape code to ISO standards. Each 7-bit character is accompanied by a test bit. The test bit is a logic 1 (hole punched) when the number of holes is uneven. Thus, the total number of holes in a character is always even. When the punched tape code is checked to ensure it is even-numbered, errors caused by damage or dirt may be detected. This method of checking for errors is called a parity check.

Because punched tape is robust in aggressive industrial environments, it is still preferred by many users. However, modern storage

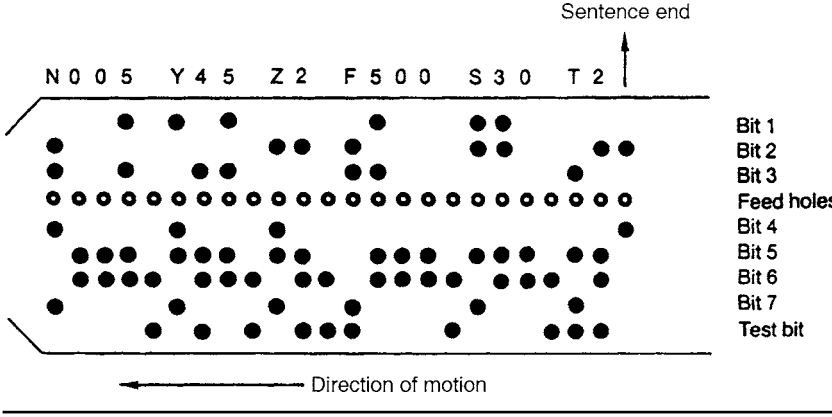


Figure 9.21 N005 sentence statement from Table 9.3 (drilling task) as a punched tape code to ISO standards.

media such as floppy disks are used more and more. It is also possible to use an economical personal computer as a development and documentation station for NC programs.

9.6.5 Operation of an NC System

Figure 9.22 is a block diagram of a simple NC system. The NC machine is equipped with a punched tape reader into which the program for machining a specific piece is inserted. After starting the program, the processor reads the first sentence. The various displacement and switching data are written into a reference-value memory. Next, the commands are processed. The comparator compares the reference values to the values supplied by the measuring devices (e.g., the current positions of the x- and y-drive units).

If there is a difference between reference values and actual values, the comparator generates suitable control signals for the final control elements of the drives. These are generated only until reference values and actual values agree. The comparator now demands a new sentence from the punched tape reader for processing. This sentence is evaluated in the manner described. Reading and processing are repeated until the final sentence has been processed. As an example of reference value/actual value comparison, Fig. 9.23 shows the adjustment of the x axis of an NC machine.

If the machine slide is moved from its current position representing the actual value to a new position representing the reference value, a continuous comparison of the reference value to the actual

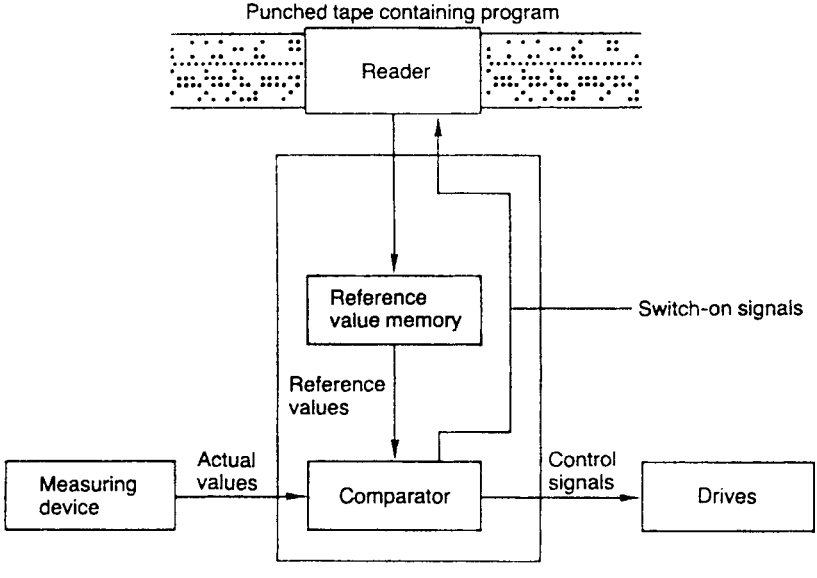


Figure 9.22 Block diagram for a simple NC system.

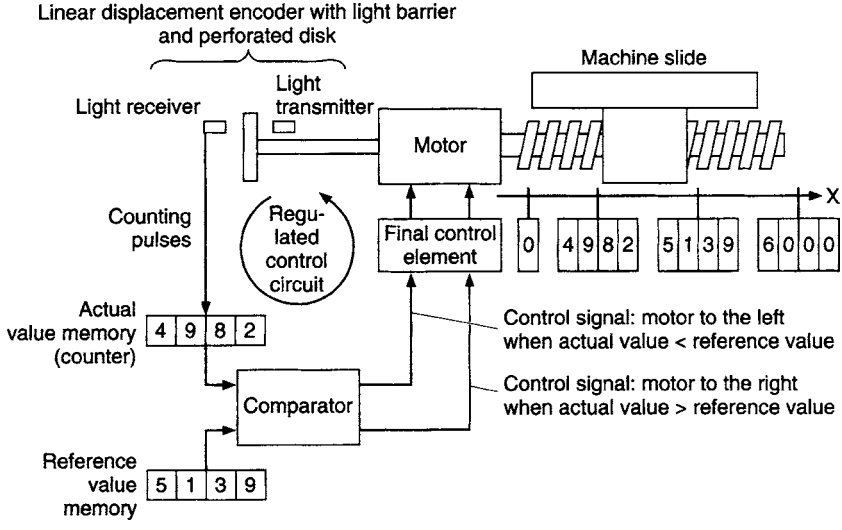


Figure 9.23 Adjustment of the *x* axis of an NC machine.

value is carried out. The comparator determines the difference between the values and supplies a control signal to the final control element of the motor in accordance with the arithmetical sign.

The motor runs in a counterclockwise direction when the actual value is smaller than the reference value. This corresponds to a displacement of the machine slide to the right. The motor runs in a clockwise direction when the actual value is greater than the reference value, which corresponds to a displacement of the machine slide to the left.

The number of rotations of the motor, which is proportional to the displacement of the slide, is counted for the actual-value memory by a measuring device which, in the example, is a light transmitter, a perforated disk on the motor shaft, and a light receiver.

The movement of the slide by the motor drive reduces the difference between the reference value and actual value. When the difference is 0, the comparator sets the control signal to 0, and the motor is stopped.

The type of programming in this case is machine-dependent; it corresponds to an assembler language. Application-oriented programming, with which a more general formulation can be written, is also developed for NC systems. An example is the NC programming language *Advanced Programming Tool* (APT), which can cope with several hundred commands. APT commands are classified into three categories:

- Geometric commands
- Motion commands
- Auxiliary commands

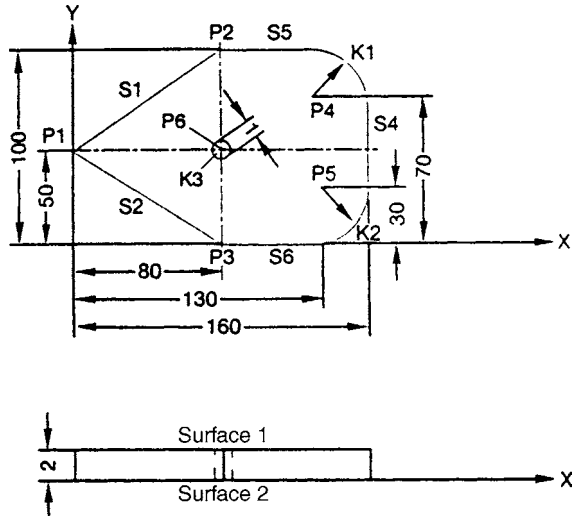


Figure 9.24 NC coordinate system for a workpiece.

9.6.5.1 Geometric Commands

In the NC coordinate system in Fig. 9.24, each point is clearly determined through specification of coordinates x, y, and z. In the APT, each point can be defined symbolically—for example, the definition for the point $x = 15.7$, $y = 25$, and $z = 13$ is

P1 = POINT/15.7,25,13

Thus, the variable to the left of the equal sign can be freely selected. Distances from one point to another are defined as follows:

S5 = LINE/P2,P3

P2 and P3 must be defined in advance. Points may also be defined as the intersection of two distances, as in

P1 = POINT/INTOF L6,L8

A circle can be defined in the following ways:

CIRCLE1 = CIRCLE/CENTER,POINT,RADIUS,5.8

CIRCLE2 = CIRCLE/P4,P5,P6

In the first instance, the circle is defined by specifying the center and the radius, and in the second, by specifying three points on the perimeter. A plane surface is defined by three points located on it—for example:

SURFACE1 = PLANE/P2,P2,P3

Another way of specifying a plane is to indicate a point on a surface and a second surface lying parallel to the first one—for example:

SURFACE2 = PLANE/P1,PARALLEL,SURFACE1

The geometric dimensions of a workpiece given in the technical drawing can be fully described using these geometric commands. The workpiece shown in Fig. 9.24 is to be described using the APT geometric commands:

P1 = POINT/0.0, 50.0, 0.0
P2 = POINT/80.0, 100.0, 0.0
P3 = POINT/80.0, 0.0, 0.0
P4 = POINT/130.0, 70.0, 0.0
P5 = POINT/130.0, 70.0, 0.0
P6 = POINT/80.0, 50.0, 0.0
S1 = LINE/P1, P2
S2 = LINE/P1, P3
K1 = CIRCLE/CENTER, P4, RADIUS, 30.0
K2 = CIRCLE/CENTER, P5, RADIUS, 30.0
K3 = CIRCLE/CENTER, P6, RADIUS, 0.5
S5 = LINE/P2, LEFT, TANTO, K1
S6 = LINE/P3, RIGHT, TANTO, K2
S4 = LINE/LEFT, TANTO, K2, RIGHT, TANTO, K1
SURFACE1 = PLANE/0.0, 0.0, –2

The commands for defining paths S4, S5, and S6 form tangents to circles K1 and K2 and are for this reason given the additional functions LEFT, RIGHT, and TANTO. The line S5 starts at P2 and forms a left (LEFT) tangent to (TANTO) circle K1.

9.6.5.2 Motion Commands

These commands are necessary to describe the path the tools must travel. A distinction is made between point-to-point (PTP) controls and path controls. The commands for point-to-point controls are as follows:

GOTO Absolute statement of the point to be approached
GODTLA Relative statement of the point to the approached

Example: To travel from point 2,5,10 to point 8,10,20, the two commands would have to be used as follows:

Either GOTO/8,10,20
Or GODTLA/6,5,10

The drilling operation in this example is programmed as follows:

```
GOTO/P6
GODTLA/0.0, 0.0, –2.5
GODTLA/0.0, 0.0, 2.5
```

A path control is programmed with the following commands:

GOFWD	(go forward)
GOLFT	(go left)
GODWN	(go down)
GOBACK	(go back)
GORGT	(go right)
GOUP	(go up)

With these commands, it is possible, for example, to program the movement of a cutting tool in accordance with the contour in the example.

GOTO/S1,TO,SURFACE1,TO,S2	Go to start point P1
GOFWD/S1,PAST,S5	Go along line S1 to S5
GOFWD/S5,TANTO,K1	Go along line S5 to K1
GOFWD/K1,TANTO,K2	Go from K1 via S4 to K2
GOFWD/K2,PAST,S6	Go from K2 via S4 to S6
GOFWD/S2,PAST,S1	Go along line S2 to S1

9.6.5.3 Auxiliary Commands

Various switching functions are programmed using commands from this group—for example:

MACHIN	Specifies the tool
FEDRAT	Determines the feed speed
COOLNT	Connects coolant
CUTTER	Diameter of the tool
PARTNO	Part number of the workpiece
FINI	Program end

Other application-oriented programming languages for NC programming are EXAPT, ADAPT, and AUTOSPOT. These languages are constructed like APT, and in part build on that language. An NC program written in application-oriented language must, of course, also be translated into machine code and produced in punched tape (Fig. 9.25).

9.6.6 Computer Numerical Control System

The program for a specific machining task of an NC machine is located on a punched tape. This punched tape must be read separately for

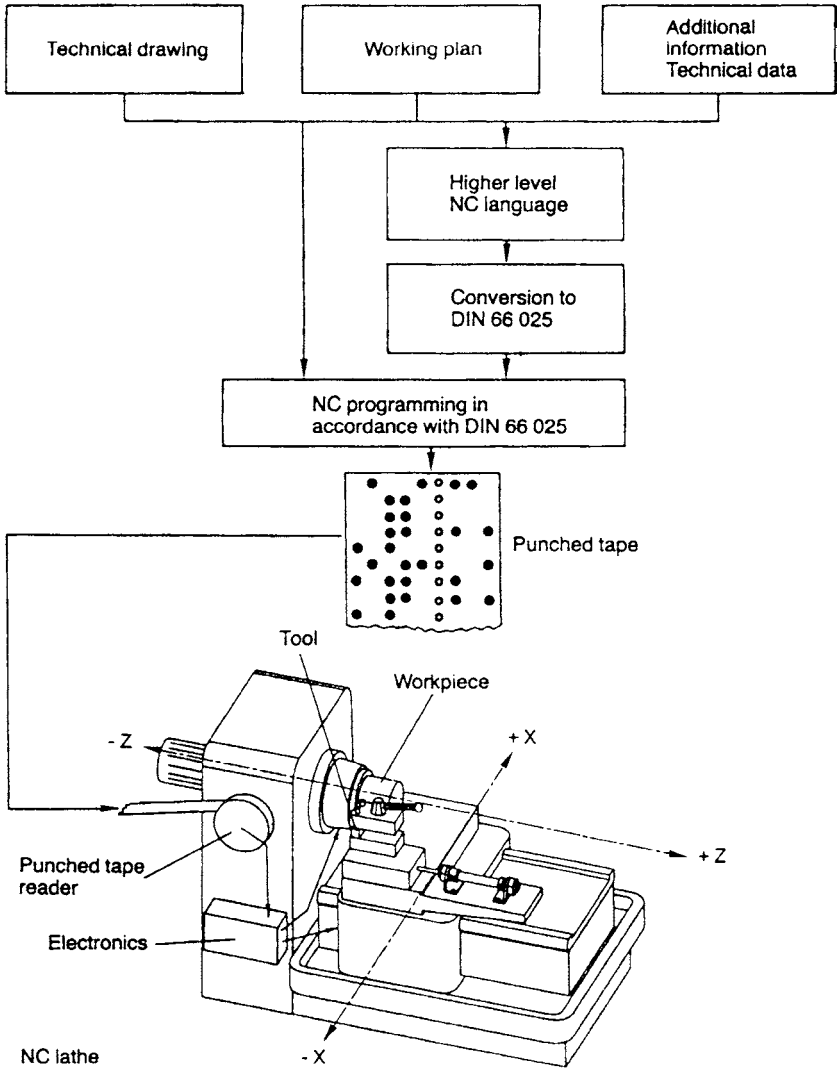


Figure 9.25 NC program translated into machine code and produced in punched tape.

every single workpiece to be machined. Reading of punched tape is a mechanical process that subjects the tape to wear, which can lead to errors.

With the development of computer numerical control, it became possible to overcome these disadvantages and, in addition, to create many advantages. A CNC system contains its own computer system based on modern microcomputer technology. A semiconductor memory is used as program memory. The punched tape is in fact also

available with the system. However, it is only required when the program is transferred into the semiconductor memory. In addition to this program entry technique, CNC systems offer the possibility of transferring programs directly from external computers via appropriate interfaces.

Moreover, most CNC control systems are equipped with a keyboard and a screen that make it possible to program directly on the machine. This is referred to as *workshop programming*. Keyboard and screen also offer the possibility of editing and improving programs on site. This is a decisive advantage when compared to NC systems, where modifications and adaptations are much more time-consuming.

With microcomputers, it is easy to program the frequently complicated arithmetical operations required for machining complicated curves. Cutting and feed speed can be matched to the best advantage. The computer can check and adapt the values continually and thus determine the most favorable setting values.

Integrated test and diagnostic systems, which are also made possible by microcomputers, guarantee a high degree of availability for CNC systems. The control system monitors itself and reports specific errors independently to the operator.

The programming languages correspond in part to those discussed for NC systems but are, in fact, frequently manufacturer-specific. The most important components of a CNC system are illustrated in Fig. 9.26.

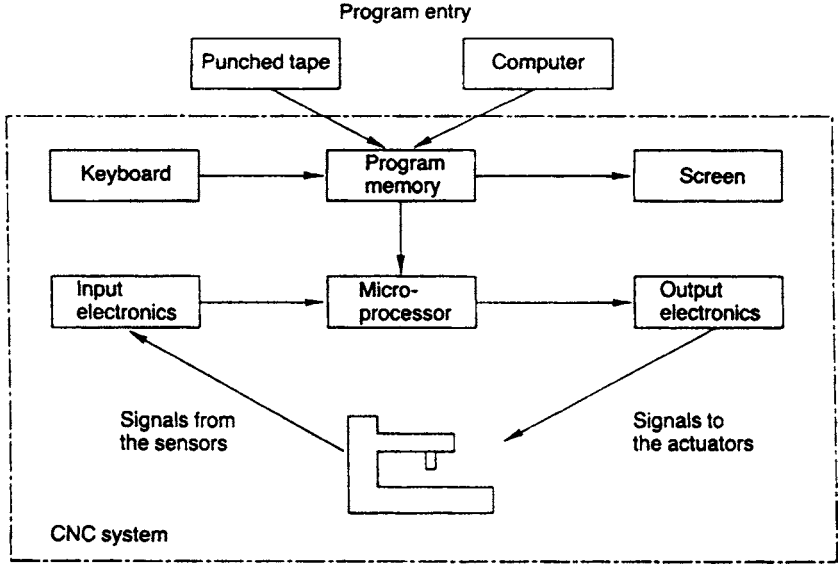


Figure 9.26 CNC system components.

9.7 Industrial Handling

The modern term for handling technology is *industrial handling* or simply *handling*. This concept embraces simple and complex handling and clamping devices.

The best and most versatile handling component is the human hand. Within its physical limitations of weight and size, many possibilities are available: gripping, allocating, arranging, feeding, positioning, clamping, working, drawing, transferring, and so on.

Unlike the human hand, a mechanical handling device is able to perform only a very limited number of functions. For this reason, several handling devices have been developed for automatic handling and clamping functions.

It is not necessary to make a full copy of the human hand. A mechanical device is more or less limited to the shape, size, construction material, and characteristics of a specific workpiece to be handled.

Fig. 9.27 shows a device made up of four pneumatic handling units that can insert parts into a lathe for turning.

Cylinder A separates the parts to be turned and assumes the function of allocating parts in the machine-ready position. The part to be turned is brought to the axial center of the machine. At this position, the part can be taken over by the gripping cylinder D. Cylinders B and C are used to insert the workpiece into the clamping chuck and remove it again after machining.

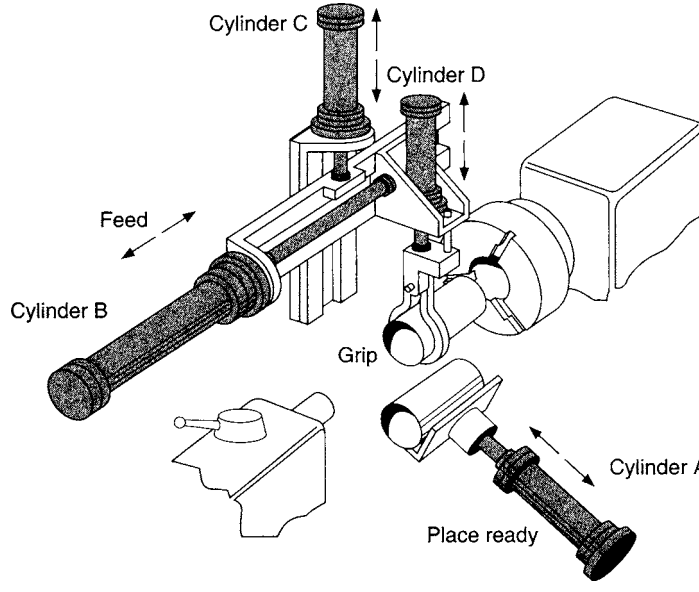


Figure 9.27 Device for inserting parts into a lathe for turning.

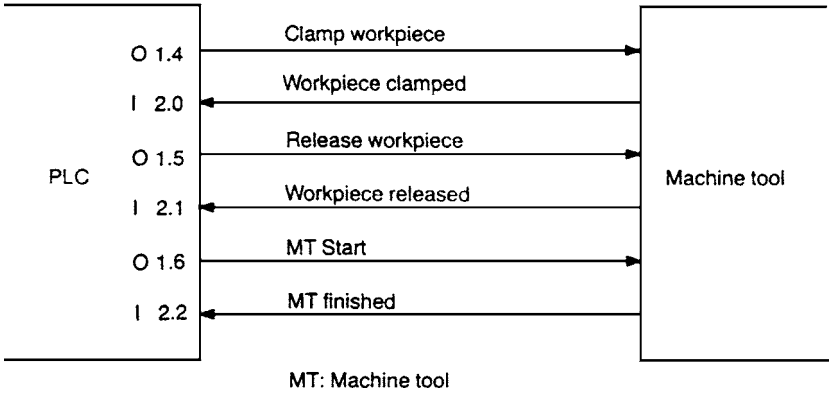


Figure 9.28 Interface signals between PLC and the machine tools.

The fully turned part may be either dropped or picked up by a handling mechanism that is located opposite cylinder A and directly passed on after being oriented.

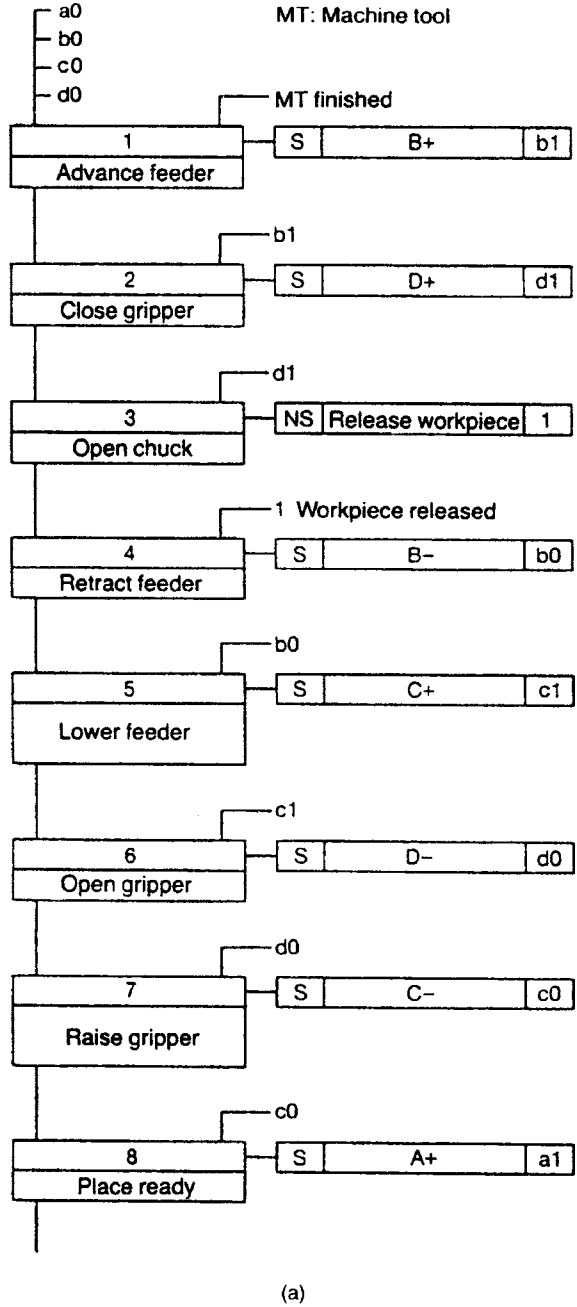
A PLC is suitable as a control system for the feed unit. A program must be developed for this purpose.

First, the subtasks (1) "pass on workpiece to the machine tool" and (2) "take workpiece from the machine tool" are to be analyzed. In order to be able to solve these tasks, signal exchange between the PLC and the machine tool (MT) is necessary. Figure 9.28 shows the necessary interface signals.

The subtasks are:

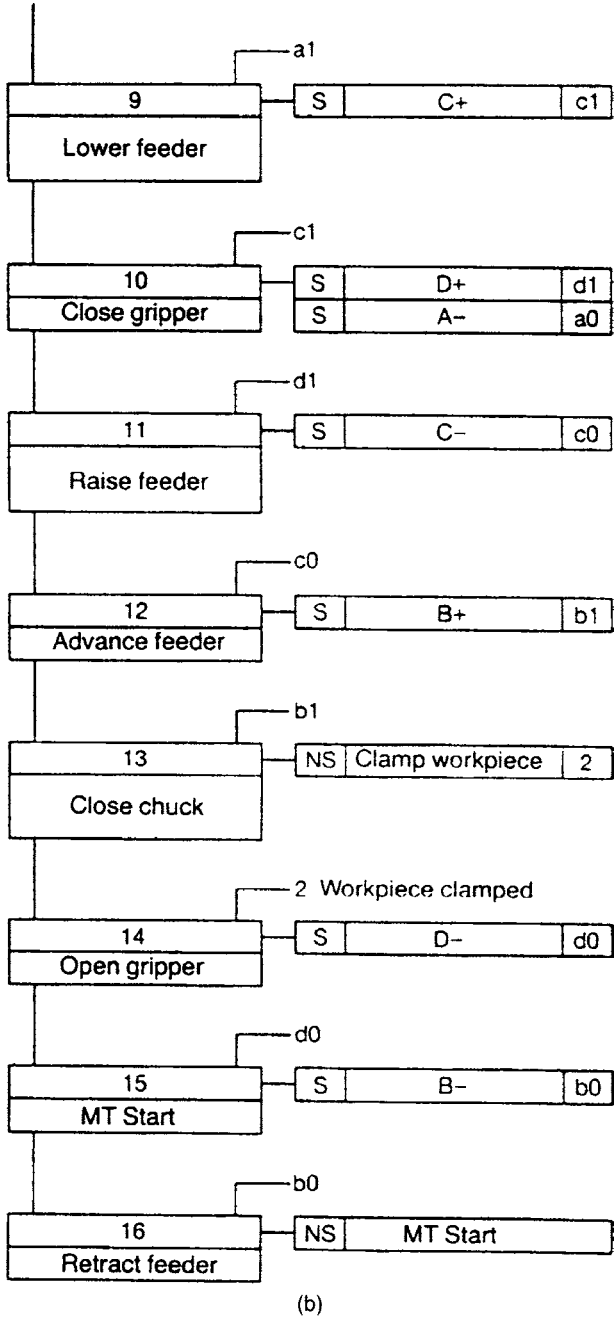
1. The MT informs the PLC that a workpiece has been completed and can be removed. Signal: MT complete.
2. The gripper travels to the MT and grips the workpiece. The PLC requires the MT to release the workpiece. Signal: release workpiece.
3. The MT releases the workpiece and replies with the signal: workpiece released.
4. The gripper can now advance from the MT with the workpiece.
5. The gripper travels with the workpiece into the MT and requires it to clamp. Signal: clamp workpiece.
6. The MT clamps the workpiece and replies with the signal: workpiece clamped.
7. The gripper opens and retracts from the MT.
8. Once the gripper has left the collision area, the PLC reports that the MT can begin machining. Signal: MT start.

The function chart in Fig. 9.29 illustrates the specified cycle in detail form. Using this chart is necessary to develop a PLC program.



(a)

Figure 9.29 Function chart.



(b)

Figure 9.29 (*Continued*)

```
          INSTRUCTION              COMMENT
=============================================================
0000 ALLOCATION LISTING MACHINE  V 01
0001                  I 1.0      " LIMIT SWITCH A0
0002                  I 1.1      " LIMIT SWITCH A1
0003                  I 1.2      " B0 HOR. FEED BACK
0006                  I 1.3      " B1 HOR. FEED FORWARD
0007                  I 1.4      " C0 VERT. FEED UP
0008                  I 1.5      " C1 VERT. FEED DOWN
0009                  I 1.6      " D0 GRIPPER OPEN
0010                  I 1.7      " D1 GRIPPER CLOSED
0011
0012                  I 2.0      " MT WORKPIECE CLAMPED
0013                  I 2.1      " MT WORKPIECE RELEASED
0014                  I 2.2      " MT FINISHED
0015
0016                  O 1.0      " CYL.A PREPARE
0017                  O 1.1      " CYL.B HOR. FEED
0018                  O 1.2      " CYL.C VERT. FEED
0019                  O 1.3      " CYL.D GRIP
0020                  O 1.4      " MT CLAMP WORKPIECE
0021                  O 1.5      " MT RELEASE WORKPIECE
0022                  O 1.6      " MT START
```

Figure 9.30 Allocation list.

Inputs must be assigned to sensors. Outputs from the PLC must also be assigned to actuators. Figure 9.30 shows the appropriate allocation lists. The sequencing program for the feed unit is shown in Fig. 9.31.

9.8 Packaging Technology

Arranging and feeding of parts for insertion into boxes, including an intermediate layer, with a feed unit is illustrated in Fig. 9.32. Using this inserting machine, many other functions are carried out either simultaneously with or subsequent to feeding. The various parts that are fed by a conveyer belt are lined up independently and then arranged in rows by actuator A, according to a specific pattern. The parts are transferred by the lifting device in accordance with this pattern.

The horizontal movement of the feed unit is executed by actuator B; the vertical movement, by actuators C and D. In the position shown in the diagram, the parts picked up by the lifting device are inserted into the box by extending actuator D. At the same time, actuator C also extends and picks up a square board that is used as an intermediate layer as two layers of parts are being inserted. As soon as actuator C has retracted, holding the intermediate layer, the device is moved by the indexing actuator B. The end position actuator C located above the box and actuator D located above the collection point are used for the arrangement of parts.

When actuator C has extended, the intermediate layer is inserted and the next layer of parts is taken up by actuator D. When actuator B retracts, the position drawn is reached once again and the second layer is inserted into the box. Then, when actuator B extends once again, an intermediate layer is placed on the second layer and the next parts are picked up.

```
          INSTRUCTION                COMMENT
=============================================================
0000 PROGRAM MACHINE    0.0 V 01
-------------------------------------------------------------
0001 STEP 1
0002 IF               I 1.0          " LIMIT SWITCH A0
0003       AND        I 1.2          " B0 HOR. FEED BACK
0004       AND        I 1.4          " C0 VERT. FEED UP
0005       AND        I 1.6          " D0 GRIPPER OPEN
0006       AND        I 2.2          " MT FINISHED
0007 THEN  SET        O 1.1          " CYL.B HOR. FEED
0008       RESET      O 1.6          " MT START
-------------------------------------------------------------
0009 STEP 2
0010 IF               I 1.3          " B1 HOR. FEED FORWARD
0011 THEN SET         O 1.3          " CYL.D GRIP
-------------------------------------------------------------
0012 STEP 3
0013 IF               I 1.7          " D1 GRIPPER CLOSED
0014 THEN SET         O 1.5          " MT RELEASE WORKPIECE
-------------------------------------------------------------
0015 STEP 4
0016 IF               I 2.1          " MT WORKPIECE RELEASED
0017 THEN RESET       O 1.1          " CYL.B HOR. FEED
0018      RESET       O 1.5          " MT RELEASE WORKPIECE
-------------------------------------------------------------
0019 STEP 5
0020 IF               I 1.2          " B0 HOR. FEED BACK
0021 THEN SET         O 1.2          " CYL.C VERT. FEED
-------------------------------------------------------------
0022 STEP 6
0023 IF               I 1.5          " C1 VERT. FEED DOWN
0024 THEN RESET       O 1.3          " CYL.D GRIP
-------------------------------------------------------------
0025 STEP 7
0026 IF               I 1.6          " D0 GRIPPER OPEN
0027 THEN RESET       O 1.2          " CYL.C VERT. FEED
-------------------------------------------------------------
0028 STEP 8
0029 IF               I 1.4          " C0 VERT. FEED UP
0030 THEN SET         O 1.0          " CYL.A PREPARE
-------------------------------------------------------------
0031 STEP 9
0032 IF               I 1.1          " LIMIT SWITCH A1
0033 THEN SET         O 1.2          " CYL.C VERT. FEED
-------------------------------------------------------------
0034 STEP 10
0035 IF               I 1.5          " C1 VERT. FEED DOWN
0036 THEN SET         O 1.3          " CYL.D GRIP
0037      RESET       O 1.0          " CYL.A PREPARE
-------------------------------------------------------------
0038 STEP 11
0039 IF               I 1.7          " D1 GRIPPER CLOSED
0040 THEN RESET       O 1.2          " CYL.C VERT. FEED
-------------------------------------------------------------
0041 STEP 12
0042 IF               I 1.4          " C0 VERT. FEED UP
0043 THEN SET         O 1.1          " CYL.B HOR. FEED
-------------------------------------------------------------
0044 STEP 13
0045 IF               I 1.3          " B1 HOR. FEED FORWARD
0046 THEN SET         O 1.4          " MT CLAMP WORKPIECE
-------------------------------------------------------------
0047 STEP 14
0048 IF               I 2.0          " MT WORKPIECE CLAMPED
0049 THEN RESET       O 1.3          " CYL.D GRIP
0050      RESET       O 1.4          " MT CLAMP WORKPIECE
-------------------------------------------------------------
0051 STEP 15
0052 IF               I 1.6          " D0 GRIPPER OPEN
0053 THEN RESET       O 1.1          " CYL.B HOR. FEED
-------------------------------------------------------------
0054 STEP 16
0055 IF               I 1.2          " B0 HOR. FEED BACK
0056 THEN SET         O 1.6          " MT START
0057      JUMP TO     S 1
0058      PSE
```

Figure 9.31 Sequencing program for the feed unit.

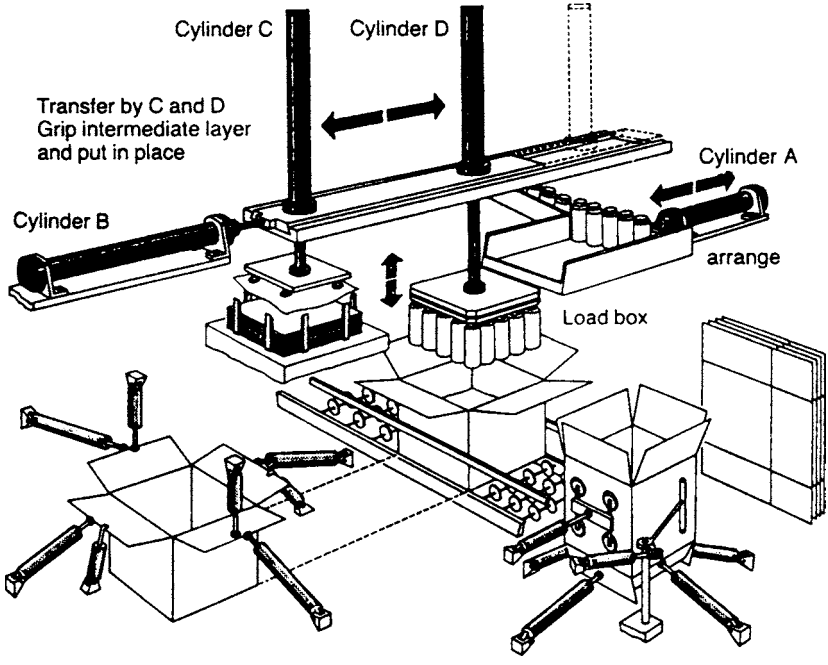


Figure 9.32 Insertion machine.

When actuators C and D retract for the second time, box changeover is triggered. Only when a new empty box is brought into position does a new insertion cycle begin.

The feed functions carried out by this inserting machine include arrangement, separation, and insertion cycles.

9.9 Linear Indexing for Manufacturing Applications

Linear indexing is best-suited in applications where strip or rod-shaped materials have to be processed in individual stages throughout their entire length. An example of linear indexing is shown in Fig. 9.33.

A machine stamps holes into a strip of sheet metal at equal intervals. After each stamping process, the metal strip is clamped, pushed forward a specified distance, and clamped again. Pushing forward and clamping are carried out by a pneumatic feed unit controlled by various sensors. This feed unit is a compact handling component specifically intended for strip-indexing tasks.

Figure 9.34 illustrates indexing conveyer belts, which are also referred to as *linear indexing systems*. Suitable devices for driving such a belt are either motors or pneumatic actuators with a mechanism such as a ratchet to convert an actuator linear stroke into a rotary

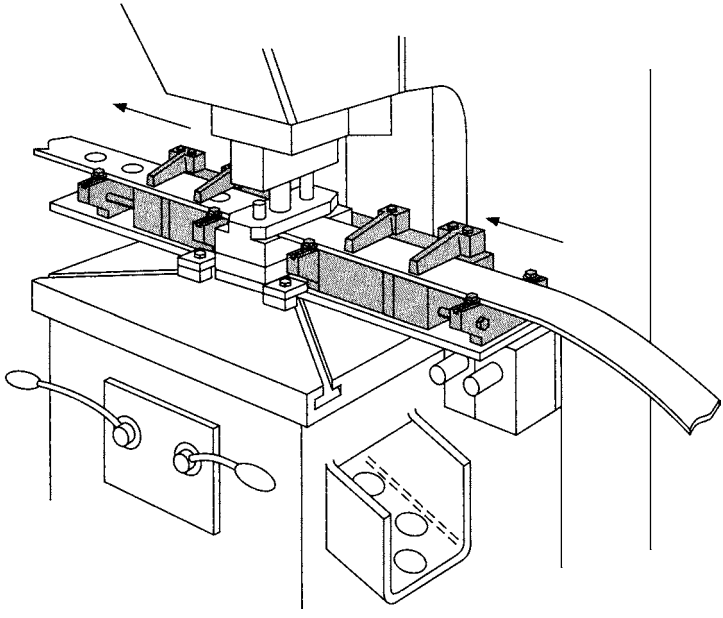


Figure 9.33 Linear indexing.

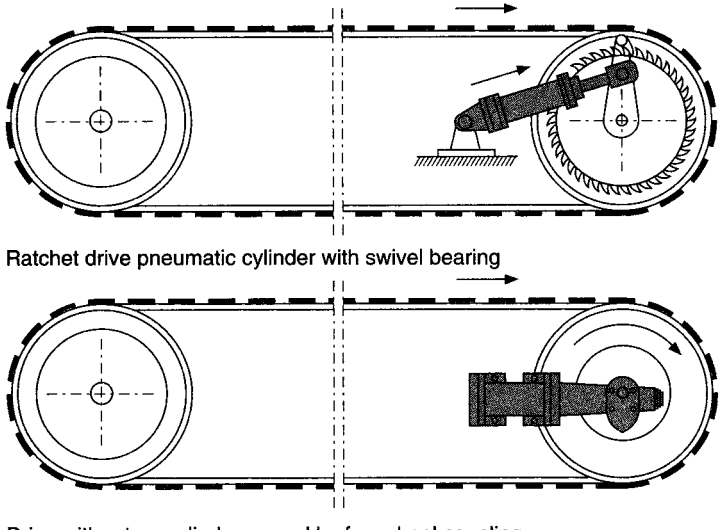


Figure 9.34 Generating indexing motion.

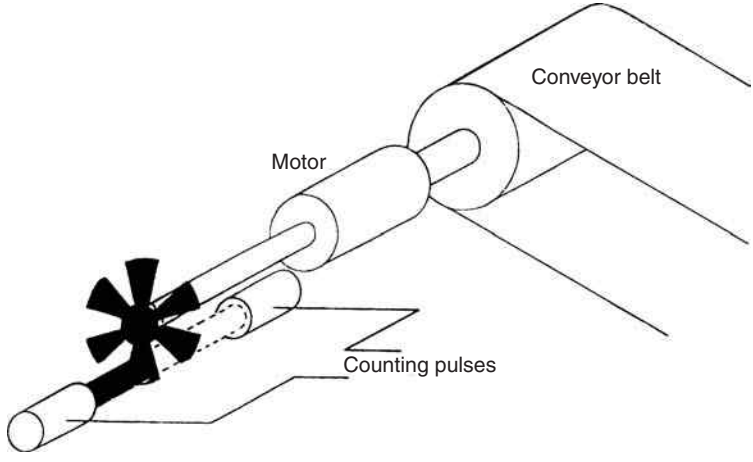


Figure 9.35 Conveyer belt with an electric servomotor as actuator and a sensor for angular displacement detection.

movement. A pneumatic device is utilized in this case as an actuator to drive the belt conveyer.

Figure 9.35 illustrates another application of an indexing driving mechanism of a conveyer belt with an electric servomotor as actuator and a sensor as an angular displacement detection.

9.10 Synchronous Indexing for Manufacturing Applications

Synchronous indexing systems are used when a number of assembly and fabrication processes are required to be carried out on a product. The product components must be fed once and remain on the synchronous indexing table until all processes are complete.

A vertical synchronous rotary indexing system picks up a minimum of two components. The complete product sequence can be carried out in one position without the need for repeated clamping and unclamping procedures. Rearrangement within the cycle becomes unnecessary. Feeding and withdrawing are limited to the last rotary indexing station; the number of manufacturing functions is thus without significance.

As illustrated in Fig. 9.36, each component is individually processed at six stations:

Station 1—insert and remove the workpiece

Station 2—drill the front face

Station 3—drill the top and base

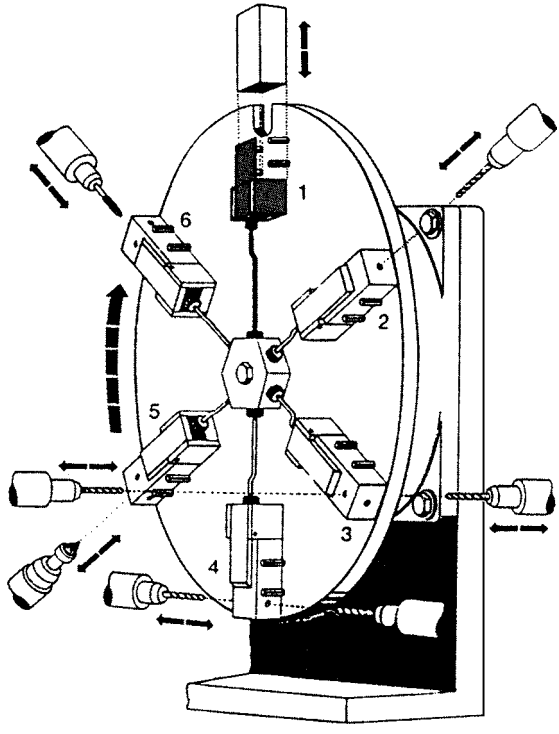


Figure 9.36 A vertical synchronous rotary indexing system.

Station 4—drill both sides at a right angle to the top and base
Station 5—countersink the front face drill hole
Station 6—cut a thread into the drill hole on the front face

A horizontal synchronous rotary indexing table is illustrated in Fig. 9.37. This system consists of eight stations. Stations 2 to 7 are the machining stations. This is just an example of the many diverse manufacturing processes that can take place in a synchronous fashion. Station 1 is the feed station; station 8 is the eject station.

The synchronous system can be driven by an electric motor with a special driver and cam configuration, by stepping motors, or by pneumatic rotary indexing units.

9.11 Parallel Data Transmission

Figure 9.38 illustrates the principle of parallel-type data transmission. Each data bit has its own transmission line. The data (e.g., alphanumeric characters) are transmitted as parallel bits from a transmitter such as the terminal to a receiver like the printer.

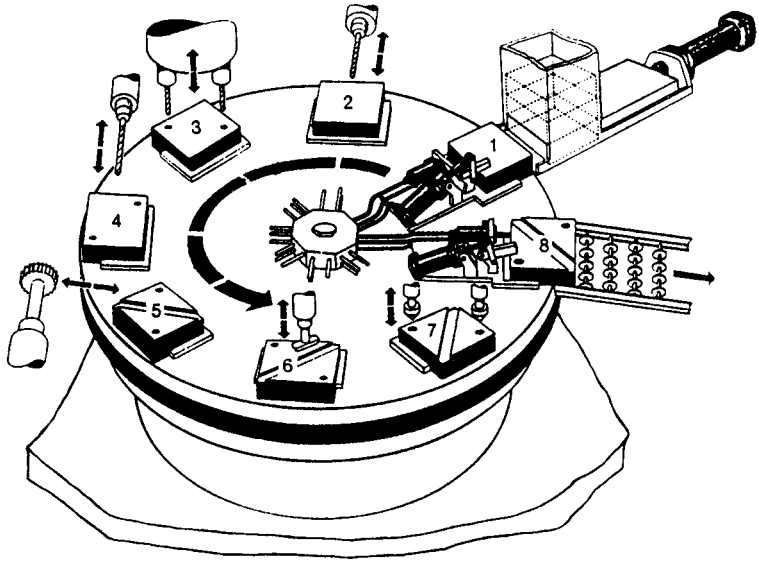


Figure 9.37 Horizontal synchronous rotary indexing table.

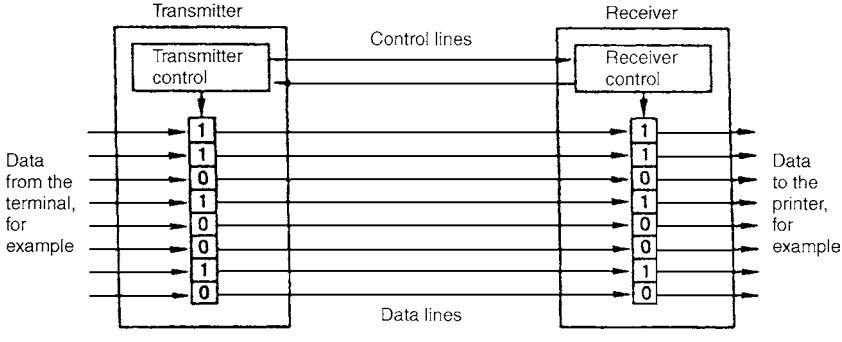


Figure 9.38 Parallel-type data transmission.

Some additional control lines are required to ensure synchronization between the transmitter and the receiver. Via these control lines, the printer informs the terminal whether it is ready to take over a character, and the terminal informs the printer that a character is being sent. The distance between the transmitter and the receiver may amount to a maximum of approximately 8 m, as line capacities may lead to coupling and signal deformation. Transmission speed is hardware-dependent and may theoretically amount to 1 Mbyte/s (i.e., 1 million characters)—the line, however, can only be 1 m in length.

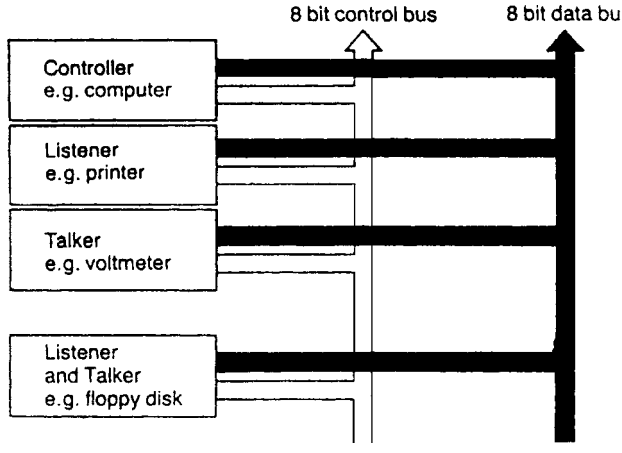


Figure 9.39 Several devices communicating with one another using parallel transmission, with precise coordination of data flow.

If several devices are to communicate with one another according to the principle of parallel transmission, precise coordination of the data flow is necessary (Fig. 9.39).

The devices are divided into four groups:

- Talkers: able only to transmit, such as sensors
- Listeners: able only to receive, such as printers
- Talkers/listeners: can transmit and receive, such as floppy disks
- Controllers: can coordinate data exchange, such as computers

Each device is allocated a specific address. The controller decides between which two devices data exchange is to be carried out. Two standards describe the technique of transmitting interface functions: standards IEC-625 and IEEE-488. The number of data and control lines is specified in the standards as eight each. Up to 30 devices can be connected to the bus; the complete system of lines may not be longer than 20 m. Maximum transmission speed is between 250 and 500 kb/s.

In general, it can be said that relatively fast data transmission, such as several hundred thousand characters per second, can be achieved with parallel interfaces. However, transmission length is limited to several meters.

9.12 Serial Data Transmission

Figure 9.40 illustrates the principle of serial data transmission. Only one line is available for serial data transmission. This means that an

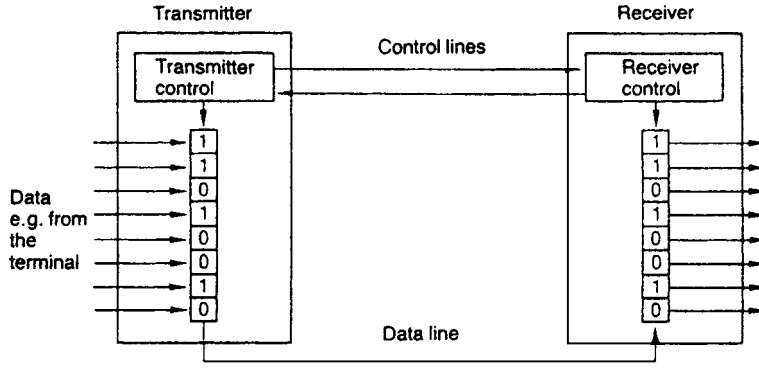


Figure 9.40 The principle of serial transmission.

8-bit piece of data, for example, must be sent bit by bit from the send register via the data line to the receiver register.

A distinction is made between synchronous and asynchronous data transmission. With asynchronous data transmission, data can be sent at any time; pulse recovery is achieved through the start and stop bits.

In the synchronous process a transmitter and receiver have the same time pulse at their disposal. Control lines are necessary for this purpose.

Functional, electrical, and mechanical characteristics of serial data transmission can also be found in several standards. The most widely used standard versions are the V-24 and RS-232 interfaces (these interfaces are almost identical). With standards, it is guaranteed that different manufacturers' devices are able to exchange data. The V-24 is an asynchronous interface.

Synchronization is achieved by ready signals. The transmitter may, for example, pass data on to the receiver only when the receiver has previously signaled its readiness to receive.

The transmission sequence of the data bits determines the data format (Fig. 9.41). Before transmission of the first data bit, a start bit is sent at the logic zero level. Then, 7 or 8 data bits follow, the one with the lowest value coming first. Afterward, a test bit (parity bit) is transmitted, followed by one or two stop bits.

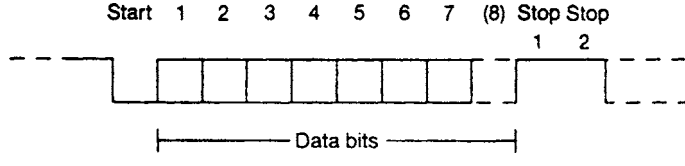


Figure 9.41 The transmission sequence of the data bits determines the data format.

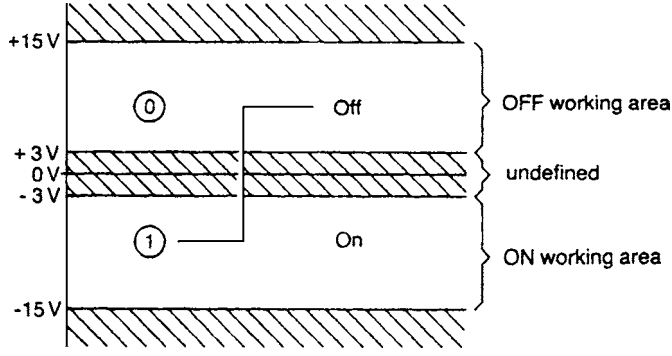


Figure 9.42 Electrical characteristics determine the signal level.

Transmission speeds are freely selectable within a specified framework. Transmitter and receiver must be set to the same speed, or baud rate. Common rates are 75, 110, 135, 150, 300, 600, 1200, 2400, 4800, 9600, and 19,200 bit/s (baud).

The electrical characteristics determine the signal level (Fig. 9.42). If the voltage of a signal on a data line opposite the signal ground is greater than 3 V and is negative, signal status 1 applies, corresponding to the binary character 1. If the voltage of a signal on a data line opposite the signal ground is greater than 3 V and is positive, signal status 0 applies, corresponding to the binary character 0.

Signal status is undefined in the transmission range +3 to –3 V. Connection of the interface line between two devices is effected with a 24-pin D plug. Figure 9.43 shows a V-24 plug with the most important signal lines.

This is one of the oldest serial interfaces. It is used to drive teleprinters [teletypewriters (TTY)]. Transmission takes place along a pair of send and receive lines. Logic 0 is realized via a 20-mA current, and logic 1, via the lack of this current. The functional characteristics of the V-24 interface may also apply to the 20-mA current loop interface.

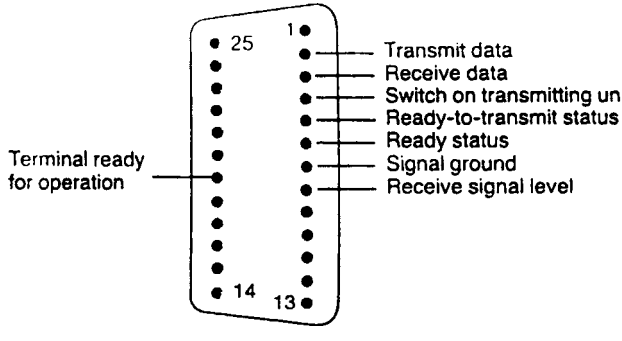


Figure 9.43 V-24 plug.

9.13 The Collection and Generation of Process Signals in Decentralized Manufacturing Systems

Not long ago, it was customary to centrally collect all process signals. The processor and control elements were housed in a central control cabinet (Fig. 9.44). Today, with automation systems, extensive control tasks can be solved effectively if these tasks are distributed among several decentralized self-sufficient systems.

A decentralized structure is illustrated in Fig. 9.45. The sensing, actuating, and processing sections of subsystems are connected to a master computer that evaluates signals applied by each subprocess and sends back control signals.

These subsystems can be realized in different ways—for example, a PLC or single-board computer could be used as the control system for the subprocesses. A process computer, personal computer, or mainframe can be used as a master computer.

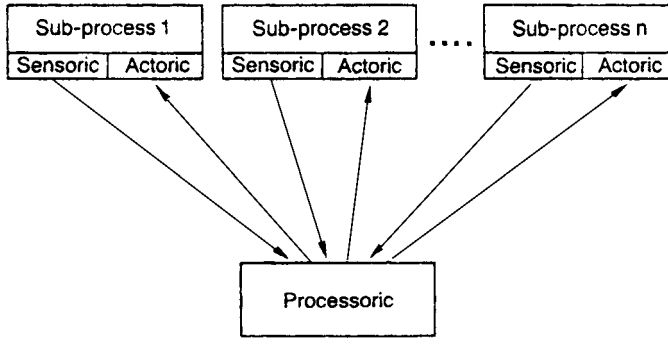


Figure 9.44 A centralized control system.

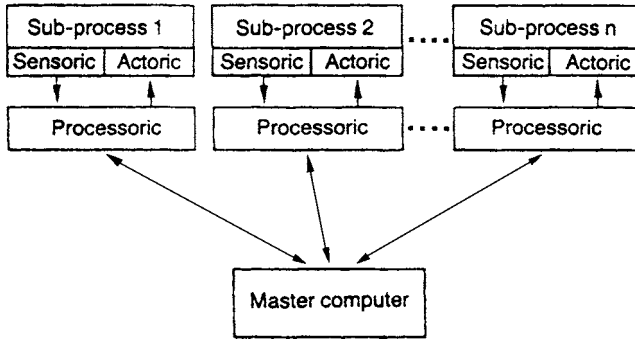


Figure 9.45 A decentralized control system.

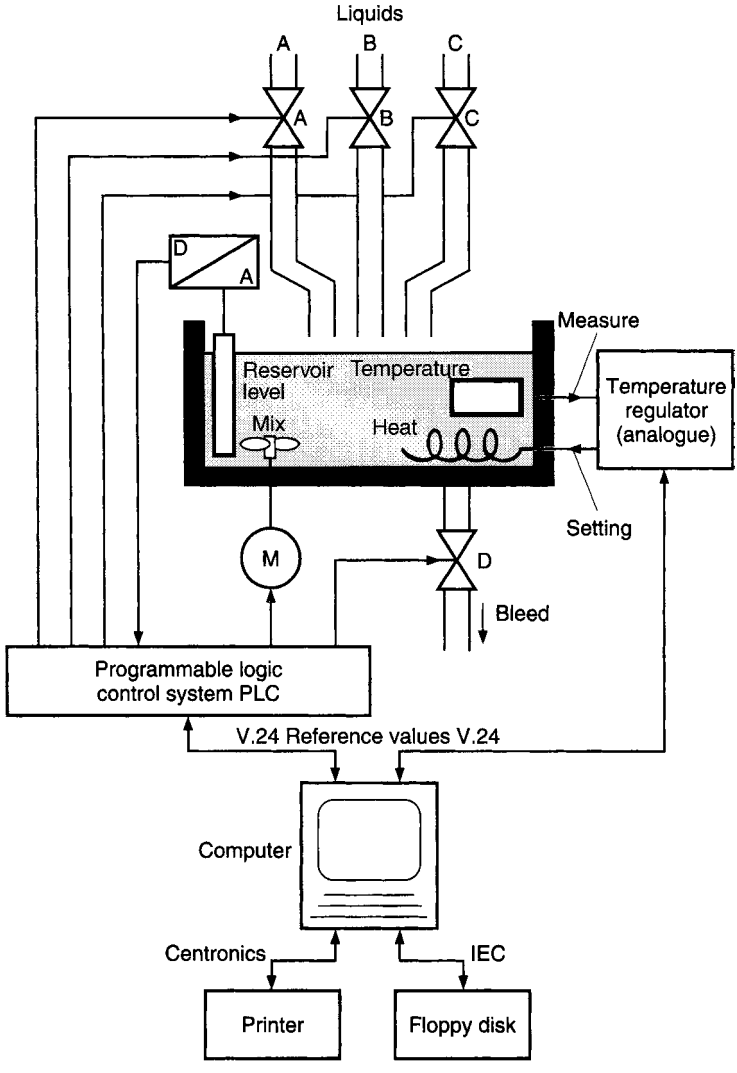


Figure 9.46 Networking of various systems in a liquid mixing operation.

Figure 9.46 shows a simple control task from chemical process technology as an example of the networking of various systems. In a liquid mixing plant, three different liquids—A, B, and C—are mixed in a specified proportion. Filling is effected by three valves—A, B, and C—and the mixture is extracted via one valve, D. Before the mixture is removed, it is heated to a prespecified reference temperature *T*.

A PLC controls the valves and the mixer. The temperature of the mixer is adjusted to the reference temperature by a two-step controller.

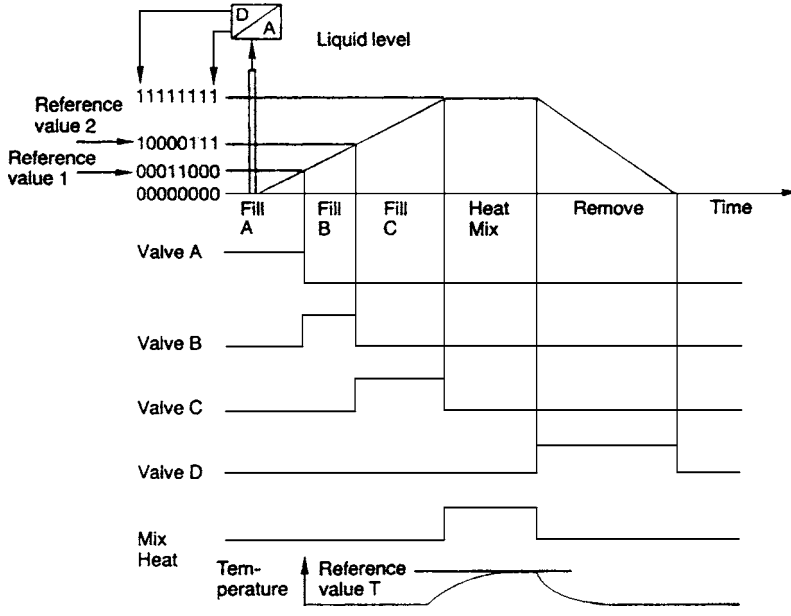


Figure 9.47 The control process in a liquid mixing operation.

An overriding computer prepares the reference values for the mixture and the mixing temperature and coordinates the PLC and regulator. In addition, the computer stores operational data and downloads the data to the printer as documentation.

Figure 9.47 illustrates the control process. The reference values 1 and 2, which determine the mixture ratio, are passed on by the computer to the PLC. The PLC controls the remainder of the sequence.

First, valve A is opened and liquid A is poured into the reservoir. The liquid level rises until the level specified by reference value 1 has been reached.

The liquid level in the reservoir is determined by analog means using a capacitive measuring probe. The analog value is transformed into 8-bit digital values. In the PLC, this value is compared with the stored reference value. If reference value and actual value agree, valve A is closed and valve B is opened.

Next, liquid B is poured in until reference value 2 has been reached. Then, valve B is closed and valve C is opened. Liquid C is poured in until the maximum value (reservoir full) is reached. Reference values 1 and 2 determine the mixture ratio A:B:C.

On reaching the maximum level of the liquid, the mixing motor switches ON and sends a signal to the computer, which in turn switches ON the temperature regulator. The regulator adjusts the temperature to the reference value and informs the computer when the reference value has been reached.

Next, the computer requires the PLC to switch ON the mixing motor and to open the bleed valve D. Once the reservoir has been emptied, the cycle is complete.

A new cycle is now started up with the same or modified reference values. Data communication between the PLC regulator and the computer is carried out serially via V-24 interfaces. Computer and printers communicate via a parallel interface, and computer and floppy disk, via a parallel IEC interface.

Further Reading

Boyes, G. S., (ed.), *Synchro and Resolver Conversion*, Memory Devices Ltd., Surrey, U.K., 1980.

Burr-Brown Corp., "Wiring and Noise Considerations," *The Handbook of Personal Computer Instrumentation*, 4th ed., sec 7, Burr-Brown, Tucson, 1989.

Cole, R., *Computer Communications*, Macmillan, London, 1986.

Grant, D., "Applications of the AD537 IC Voltage-to-Frequency Converter," application note, Analog Devices Inc., Norwood, Mass., 1980.

Hagill, A. L., "Displacement Transducers Based on Reactive Sensors in Transformer Ratio Bridge Circuits," *Journal Phys. E: Instrum.*, 16: 597–606 (1982).

Higham, E. H., *Pneumatic Instrumentation in Instrument Technology: Instrumentation Systems*, Butterworths, London, 1987.

Hilburn, J. L., and D. E. Johnson, *Manual of Active Filter Design*, 2d ed., McGraw-Hill, New York, 1983.

Huang, S. M., A. L. Scott, and M. S. Green, "Electronic Transducers for Industrial Measurement of Low Value Capacitances," *Journal of Phys. E: Sci. Instrum.*, 21: 242–250 (1988).

Netzer, Y., "Differential Sensor Uses Two Wires," *Electronic Design News*, 167, March 31, 1982.

Oliver, B. M., and J. M. Cage, "Electronic Measurements and Instrumentations," McGraw-Hill, New York, 1971.

Paiva, M. O., "Applications of the 9400 Voltage to Frequency, Frequency to Voltage Converter," application note 10, Teledyne Semiconductor, Mountain View, Calif., 1989.

Tompkins, W. J., and J. G. Webster (eds.), *Interfacing Sensors to IBM PC*, Prentice-Hall, Englewood Cliffs, N.J., 1988.

CHAPTER 10

SpectRx NIR Technology

10.1 An Introduction to SpectRx NIR Technology

This section presents certain theoretical aspects that apply to the SpectRx[1] FFT NIR Spectroradiometer. It is intended for readers who desire to acquire the working knowledge of the SpectRx™ system, shown in Fig. 10.1, in order to better understand its applications. The SpectRx™ system is an analytical system that can be implemented on pharmaceutical and medical production lines, Fig. 10.2. The System measures the chemical compositions and active ingredients of pharmaceutical tablets, capsules, and gels, Fig. 10. 3.

While it is technically impossible for the SpectRx™ NIR system to analyze or detect metallic substances, it however, detects and analyzes the organic compound deposited on the metallic surfaces such as the stainless steel medical stents, Fig. 10.4. When different identical pharmaceutical samples are taken from different batches at a different time or production location, the SpectRx™ NIR system provides identical spectra as illustrated in Fig. 10.5.

Those who are already familiar with FFT-NIR technology may want to use it as a refresher. The instruments and samples of its applications are also presented.

10.2 Abbreviations—Nomenclature

ADC	Analog-to-digital converter
DSP	digital signal processor
EMI	electromagnetic interference
FFT	fast Fourier transform
FOV	field of view
FFTIR	Fourier transform near infrared

[1] Proprietary of American SensoR$_x$, Inc.

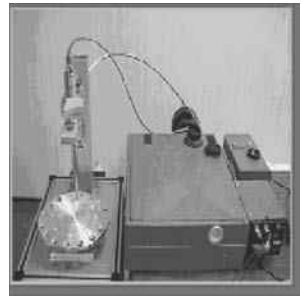


Figure 10.1
The SpectRx™ NIR system. (*Courtesy of American SensoRx, Inc.*)



Figure 10.2
The SpectRx™ NIR system installed in pharmaceutical production line. (*Courtesy of American SensoRx, Inc.*)

FWHM	full-width at half maximum
ILS	instrument line shape
NIR	near infrared
kb	kilobyte (equals 1024 bytes)
Mb	megabyte (equals 1,048,576 bytes)
MPD	maximum path difference (in one direction of the scan arm, or from ZPD)
NEP	noise equivalent power
NESR	noise equivalent spectral radiance
OD	optical density
OPD	optical path difference
RMS	root mean square
RSS	root Sum Square
SI	Sabrie's Index
SNR	signal-to-noise ratio
ZPD	zero path difference

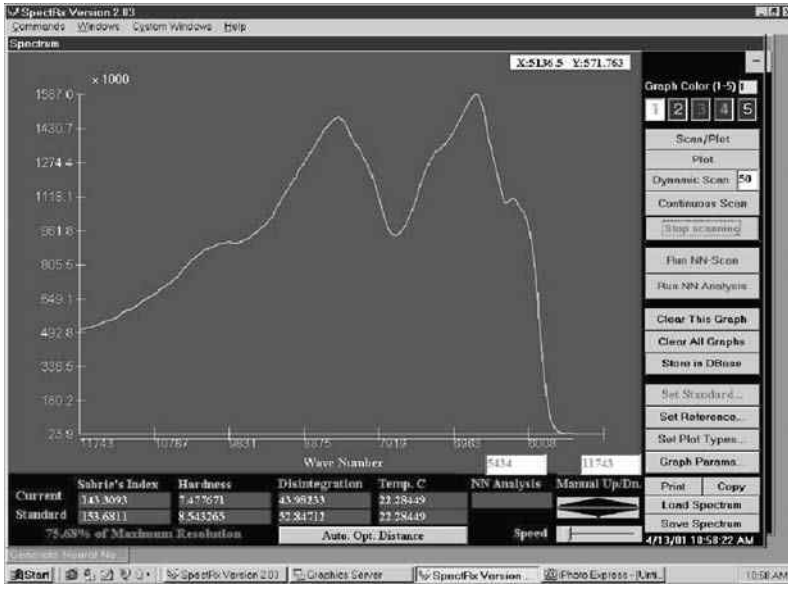


FIGURE 10.3 The SpectRx™ NIR system plotted the spectrum of a pharmaceutical tablet measuring its checmical compositions and active ingredient by weight. (*Courtesy of American SensoRx, Inc.*)

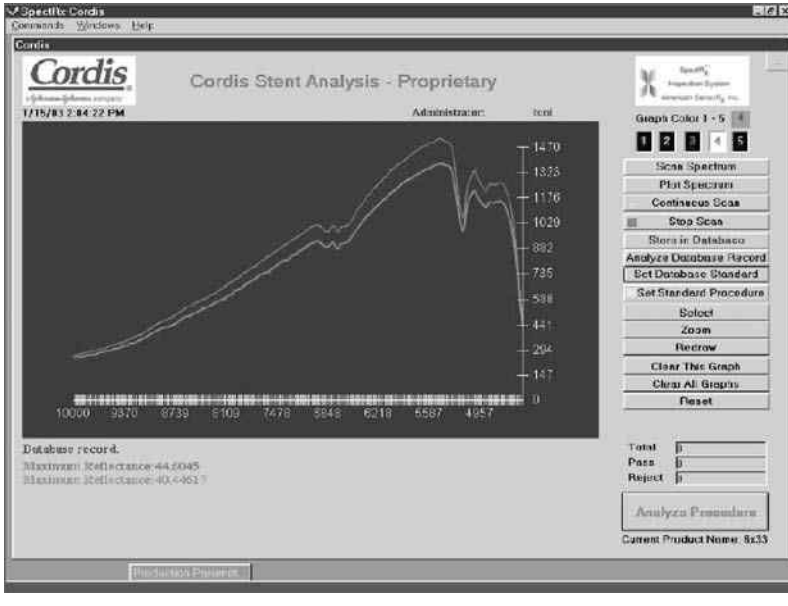


FIGURE 10.4 The SpectRx™ NIR system measured the polymer layers, and the drug layer deposited on a stent and provided a comparative data to a stent standard reference. (*Courtesy of American SensoRx, Inc.*)

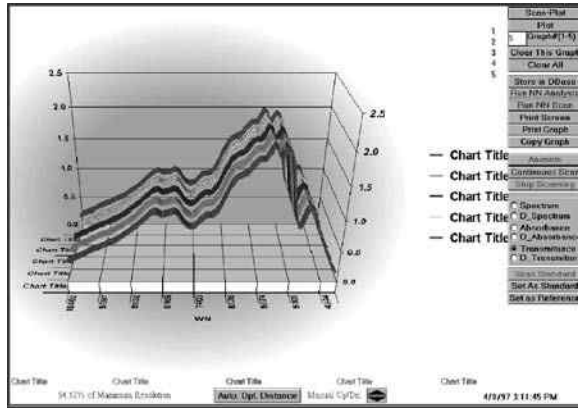


FIGURE 10.5 5 acceptable and identical tablets shows identical spectra plotted by the SpectRx™ analytical system. (*Courtesy of American SensoRx, Inc.*)

10.3 Glossary

aliasing The result of sampling that is too low in frequency to preserve the spectral content of the sampled signal. When the signal has a frequency greater than half the sampling frequency, it appears in the sampled signal at a lower frequency.

aperture stop (AS) The element limiting the size of the cone of radiation that an optical system will accept from an axial point on the source.

calibration The process by which a relation between a reference and a measurement is established.

Coaddition The coherent addition or signal averaging of interferograms. This is a technique to improve the signal-to-noise ratio by adding successive interferograms with corresponding data points digitized to exactly the same optical retardation. Spectral information will add up faster than noise.

divergence (θ) The angle under which a point on the primary optics views a source.

extended source A source filling the field of view (q.v.) of an instrument and radiating in all directions.

field of view (Ω) The solid angle under which a point on the primary optics views a source.

field stop (FS) The element that limits the field of view (q.v.) of an optical system.

intensity (I) The power emitted by a point source (q.v.) into a solid angle. Units are expressed in W/sr. The apparent intensity is the measurement of the intensity of a source without taking atmospheric transmission into account.

irradiance (E) The power collected by the primary (q.v.) of the radiometer (q.v.) per unit area. Units are expressed in W/cm^2.

point source A source of infinitesimal dimension that radiates in all directions.

primary In a telescope, the "primary" is defined as the first focusing mirror that converges incoming radiation into the optical system.

radiance (L) The power radiated from an object per solid angle unit per unit area. Units are expressed in W/cm^2 sr.

radiometer An instrument allowing the measurement of radiometric quantities such as E and L, under a given throughput (q.v.). Measurement involves comparing an unknown quantity to a known reference or standard.

Sabrie's Index (SI) The total transmitted or reflected spectral energy of an object receiving light radiation from an independent light source.

spectral intensity ($I\sigma$) The power emitted by a point source (q.v.) into a solid angle within a specified small spectral interval centered on frequency σ. Units are expressed in $W/sr\ cm^{-1}$. The apparent spectral intensity is the measurement of the spectral intensity of a source without taking atmospheric transmission into account.

spectral irradiance ($E\sigma$) The power reaching the primary (q.v.) of the radiometer (q.v.) per unit area within a specified small spectral interval centered on frequency σ. Units are expressed in $W/cm^2\ cm^{-1}$.

spectral radiance ($L\sigma$) The power radiated from an object per unit solid angle, per unit area within a specified small spectral interval centered on frequency σ. Units are expressed in $W/cm^2\ sr\ cm^{-1}$. This unit is used by the spectroradiometer for calibrating the spectrum.

spectroradiometer An instrument allowing the measurement of radiometric quantities such as $E\sigma$ and $L\sigma$, under a given throughput (q.v.). Measurement involves comparing an unknown quantity to a known reference or standard.

throughput (Θ) The factor quantifying the geometric aspect of power transfer between two objects.

wave number (σ) The optical frequency of light expressed in cm^{-1}.

10.4 Formula and Mathematical Relations

- The area of an optical element $A = \frac{\pi \times D^2}{4}$ where D is the diameter of the optical element.

- Wavelength in microns is converted from the wave number by $\lambda = \frac{10{,}000}{\sigma}$.
- The divergence of a telescope is approximated by $\theta = \frac{Dfs}{d}$ where *Dfs* is the diameter of the field stop and *d* the focal length of the primary.
- The solid angle of a beam at the entrance of a telescope is approximated by $\Omega = \frac{Afs}{d^2}$ where *Afs* is the area of the field stop and *d* the focal length of the primary.
- The relation between the solid angle and the divergence is $\Omega = 2\pi\left(1 - cos\frac{\theta}{2}\right)$. For small angles (e.g., in calculations related to the FOV of telescopes), the following formula provides a good approximation: $\Omega = \frac{\pi}{4}\theta^2$.
- The diameter of an object at a distance *d* and subtended by a divergence θ is given by $Ds = d\theta$.
- The throughput between two optical elements is given by $\Theta = A\Omega$, *Aas* being the area of the aperture stop, and Ω, the solid angle or divergence, introduced by the field stop. We can approximate Ω by taking the area of the field stop and dividing it by the square of the distance between both stops. The throughput is then expressed by $\Theta = \frac{Afs \times Aas}{d^2}$.
- Converting radiance to irradiance is done by multiplying the radiance by the solid angle of the telescope.
- Converting irradiance to apparent intensity is done by multiplying the irradiance by the square of the distance between the point source and the primary of the telescope.
- Converting apparent intensity-to-intensity is done by correcting the apparent intensity with an atmospheric transmission spectrum computed by modeling software (Fast Code, Lowtran, Modtran, and others).

10.5 The SpectRx FFT-NIR Technology Advantage

Spectral information can be obtained using several technical methods. Most of them rely on the dispersion of light to achieve spectral separation. This can be done with a prism or a grating. In this case, the spectral distribution is transformed into a spatial distribution and requires a scanning mirror in one axis. Circular variable filter radiometers are also currently used, but spatial/spectral smearing effects add to the other drawbacks of this method.

The fast Fourier transform (FFT) spectrometer does not rely on spatial dispersion, but rather on time dispersion of the spectral

information. This limits the number of moving components and provides a more reliable observation geometry. This is a further advantage that can be added to the three "classic" advantages of FFT spectrometry over dispersive spectrometry:

- Multiplexing—all wavelengths fall on the detector at all times
- Throughput—the aperture versus resolution limitation is reduced
- Wavelength accuracy—the internal laser ensures stability in ppm levels

In practical terms, the superiority of the FTIR spectroradiometer can be seen most clearly in its high spectral resolution and broadband capability.

The throughput, and thus the sensitivity, of the SpectRx FT Spectroradiometer is superior to that of a dispersive spectroradiometer—a tendency that increases as spectral resolution is increased.

SpectRx FT spectroradiometer has inherent broadband capability. When no optical filtering is used, the spectral bandwidth of the instrument is the same as that of the detector used, for example 1 to 5.5 μm in the case of an InSb detector. It is also important to note that there is no time penalty when acquiring a large bandwidth as opposed to a narrow one.

Finally, the most useful characteristic of a SpectRx FFT spectroradiometer is its flexibility. The spectral resolution of the SpectRx FFT Spectroradiometer can be instantly changed by computer. Also, because of the complete separation of spectral information (performed in the time domain) and spatial information (performed in the spatial domain), the "Field of View" configuration of the instrument can be changed in the field, simply by changing telescope modules.

10.5.1 Software Philosophy

Multiple software modules are developed with the SpectRx systems, including two vital modules: the C++/Visual Basic Windows-based acquisition module—Research-Acquire; and a Windows-based data manipulation and processing module.

The Research-Acquire routine provides the functions necessary for general SpectRx applications. It includes functions to acquire raw interferograms, uncalibrated spectra, radiometric reference spectra using calibrated sources, and calibrated spectra. It also acquires Fast Fourier Transformation—Near Infrared FFT-NIR or FFTIR-type data in transmission, such as raw, reference, transmittance and absorbance spectra, which can be useful for quick diagnostics in such things as the measuring of stability, the transmission capacity of optical filters, and a general examination of the spectral content of the source observed.

Research-Acquire also includes a basic display system and a set of post-processing functions to perform radiometric calibration, conversion to temperature units, integration over programmable bands, computation of theoretical Planck functions, addition, subtraction, multiplication, and ratios.

American SensoRx SpectRx software is used to view, process, and organize data. It allows easy data and graphics exchange with other Windows' applications such as spreadsheet and word processors. Data processing capabilities are fast and accurate. It includes a built-in programming language, Array Basic, so that advanced users can modify existing algorithms or create new ones for the individual requirements of specific experiments.

10.5.2 Data Flow Outlines

Data acquisition involves notions of size, type, and transfer rate. These are functions of scanning speed, spectral resolution, and spectral range. Temporal and spectral data represent the two types involved in an FT spectroradiometer and are linked by the Fourier transform.

The following points will help the user understand how data is generated and help plan an experiment by taking data flow into account.

- The scan length is a function of resolution.
- All wavelengths are acquired simultaneously whatever the scan length is.
- SpectRx systems acquire double-sided interferograms.
- The sampling rate is a function of the scanning speed and of the number of samples per fringe or over-sampling factor.
- The maximum observable spectral range is determined by the oversampling factor. In the SpectRx, the over-sampling factor can be 1 (no over-sampling) or 2, which provides for spectral ranges of 0 to 7899 cm^{-1} and 15798 cm^{-1}, respectively.
- Each data point is stored as a 16-bit number. Coadding or signal averaging is done with 32-bit numbers.
- The maximum number of interferograms, which can be recorded in kinetic mode, depends on the size of the DRAM buffer. This is 4 MB (expandable to 64 MB) in the case of the SpectRx 100, and 64 MB with the SpectRx 200.
- The required scanning speed is dictated by how fast an event is to be resolved. It is limited by the ADC speed, the detector response time, and the throughput of the electronics, including the communication links. It is 1 cm/s in the SpectRx 100 and 25 cm/s in the SpectRx 200.

10.6 Radiometric and Spectral Theory

Efforts are made to illustrate the presented theory with examples pertaining to the SpectRx system whenever possible.

10.7 Sensitivity

The sensitivity of the SpectRx system is its ability to transform input radiance into signal form. This sensitivity is a function of many parameters such as noise, transmission, modulation, throughput, and detector responsivity. The important quality of the system is its ability to distinguish the signal from the noise. This is expressed in terms of its signal-to-noise ratio (SNR), Fig. 10.6. In the context of detection, it is often stated that the *signal can be detected when its amplitude is at least three times the peak noise amplitude*. The computational steps are used to predict the signal-to-noise ratio of an interferometer system. The presented computations were performed using in-house simulation software. The evaluation of the SNR is described in the following sections.

10.7.1 Object View Spectral Radiation

Figure 10.7 shows the radiation from an ideal blackbody source at various temperatures between 100°C and 1600°C on a logarithmic

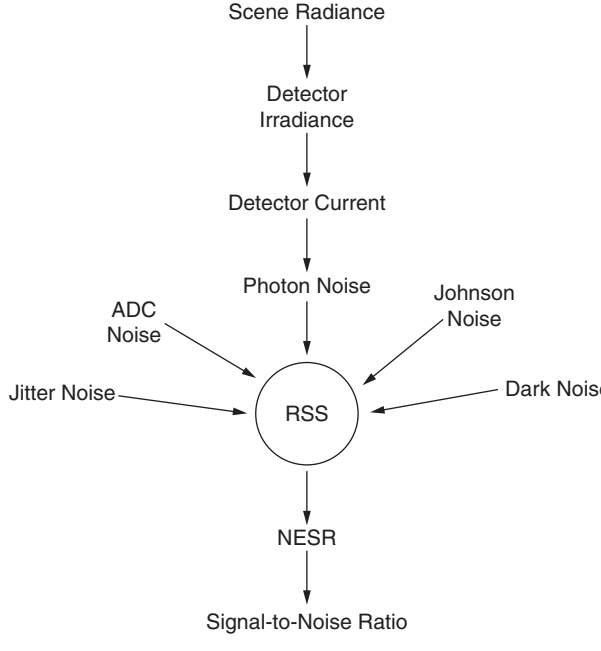


Figure 10.6 Signal-to-noise ratio computation steps.

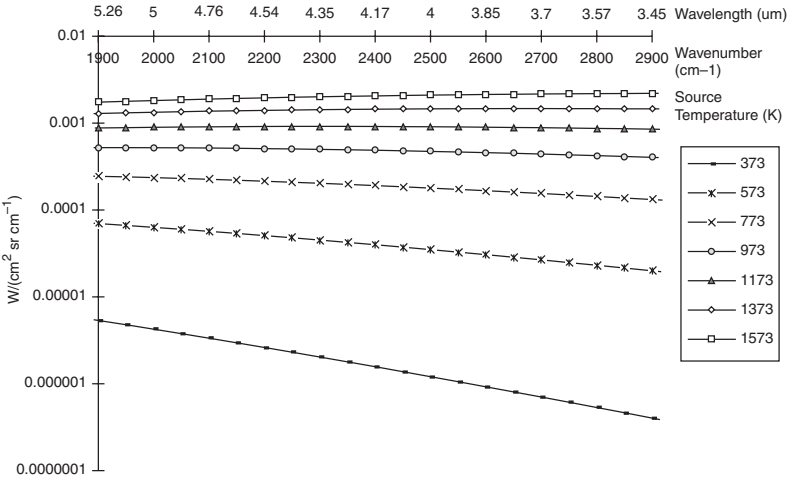


Figure 10.7 Object view spectral radiance versus temperature.

scale. It can be seen that the dynamic range requirement of the system is quite large, spanning almost four orders of magnitude at 3.5 μm.

Spectral radiance from the object view is given by Planck's equation:

$$L(\sigma) = \frac{C_1 \sigma^3}{\exp\left(\frac{C_2 \sigma}{T}\right) - 1} \tag{10.1}$$

where $C_1 = 1.191 \times 10^{-12}$ W/(cm^2 sr [cm^{-1}]^4) is the first radiation constant

$C_2 = 1.439$ K cm is the second radiation constant

σ = wave number in cm^{-1}

T = temperature of the object view in K

$L(\sigma)$ = radiated power density or *spectral radiance* in W/(cm^2 sr cm^{-1}).

This version of Planck's equation expressed in wave-numbers is different from the usual equation expressed in wavelengths:

$$M(\lambda) = \frac{C_1 \lambda^{-5}}{\exp\left(\frac{C_2}{\lambda T}\right) - 1} \tag{10.2}$$

The difference comes from the fact that Planck's function is a power density. Because $\sigma = 1/\lambda$, the infinitesimal wave-number element $d\sigma$ is equal to $-d\lambda/\lambda^2$. Thus:

$$\int_{\sigma_1}^{\sigma_2} L(\sigma)\, d\sigma = \int_{\lambda_1}^{\lambda_2} M(\lambda)\, d\lambda \tag{10.3}$$

and the apparent disagreement between Eqs. (10.1) and (10.2) disappears.

10.7.2 Transmission of the Spectroradiometer

Transmission of the spectroradiometer is the ratio of the radiant signal at the detector to that at the input port, taking into account the reflectivity of mirrors and transmission through windows, beam-splitters, and lenses. The description and transmission coefficient of these elements is shown in Table 10.1. The system transmission $T(\sigma)$ (no units) is given by:

$$T(\sigma) = \prod_{i=1}^{n} \varepsilon_i(\sigma) \tag{10.4}$$

where $\varepsilon_i(\sigma)$ is the spectral transmission efficiency of element i (no units).

For the SpectRx system, the standard beam-splitter is ZnSe, a nonhygroscopic material well adapted to field applications. The optical components of the SpectRx systems along with their transmittance is presented in Table 10.1. The total estimated system transmission is about 30 percent. This transmission can be improved to about 50 percent by using an anti-reflection coating on the protective windows; however, this will have a detrimental effect on the spectral range coverage. Another way to improve the transmission is to use different window materials. Unfortunately, most acceptable infrared window materials (Kbr, Kcl, etc.) are hygroscopic and require more care than ZnSe.

10.7.3 Throughput

The throughput of the SpectRx system is 0.004 cm^2. The system can be used without having to reduce the throughput to a spectral resolution below 1 cm^{-1}.

10.7.4 Power at the Detector

The power at the detector (W/cm^{-1}) is given by:

$$P^{\text{detector}}(\sigma) = \frac{L(\sigma) \cdot T(\sigma) \cdot \Theta}{2} \tag{10.5}$$

where $T(\sigma)$ = spectral instrument transmission efficiency (no units)
$L(\sigma)$ = spectral radiance from the object view ($W/cm^2\,cm^{-1}$ sr)
Θ = instrument throughput (cm^2 sr)

The SpectRx system is equipped with a single input output port and dual output ports. The evaluation of the power at the detector takes into account that the energy from a single input port is divided

equally between the two output ports, thus representing a factor of 2 in Eq. (10.5).

10.7.5 Detector Current

The detector's responsivity is the detector output current for a given input illumination power. The responsivity is directly proportional to the wavelength multiplied by the quantum efficiency $\eta(\sigma)$. This is because shorter wavelength photons have more energy per photon; therefore, there are fewer photons per watt of flux of either the signal or the background. The responsivity is thus dependent on the spectral frequency.

10.7.6 Sources of Noise

Noise refers to any electrical output other than the desired signal. Some noise sources are fundamental, and for several reasons cannot be avoided. A few of these reasons are:

- Photons do not arrive at a constant rate.
- Atoms in the detector vibrate slightly.
- Electrons move randomly.

Other noise sources arise externally and can be minimized, such as:

- Electrical interference
- Temperature fluctuation
- Vibration that causes electrical components to shift

10.7.7 Photon Noise

Photon noise or shot noise is unavoidable. It is due to the random arrival of photons at the detector. Photon noise $(A/\sqrt{Hz})$ for a photovoltaic detector (InSb) is given by:

$$I^{\text{photon}} = \sqrt{2 \cdot e \cdot I^{\text{detector}}} \tag{10.6}$$

It becomes $I^{\text{photon}} = \sqrt{e \cdot I^{\text{detector}}}$ for a photoconductor (MCT).

Although photon noise seems smaller in the case of a photoconductor detector than in a photovoltaic one, the sensitivity of photoconductors is, in general, relatively small.

Because photon noise is unavoidable, it is the ultimate limiting factor of the system's performance (SNR). Whenever possible, all other sources of noise should be kept lower than the photon noise.

Other sources of noise, such as the Johnson noise from the first amplifier stage feedback resistor, quantization noise, and detector dark noise contribute to the total noise.

10.7.8 Johnson Noise (InSb Detectors)

Johnson noise, roughly defined, is the random variation of voltage due to the thermal agitation of charge carriers in a resistor. Originally described and measured by J. B. Johnson, a physicist at Bell Laboratories who performed his experiments with a vacuum tube amplifier and thermocouple, the effect was first explained theoretically by H. Nyquist in 1928.

The Johnson noise $(A/\sqrt{Hz})$ of the feedback resistor of the first stage of amplification is given by:

$$I^{\text{Johnson}} = \alpha \cdot \sqrt{\frac{4kT}{R_f}} \tag{10.7}$$

where R_f = feedback resistor (Ω)
$k = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant
α = noise factor for imperfect resistors (no units), estimated by the manufacturer to be 1.5
T = temperature of the resistor (K)

In the SpectRx design, Johnson noise never dominates the total noise of the system. It increases the total system noise by less than 10 percent with the coldest object view temperature and by even less as the object view temperature is increased.

10.7.8.1 In depth study of Johnson Noise (InSb Detectors)

The circuit model of Nyquist is adopted in order to understand the behavior of the resistor. Imagine a resistor R at temperature T connected in series with an identical resistor R at the same temperature, through a transmission line of length L. To avoid losses due to radiation from the wires connecting the resistors, one line may be enclosed by the other, as in a coaxial cable. Imagine the first resistor to be held at a finite temperature T, so that thermally agitated electrons within produce a fluctuating voltage signal, and then represent this signal with the source V, as shown in Fig. 10.8.

By placing the second resistor at the distant end of the transmission line, a boundary condition has been achieved. The transmission

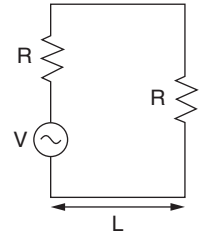


Figure 10.8 Thermally agitated electrons to produce voltage fluctuation.

line between 0 and L ought to accommodate voltage waves $V(x, t)$ obeying boundary conditions

$$V(0, t) = V(L, t) \tag{10.8}$$

where, generally,

$$V(x, t) = V_0 e^{i(kx-\omega t)} \tag{10.9}$$

The voltage waves propagate at the speed of light,

$$c = \frac{\omega}{|k|} \tag{10.10}$$

The boundary condition implies

$$kL = 2\pi n,\ n \in Z \tag{10.11}$$

or, assuming the transmission line to be very long (so that many modes n are permitted),

$$\begin{aligned} L\,\mathrm{d}k &= 2\pi\,\mathrm{d}n \\ \Rightarrow \frac{1}{L}\mathrm{d}n &= \frac{1}{2\pi}dk \end{aligned} \tag{10.12}$$

This is simply the density of modes—the number of modes, or "voltage states" allowed per unit length—in the line. Both right- and left-propagating voltage waves are admitted in the solution, so that both k and its associated frequency w range over positive and negative values.

Propose that each mode of voltage oscillation is a mode in which the resistor on the right can absorb radiation from the (quantized) electromagnetic field. Since one imagines the resistor to be an idealized one-dimensional "lumped element," propose also that it absorbs radiation in the single dimension of the transmission line. Then the power—energy per unit time—absorbed by the resistor is the energy of a certain mode multiplied by the number of quanta in that mode, integrated over all modes, multiplied by the rate at which the quanta are absorbed (c/L),

$$\begin{aligned} P_{\text{absorbed}} &= \frac{c}{L}\int_{-\infty}^{\infty} \hbar|\omega| \frac{1}{e^{\frac{\hbar|\omega|}{kB^T}} - 1}\frac{L}{2\pi}dk \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \frac{\hbar|\omega|}{e^{\frac{\hbar|\omega|}{kB^T}} - 1} \end{aligned} \tag{10.13}$$

The current generated in the circuit by the fluctuating voltage is simply

$$I = \frac{V}{2R} \tag{10.14}$$

so that the average power dissipated (emitted) by the resistor is

$$P_{\text{emitted}} = \frac{\langle V^2 \rangle}{4R} \tag{10.15}$$

The time average is

$$\langle V^2 \rangle = \lim_{T\to\infty} \frac{1}{T} \int_{-T}^{T} dt\, V(t) V^*(t) \tag{10.16}$$

In Eq. (10.16) and the expressions that follow, the voltage function $V(t)$ is considered at frequency ω, Eq. (10.9).

Then, the *power spectrum*—the distribution of power, or square voltage, over many frequencies—is defined as follows:

$$S(\omega) = \int_{-\infty}^{\infty} \mathrm{d}\tau\, R(\tau) \mathrm{e}^{-i\omega T}, \tag{10.17}$$

where

$$R(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T}^{T} \mathrm{d}t\, V(t) V^*(t-\tau) \tag{10.18}$$

is the *autocorrelation function* of the function $V(t)$ over the period $[-T, T]$. Then, integrating $S(w)$ over all frequencies,

$$\begin{aligned}
\int_{-\infty}^{\infty} \mathrm{d}\,\omega\, S(\omega) &= \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} \mathrm{d}\omega \int_{-\infty}^{\infty} \mathrm{d}\tau \int_{-T}^{T} \mathrm{d}t\, V(t) V^*(t-\tau) \mathrm{e}^{-i\omega T} \\
&= \lim_{T\to\infty} \frac{2\pi}{T} \int_{-\infty}^{\infty} \mathrm{d}\tau \int_{-T}^{T} \mathrm{d}t\, V(t) V^*(t-\tau) \delta(\tau) \\
&= \lim_{T\to\infty} \frac{2\pi}{T} \int_{-T}^{T} \mathrm{d}t\, V(t) V^*(t) \\
&= 2\pi \langle V^2 \rangle
\end{aligned} \tag{10.19}$$

Equating the power dissipated by the resistor, Eq. (10.15) and that absorbed due to thermal radiation Eq. (10.13), it leads to

$$\frac{\langle V^2 \rangle}{4R} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega \frac{\hbar |\omega|}{e^{\frac{\hbar|\omega|}{kB^T}} - 1} \tag{10.20}$$

By the relationship in Eq. (10.19),

$$\frac{1}{4R} \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega\, S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega \frac{\hbar |\omega|}{e^{\frac{\hbar|\omega|}{kB^T}} - 1} \tag{10.21}$$

so that

$$S(\omega) = 4R \frac{\hbar |\omega|}{e^{\frac{\hbar|\omega|}{kB^T}} - 1} \tag{10.22}$$

Note: Equating the integrands of two definite integrals cannot be justified mathematically. A strict derivation would then appeal to the balance of emitted and absorbed power over an arbitrarily small frequency interval, rather than simply over the full range; this is the argument Nyquist constructs in Fig. 10.9. There, he considers the following "circuit thought experiment": Imagine that over all frequencies, the exchange of power between the two resistors (at equal temperature) is equal, but over a small range of frequencies, the resistor on the left radiates more power than it absorbs from that on the right. Then, introduce a non-dissipative circuit element between the two that interferes more with the transfer of energy in the frequency range of interest than in any other range, so that now more total power flows from the right resistor to the left than in the other direction. But since the resistors are initially at the same temperature, to let the system evolve would be to raise the temperature of a hotter thermal body using the heat of a colder body, by simply introducing a passive circuit element. So, there must be a "detailed balance" of power over each tiny range of frequency, in accordance with the second law of thermodynamics.)

Then, in the classical limit, $\hbar\omega \quad k_B T$,

$$S(\omega) \equiv S_V(\omega) = 4Rk_B T. \tag{10.23}$$

Equation (10.23) shows the Nyquist relation between the voltage power spectrum and temperature. Accordingly, the current power spectrum is closely related to the following equation,

$$S_I(\omega) = \frac{1}{R^2} S_V(\omega) = \frac{4k_B T}{R} \tag{10.24}$$

Since the actual resistor will be situated in a circuit with an amplifier designed to enhance the minuscule voltage fluctuations, and since the transmission line will exhibit capacitive effects, a realistic configuration of the circuit diagram would therefore, appear as in Fig. 10.9.

The parallel combination of current source i and resistor R is the "Norton equivalent" of R and a small fluctuating voltage source,

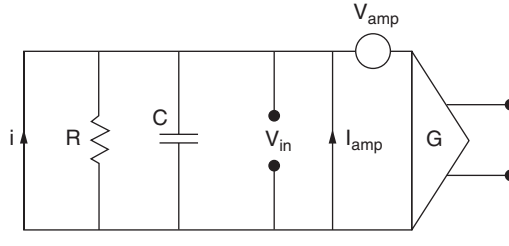


Figure 10.9 Enhancing small voltage fluctuations.

illustrated in Fig. 10.9. The net input voltage is then denoted by V_{in}, and the additional current and voltage associated with the operation of the amplifier itself are labeled I_{amp} and V_{amp}. The input voltage power spectrum takes the simple form

$$S_{V_{\text{in}}} = S_I[\Re(R \parallel C)]^2 + S_{I,\text{amp}}[\Re(R \parallel C)]^2 \tag{10.25}$$

where S_I is the current power spectrum due to the resistor as illustrated in Eq. (10.24), $S_{I,\text{amp}}$ is the current power spectrum due to the amplifier, and $[\Re(R \parallel C)]^2$ is the squared real part of the parallel resistor–capacitor combination.

The voltage power spectrum output generated through the amplifier can be represented by:

$$S_{V\,\text{out}} = (S_{V_{\text{in}}} + S_{V,\text{amp}})G^2 \tag{10.26}$$

where $S_{V,\text{amp}}$ is the voltage power spectrum of the amplifier itself, and G is its gain.

Eqs. (10.23) through (10.26) relate the measured quantity $S_{V,\text{out}}$.

Another important feature of Nyquist's model is that *Johnson noise is independent of the material* chosen for the resistor. This feature is exploited in accurate thermometry, allowing precise temperatures measurements without worrying about the particular type of sensor material having a contaminating effect. Additionally, Johnson noise found many applications in classical information cryptography.

10.7.9 Quantization Noise

Quantization noise $(A/\sqrt{Hz})$ arises from the finite size of the quantization steps of the *analog-to-digital converter* (ADC). It is given by:

$$I^{\text{quantization}} = \frac{R}{G \cdot 2^N \cdot \sqrt{12} \cdot \sqrt{f_N}} \tag{10.27}$$

where R = range of the ADC (V)
G = gain from detector to quantization (Ω)
N = number of bits for quantization (no units)
f_N = Nyquist frequency, which is half the sampling frequency (Hz)

Quantization noise is of concern in photometric systems utilized for observing object views of high radiance, such as the sun.

10.7.10 Dark Noise

Dark noise is the electronic noise observed when no radiation reaches the detector. This noise depends on the type of detector and the electronic circuitry used. The values used in the case of an InSb detector are $1.5 \times 10^{-14}\,A/\sqrt{Hz}$ for each detector. As for the quantization noise,

dark noise in SpectRx systems remains a negligible contributor to system noise.

10.7.11 Jitter Noise

Jitter noise is noise caused by jitter during sampling of the interferogram. Sampling jitter is mainly due to ADC jitter, laser intensity fluctuations, and the noise from the electronics that trigger the ADC.

Jitter noise can become troublesome at elevated scanning speeds. This is because at high speeds, a given jitter becomes non-negligible as compared to the sampling interval. Jitter noise is especially annoying because it is proportional to the measured signal; therefore, the SNR due to jitter cannot be improved by observing a warmer object view. Because of this, rapid scanning systems require carefully designed electronic circuits to trigger the ADC. The reader is referred to (RD1) for more details.

10.7.12 Scanning Instability Noise

Scanning instability noise can be attributed to scanning speed instabilities together with uneven spectrometer response and/or a delay mismatch between the ADC triggering and the IR signal. Mainly because of the analog filter response at high frequencies, the SpectRx spectroradiometer is operated with a non-zero slope gain. This is not a problem if the scanning speed that converts optical speeds to electrical speeds is constant. This is not the case, however, and some noise-equivalent current results from the peak amplitude of the speed variation and the analog filter slope. Filter slope noise is often a non-negligible contributor to FTIR system noise. This is the main reason why analog filter design is carefully analyzed.

Speed variations can also introduce noise if the ADC triggering circuitry is not perfectly synchronized with the IR signal. Because it has a larger bandwidth than the IR channel, the sampling laser fringe detection circuitry usually has a smaller associated delay. If this delay-mismatch is not corrected, the spectroradiometer will sample the interferogram with a constant negative time offset: $-\Delta t$. This is not a problem if the scanning speed is constant, but as this is not the case, the uneven sampling resulting from the delay-mismatch and speed variations converts into a noise-equivalent current.

Like jitter noise, these two sources of noise are of great concern, especially as they are proportional to the measured signal, so that the SNR due to these sources cannot be improved by observing warmer object views. Careful design of the analog filter cut-off frequency as well as delay-mismatch compensation is therefore applied.

10.7.13 Instrument Efficiency

The modulation efficiency of the interferometer depends on many factors, such as the imperfections of the beam-splitter, and the shear

and tilt of the recombining wavefronts. Modulation due solely to the imperfections of the beam-splitter is given by:

$$M_{BS}(\sigma) = 4\, r_{BS}(\sigma) \cdot t_{BS}(\sigma) \tag{10.28}$$

where M_{BS} = beam-splitter modulation
r_{BS} = beam-splitter reflectivity
t_{BS} = beam-splitter transmissivity

Maximum beam-splitter modulation (i.e., $M_{BS} = 1$) occurs when r_{BS} and t_{BS} are equal to 0.5.

The typical interferometer modulation for input port A is 70 percent using five-second precision corner-cube retroreflectors. Better modulation can be expected with the use of one-second precision retroreflectors.

It should be noted that the total energy is split between the two output ports. This factor is normally included in the instrument efficiency ξ together with beam-splitter modulation and system transmission.

$$\xi = \frac{T(\sigma) \cdot M_{BS}}{2} \tag{10.29}$$

Input port B can be set to look at a cold source to reduce the photon noise.

10.7.14 NESR

The system's noise equivalent spectral radiance (NESR) (W/cm² sr cm⁻¹) is given by:

$$NESR = \frac{\sqrt{((I^{\text{photon}})^2 + (I^{\text{quantization}})^2 + (I^{\text{darknoise}})^2 + (I^{\text{Johnson}})^2 + (I^{\text{jitter}})^2 + (I^{\text{scan.inst}})^2)}}{\Delta\sigma \cdot \Theta \cdot \xi \cdot \sqrt{t} \cdot R(\sigma) \cdot F} \tag{10.30}$$

where I^{photon} = noise current due to photon noise $(A/\sqrt{Hz})$
$I^{\text{quantization}}$ = noise current due to quantization noise $(A/\sqrt{Hz})$
I^{Johnson} = noise current due to the Johnson noise of the resistance of the first amplification stage $(A/\sqrt{Hz})$
$I^{\text{darknoise}}$ = noise current due to dark noise $(A/\sqrt{Hz})$
I^{jitter} = noise current due to the jitter of the sampling $(A/\sqrt{Hz})$
$I^{\text{scan. inst.}}$ = noise current due to scanning speed variations in the spectroradiometer, coupled with the analog filter slope and the delay mismatch $(A/\sqrt{Hz})$
t = acquisition time (s)
$\Delta\sigma$ = spectral interval (cm⁻¹)

ξ = instrument efficiency (no units)
$R(\sigma)$ = responsivity of the system (A/W)
Θ = instrument throughput (cm^2 sr)
F = noise reduction apodization factor (no units); equal to 1 if no apodization is used

Being incoherent, the various sources of noise are added in an RSS manner in Eq. (10.11).

10.7.15 Signal-to-Noise Ratio

Finally, the signal-to-noise ratio SNR is given by:

$$SNR(\sigma) = \frac{P(\sigma)}{NESR} \tag{10.31}$$

10.8 Instrument Line Shape and Spectral Resolution

The ILS (instrument line shape) is the response of the spectrometer to a monochromatic spectral stimulus. This is illustrated schematically in Fig. 10.10. A distinct ILS spanning from $\sigma = -\infty$ to $+\infty$ exists for each wave-number σ', although it changes from one σ' to the next in a slow and progressive manner. The ILS at a given wave number σ' represents the noiseless spectrum that would be obtained if the spectrometer were submitted to monochromatic radiation—for example, that of a laser at $\lambda' = 1/\sigma'$. The ILS function $ILS_{\sigma'}(\sigma)$ is thus characterized by two indices, σ and σ', as illustrated in Fig. 10.10.

An observed spectrum $S^{\text{observed}}(\sigma)$ is the convolution of the ILS functions with the spectral distribution of photon incidence $B(\sigma)$.

$$S^{\text{observed}}(\sigma) = \int_{-\infty}^{\infty} B(\sigma - \sigma')\, ILS_{\sigma-\sigma'}(\sigma')\, d\sigma' \tag{10.32}$$

If the spectral features or variations of $B(\sigma)$ are much wider than the width of the ILS peak, then $S^{\text{observed}}(\sigma)$ will be very similar to $B(\sigma)$.

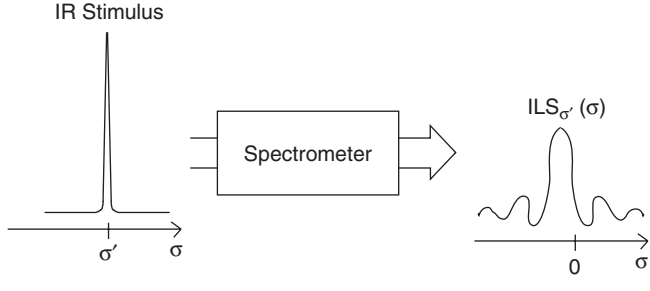


Figure 10.10 Monochromatic IR stimulus and instrumental response function (ILS).

On the other hand, if $B(\sigma)$ has features much narrower than the ILS, $S^{observed}(\sigma)$ will be greatly affected by the response of the system. This is often referred to as under-resolving. In this case, the narrow lines from $B(\sigma)$ would all have widths very similar to that of the ILS. In the limiting case of an isolated monochromatic line, the measured spectrum is the ILS itself, as illustrated in Fig. 10.10.

With respect to the instrument structure, many parameters influence the ILS, such as the finite mirror scan length, the finite divergence in the interferometer, the IR alignment, and so on.

10.8.1 Spectral Resolution

In the present system, the width of the ILS is mainly governed by the extent of the finite length of the interferogram. For a system with negligible divergence that is equipped with a round detector, the normalized ILS is given by:

$$ILS(\sigma',\sigma) = \frac{\mathrm{Sin}(2\pi\sigma MPD)}{2\pi\sigma\, MPD} \tag{10.33}$$

The ILS function due to the finite interferogram length is *independent of* σ (i.e., it is the same for all wave numbers). This ILS function is illustrated in Fig. 10.10. The line shape seen in Fig. 10.11 is very similar to the actual simulated ILSs displayed in Fig. 10.10.

The FWHM of the ILS given by Eq. (10.15), $FWHM_{\mathrm{boxcar}}$, can be derived analytically and is:

$$FWHM_{\mathrm{boxcar}} = \frac{1.207}{2MPD} \tag{10.34}$$

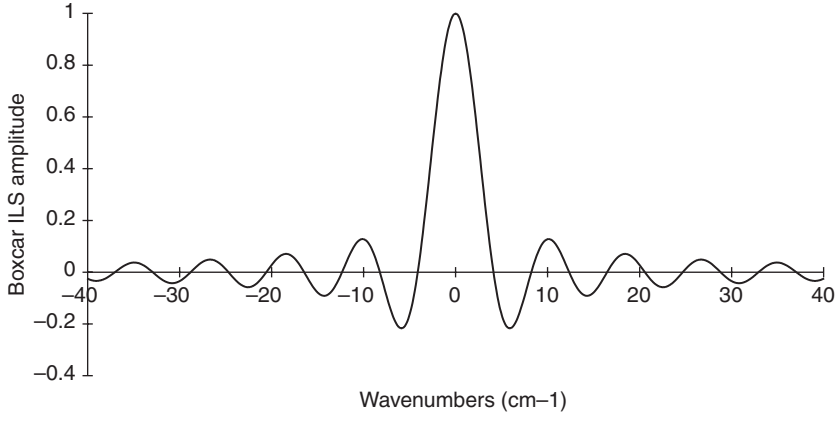


Figure 10.11 Plot of $\sin(2\pi\sigma L)/(2\pi\sigma L)$ (where L = 0.122 cm).

For an MPD of 0.122 cm, the sole contribution of the finite interferogram length to the resolution yields 4.96 cm^{-1}. The actual FWHM of the ILS can be written as:

$$FWHM_{ILS} = \beta \frac{1.207}{2MPD} \tag{10.35}$$

where β is a broadening factor to account for parameters other than the finite interferometer length.

The broadening factor β is mainly due to the finite divergence in the interferometer, a contribution that varies with the wave number, being larger for higher wave numbers.

10.8.2 Apodization

As stated in the previous sections, the ILS is governed mainly by the finite interferogram length and thus closely resembles a Sin(x)/x function as seen in Fig. 10.11. This function is sometimes considered a disadvantage because it has amplitude oscillations, which are large (21 percent for the first ripple), and which die off very slowly. It is possible to reduce the oscillations of the ILS by using post-acquisition digital processing on the acquired interferogram prior to Fourier transformation. This is referred to as *apodization*, or "removal of feet," since the purpose of the operation is to remove or reduce the oscillations. This improvement of the ILS is gained at the cost of spectral resolution or increased ILS FWHM. The effect of various types of apodizing functions on the ILS FWHM is illustrated in Fig. 10.11. Triangular, Hamming, and Gaussian functions are commonly used apodization functions. Boxcar apodization is the use of no apodizing function. For a given MPD, one can see in Fig. 10.12 that the ILS

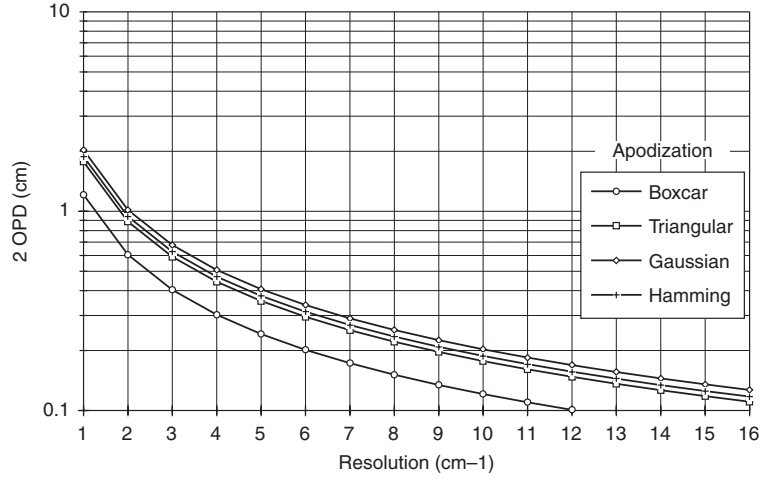


Figure 10.12 Optical path versus resolution setting.

FWHM for boxcar apodization is better than that obtained using triangular, Hamming, or Gaussian apodization by a factor of 1.47, 1.52, and 1.68, respectively.

10.9 Radiometric Calibration

Radiometers with no spectral capability, whether they are single-pixel or imaging systems, are extremely difficult to calibrate radiometrically. The main cause for this is that at the detector level, a radiometer integrates the radiant signal over a certain spectral range producing a single output. The integration translates into a loss of information, which limits the interpretation of the measurement.

The measurement produced by a radiometer is intended to be the radiance of an object view integrated over a specific spectral range. The actual output of a radiometer, however, is the radiance of the object view integrated over a specific spectral range but weighted by the *spectral response of the system*. Even after calibrating with a source of known radiance (e.g., a calibration blackbody), the uncertainty of the spectral response of the system induces an uncertainty in the intended measurement. This is because the response of the system is never flat within the spectral band pass, and thus spectrally varying object views of different integrated radiances can produce the same output. In other words, uneven spectral responses are linked with radiometric errors.

The SpectRx Spectroradiometer represents an improvement over non-spectral radiometers. First, the addition of many spectral channels contributes to the narrowing of each channel, thus potentially making each of them more uniform. Another advantage is that it may be possible to restore distorted signals. If spectral channels are located in a contiguous manner, spectrally side-by-side, an over-evaluation of a given channel is accompanied by the complementary under-evaluation in neighboring channels. This means that the correct radiance of a feature in the spectrum can be evaluated by spectrally integrating over all channels that responded to the signal. Spectroscopically, this is a known phenomenon. When a spectrometer is operated at a resolution lower than the intrinsic features of an object view, the representation of these features in the spectrum is broader and weaker, but the integrated intensity is correct.

Consequently, the radiometric characterization of a FTIR-based spectroradiometer is a straightforward operation that leads to a radiometrically sound calibration. This is due to the intrinsically large number of spectral channels in FTIR-based systems. In fact, the spectral response of a channel is given by the ILS function described in the previous section. An FT-based spectroradiometer has a large number of overlapping and smoothly varying channels.

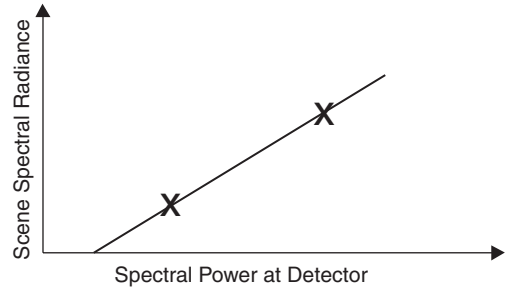


Figure 10.13 Linear relationship between the object view spectral radiance and the power at the detector.

10.9.1 Calibration Theory

The ideal radiometer is a *linear instrument* (i.e., the measured signal for each pixel and each spectral channel is proportional to the radiant spectral power at the detector). This is illustrated in Fig. 10.13. The power at the detector is composed of two contributions, the spectral power from the object view, and that from the thermal emission of the spectrometer itself. Because of this, the response line in Fig. 10.13 does not cross the abscissa at x = 0 but rather at some value corresponding to the thermal emission of the spectrometer. A calibration for such a system thus requires at least two measurements, as illustrated by the ×'s in Fig. 10.13. If the emission of the instrument could be neglected, only one characterization measurement would be necessary.

Let us now look at the calibration equations. An uncalibrated measurement can be expressed as:

$$S^{\text{Measured}}(\sigma) = K(\sigma)(L^{\text{Source}}(\sigma) + M^{\text{Stray}}(\sigma)) \tag{10.36}$$

where $S^{\text{Measured}}(\sigma)$ = measured complex spectrum (arbitrary)
$K(\sigma)$ = complex instrument response function (arbitrary cm^2 sr cm^{-1}/W)
$L^{\text{Source}}(\sigma)$ = source spectral radiance (W/cm^2 sr cm^{-1})
$M^{\text{Stray}}(\sigma)$ = spectral power of the stray radiance (W/cm^2 sr cm^{-1})

An interferogram always has a certain degree of asymmetry due to dispersion present in the beam-splitter (the wavelength-dependent refractive index) and in the amplification stages of the detectors (the frequency-dependent electronic delays). This asymmetry causes the Fourier transform of the interferogram (i.e., the spectrum) to have an imaginary part. At this point, a phase correction can be applied to the complex spectrum to yield a real spectrum. With the present calibration algorithm, however, the phase correction is not needed and instead we can work with complex spectra.

A radiometric calibration is a two-step process. First, the two unknowns per detector *radiometric gain* $K(\sigma)$ and *radiometric offset* $M^{\text{Stray}}(\sigma)$, are determined using an experimental step called a *characterisation*. Second, the *calibration* is applied to an un-calibrated measurement to produce a calibrated spectrum.

The gain and offset characterisation requires two measurements, one hot blackbody measurement and one cold blackbody measurement, ideally both uniformly filling the field of view. This is referred to as a *two-point calibration*, as shown in Fig. 10.13. The temperatures of the calibration blackbody measurements are judiciously set in order to minimize calibration error.

Applying Eq. (10.17) to the measurements of the hot and cold blackbodies, we obtain the following equations:

$$L_H(\sigma) = S_H^{\text{Measured}}(\sigma) / K(\sigma) - M^{\text{Stray}}(\sigma) \tag{10.37}$$

$$L_C(\sigma) = S_C^{\text{Measured}}(\sigma) / K(\sigma) - M^{\text{Stray}}(\sigma) \tag{10.38}$$

where $L_H(\sigma)$ and $L_C(\sigma)$ are the theoretically calculated spectral radiances using the emissivity of the blackbody and the Planck function at the temperature of the blackbody:

$$L_x(\sigma) = \varepsilon_x(\sigma) P_{T_x}(\sigma) = \varepsilon_x(\sigma) \frac{C_1 \sigma^3}{\exp\left(\frac{C_2 \sigma}{T_x}\right) - 1} \tag{10.39}$$

where $\varepsilon_x(\sigma)$ = emissivity of the x blackbody
x = C for the cold blackbody and H for the hot blackbody
$C_1 = 1.191 \times 10^{-12}$ W cm^2 and is the first radiation constant
$C_2 = 1.439$ K cm and is the second radiation constant
σ = wave-number (cm^{-1})
T_x = temperature of the x blackbody (K)

The solution to Eqs. (10.18) and (10.19) yields Eqs. (10.21) and (10.22). It is then simple to solve for the two unknowns $K(\sigma)$ and $M^{\text{stray}}(\sigma)$.

$$K(\sigma) = \frac{S_H^{\text{Measured}}(\sigma) - S_C^{\text{Measured}}(\sigma)}{L_H(\sigma) - L_C(\sigma)} \tag{10.40}$$

$$M^{\text{Stray}}(\sigma) = \frac{L_H(\sigma) \cdot S_C^{\text{Measured}}(\sigma) - L_C(\sigma) \cdot S_H^{\text{Measured}}(\sigma)}{S_H^{\text{Measured}}(\sigma) - S_C^{\text{Measured}}(\sigma)} \tag{10.41}$$

The calibrated spectrum is given by:

$$S^{\text{Calibrated}}(\sigma) = S^{\text{Measured}}(\sigma) K^{-1}(\sigma) - M^{\text{Stray}}(\sigma) \tag{10.42}$$

This method assumes that the operating parameters are not changed between the measurement of the calibration spectra and that of the object view spectra. The scanning speed and the optical configuration must remain the same. Since the experiment requires the acquisition of data in both directions of the interferometer stroke, two sets of $M^{Stray}(\sigma)$ and $K(\sigma)$ must be provided (one for each direction) for the calibration.

The radiometrically corrected measurements are also complex spectra. However, all spectral information should be contained in the real part of the complex spectra. The imaginary part should only contain noise. The imaginary part can be used to validate the spectra and to estimate the noise levels. After verification, the imaginary part can be dropped to free up storage space.

10.9.2 Multiple Point Calibrations

The calibration procedure can be extended to use multiple reference points instead of two points. Multiple point calibrations are used to improve the quality of the obtained calibrated spectra. Two types of errors can be addressed with multiple point calibrations: random errors due to noise, and systematic errors due to the nonlinear response of the spectroradiometer.

10.9.3 Linear Multiple Point Calibrations

Linear multiple point calibrations are used to reduce the impact of random uncertainty on the calibration curve. When a series of measurements are taken for blackbodies at different temperatures, there is always a fluctuation around the ideal calibration line, as shown in Fig. 10.14.

The best-calibrated spectra will be obtained by using the multiple point calibration, which is a linear regression that uses all the available points in order to more accurately characterize the system. If we

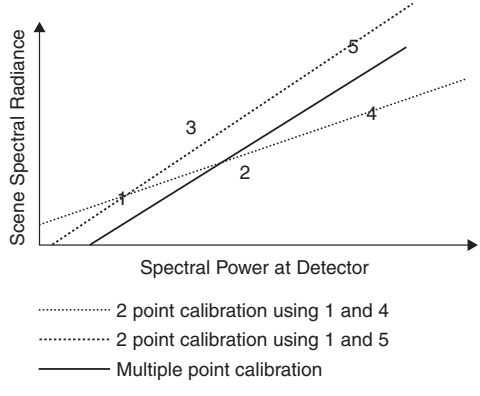


Figure 10.14 The multiple point linear relationship between the object view radiance and the detector response.

apply Eq. (10.17) to the measurements of multiple blackbodies, we get the following system of equations:

$$\begin{aligned} L_1(\sigma) &= S_1^{\text{Measured}}(\sigma) / K(\sigma) - M^{\text{Stray}}(\sigma) \\ L_2(\sigma) &= S_2^{\text{Measured}}(\sigma) / K(\sigma) - M^{\text{Stray}}(\sigma) \\ &\vdots \\ L_n(\sigma) &= S_n^{\text{Measured}}(\sigma) / K(\sigma) - M^{\text{Stray}}(\sigma) \end{aligned} \tag{10.43}$$

The preceding equations form a system of linear equations. The general solution to such a system for n equations is as follows:

$$K(\sigma) = \frac{n\sum_i S_i^{\text{Measured}}(\sigma)^2 - \left(\sum_i S_i^{\text{Measured}}(\sigma)\right)^2}{n\sum_i S_i^{\text{Measured}}(\sigma) L_i(\sigma) - \sum_i S_i^{\text{Measured}}(\sigma)\sum_i L_i(\sigma)} \tag{10.44}$$

$$M^{\text{Stray}}(\sigma) = \frac{n\sum_i S_i^{\text{Measured}}(\sigma)\sum_i S_i^{\text{Measured}}(\sigma) L_i(\sigma) - \left(\sum_i S_i^{\text{Measured}}(\sigma)\right)^2 \sum_i L_i(\sigma)}{n^2\sum_i S_i^{\text{Measured}}(\sigma)^2 - n\left(\sum_i S_i^{\text{Measured}}(\sigma)\right)^2} - \frac{\sum_i L_i(\sigma)}{n} \tag{10.45}$$

Eqs. (10.25) and (10.26) can be used to solve for systems of equations in the form of Eq. (10.24). In the case illustrated in Fig. 10.14, n equals 5 and i runs from 1 to 5. The calibrated spectrum is then given by Eq. (10.23) as before.

10.9.4 Nonlinear Multiple Point Calibrations

Nonlinear multiple point calibrations are used to correct for nonlinearity in the response function of the FT spectroradiometer, Fig. 10.15.

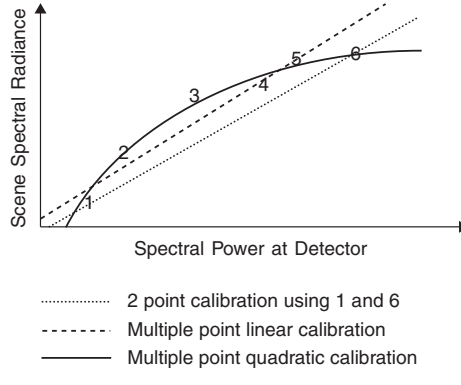


Figure 10.15 The nonlinear relationship between the object view radiance and the detector response.

The idea is to model the nonlinear response of the system by using a quadratic function instead of a linear function. The uncalibrated measurement is now expressed as:

$$S^{\text{Measured}}(\sigma) = Q(\sigma)(L^{\text{Source}}(\sigma) + M^{\text{Stray}}(\sigma))^2 + K(\sigma)(L^{\text{Source}}(\sigma) + M^{\text{Stray}}(\sigma)) \tag{10.46}$$

where $S^{\text{Measured}}(\sigma)$ = measured complex spectrum (arbitrary)
$Q(\sigma)$ = nonlinear instrument response function (arbitrary $\text{cm}^2\ \text{sr}\ \text{cm}^{-1}/\text{W}^2$)
$K(\sigma)$ = complex instrument response function (arbitrary $\text{cm}^2\ \text{sr}\ \text{cm}^{-1}/\text{W}$)
$L^{\text{Source}}(\sigma)$ = source spectral radiance ($\text{W}/\text{cm}^2\ \text{sr}\ \text{cm}^{-1}$)
$M^{\text{Stray}}(\sigma)$ = spectral power of the stray radiance ($\text{W}/\text{cm}^2\ \text{sr}\ \text{cm}^{-1}$)

The radiometric calibration is performed as before but there are now three unknowns per detector, the *nonlinear radiometric gain* $Q(\sigma)$, the *linear radiometric gain* $K(\sigma)$, and the *radiometric offset* $M^{\text{Stray}}(\sigma)$. This nonlinear *calibration* is then applied to an uncalibrated measurement to obtain a calibrated spectrum.

The characterization of the preceding system requires a minimum of three different temperature blackbody measurements. If we apply Eq. (10.27) to the three measured blackbodies, we get the following system of equations.

$$S_1^{\text{Measured}}(\sigma) = Q(\sigma)\left(L_1^{\text{Source}}(\sigma) + M^{\text{Stray}}(\sigma)\right)^2 + K(\sigma)\left(L_1^{\text{Source}}(\sigma) + M^{\text{Stray}}(\sigma)\right) \tag{10.47}$$

$$S_2^{\text{Measured}}(\sigma) = Q(\sigma)\left(L_2^{\text{Source}}(\sigma) + M^{\text{Stray}}(\sigma)\right)^2 + K(\sigma)\left(L_2^{\text{Source}}(\sigma) + M^{\text{Stray}}(\sigma)\right) \tag{10.48}$$

$$S_3^{\text{Measured}}(\sigma) = Q(\sigma)\left(L_3^{\text{Source}}(\sigma) + M^{\text{Stray}}(\sigma)\right)^2 + K(\sigma)\left(L_3^{\text{Source}}(\sigma) + M^{\text{Stray}}(\sigma)\right) \tag{10.49}$$

It is now possible to solve for the three unknowns $Q(\sigma)$, $K(\sigma)$, and $M^{\text{Stray}}(\sigma)$. The calibrated spectrum is then given by:

$$L^{\text{Source}}(\sigma) = \frac{(K(\sigma) - 2Q(\sigma)M^{\text{Stray}}(\sigma))}{2Q(\sigma)} \pm \frac{\sqrt{(2Q(\sigma)M^{\text{Stray}}(\sigma) + K(\sigma))^2 - 4Q(\sigma)(Q(\sigma)M^{\text{Stray}}(\sigma) + K(\sigma)M^{\text{Stray}}(\sigma) - S^{\text{Measured}}(\sigma))}}{2Q(\sigma)} \tag{10.50}$$

This algorithm can be extended to more than three points using statistical regression analysis on the polynomial much in the same way as was done for the linear case.

10.10 Radiometric Accuracy

Radiometric accuracy is the deviation of the measured spectral radiance from the actual object view radiance. This is illustrated schematically in Fig. 10.16.

However, even if noise[2] affects the radiometric accuracy, it is treated as a separate parameter, the NESR, which is not included in the radiometric accuracy.

In general, radiometric accuracy is an arbitrary function, as shown in Fig. 10.16, but is usually described in two parts, one absolute, the other relative. Radiometric error is absolute, if it does not vary with the object view radiance, or it is relative, if the error does vary with object view radiance. Absolute errors are more difficult to estimate than relative ones. In the following paragraphs, we will only discuss relative radiometric errors, converting absolute errors into relative equivalents. This is a more convenient way to predict the system accuracy for a given object view.

Three types of errors influence radiometric accuracy. The first type of *calibration source errors* are deviations in the production of a perfect (i.e., perfectly known) calibration source. These errors include the accuracy of blackbody temperature and the accuracy of its emissivity over the operational spectral range. The second type of error is *calibration drift*. Calibration drift includes everything that changes the radiometric gain and offset during the time interval between performing calibration measurements and object view measurements. Calibration drift is influenced by many factors, including the ambient temperature, the stability of the electrical gain, the stability of detector responsivity, and optomechanical stability. Finally, the third type of error that influences radiometric accuracy is the spectrometer *intrinsic linearity.* This includes system parameters such as detector linearity, channel spectrum, and spectral aliasing.

Because they are uncorrelated, the contributions from all predicted errors add up in a root-sum-square fashion.

10.10.1 Calibration Source Errors

Calibration source errors are misevaluations of the spectral radiance supplied by the calibration source. The effect of calibration source

[2]The distinction between noise and radiometric error is somewhat arbitrary. It is assumed that noise is the spectral-element-to-spectral-element uncorrelated intensity variations, which statistically average out with time. Radiometric errors stem from system imperfections and stay in the calibrated spectra even after all visible noise has been washed out.

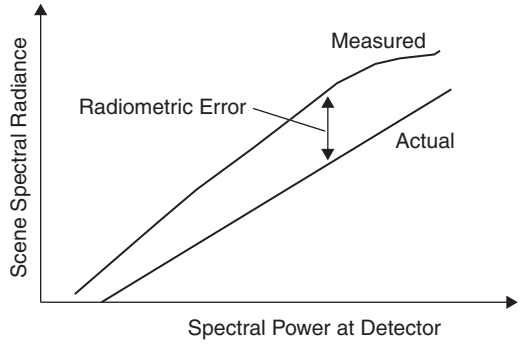


Figure 10.16 Actual and measured object view spectral radiance versus spectral power at the detector.

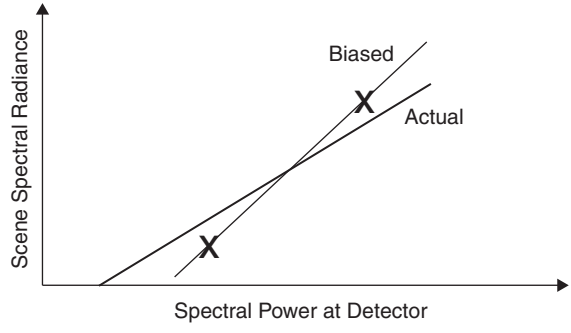


Figure 10.17 The effect of calibration source errors on radiometric calibration.

errors is illustrated in Fig. 10.17. The calibration points on the graph are misplaced and the resulting calibration line is thus skewed.

There are two types of calibration source errors. The first is a misevaluation of the radiance supplied by the calibration source. This is known as *blackbody error*. Calibration source emissivity error and temperature error are of this type.

Emissivity is a factor contributing to blackbody errors. To analyze the effect of underestimating or overestimating blackbody emissivity, it is interesting to examine Eqs. (10.21) and (10.22). When everything else is constant, the calculated radiometric gain is inversely proportional to the emissivity used for the blackbody. The calculated radiometric offset is, for its part, directly proportional to the emissivity used for the blackbody. These dependencies translate into a direct proportionality of a calibrated spectrum [Eq. (10.23)] on the emissivity used for the blackbody. In other words, the use of 1 percent-inflated emissivity values leads to 1 percent-overestimated

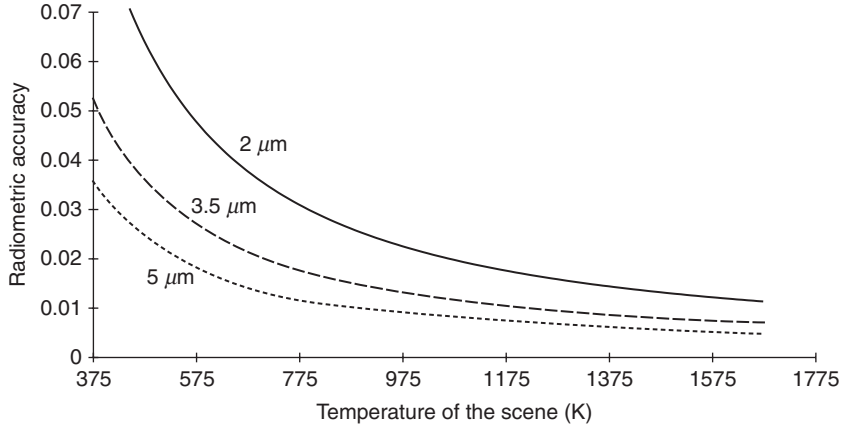


Figure 10.18 Relative radiometric errors at 5000 cm^{-1}, 2860 cm^{-1}, and 2000 cm^{-1} (2 μm, 3.5 μm, and 5 μm) due to the uncertainty of the calibration source temperature, assuming a blackbody relative temperature accuracy of 0.2 percent and an absolute accuracy of 1°.

calibrated spectra, and thus 1 percent radiometric errors. Good emissivity for available blackbodies in the 2000 to 2860 cm^{-1} (3.5 μm to 5 μm) spectral region is typically 0.98 ± 0.01. This is 1 percent emissivity accuracy.

The error contribution due to temperature uncertainty is also important. Estimating the error in this case is more difficult, however, since it involves three Planck functions [Eq. (10.20)], one evaluated at T_H (the temperature of the hot blackbody), one evaluated at T_C(the temperature of the cold blackbody), and one evaluated at T_S (the temperature of the object view). For a particular choice of T_S, it is possible to find a certain combination of T_H and T_C to minimize the error. In general, however, we can say that the relative radiometric error due to calibration source temperature uncertainty will always be less than the values displayed in Fig. 10.18, as long as $T_H > T_S > T_C$. In other words, the values displayed in Fig. 10.18 are upper-limit relative radiometric errors due to calibration source temperature uncertainty. Here, we have assumed a relative blackbody temperature accuracy of 0.2 percent and an absolute accuracy of 1°. The maximum error is 5 percent for the coldest source (T_S = 373 K) at the highest frequency (σ = 2860 cm^{-1}). This is a very pessimistic value.

10.10.2 Calibration Drift

Calibration drift is the variation of radiometric gain and offset, between the time they are acquired (the characterization) and the time they are applied (the object view measurement). This is illustrated schematically in Fig. 10.19.

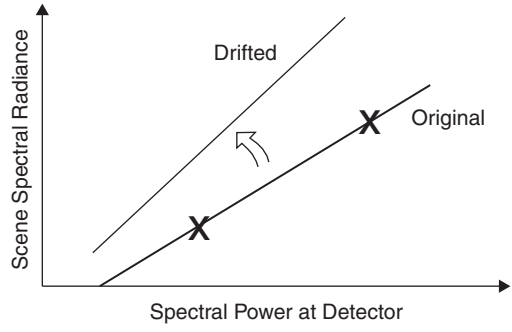


Figure 10.19 Relative radiometric error due to calibration drift.

We can write this relative radiometric accuracy, due to calibration drift, as:

$$RA^{\text{Drift}} \approx t\sqrt{\left(\frac{dG_E}{dt}\right)^2+\left(\frac{dR_D}{dt}\right)^2+\left(\frac{dG_O}{dt}\right)^2} \tag{10.51}$$

where t = time interval between the characterization and the object view measurement (s)
dG_E/dt = relative rate of change of the electronic gain of the system (Hz)
dR_D/dt = relative rate of change of the responsivity of the detector (Hz)
dG_O/dt = relative rate of change of the system response due to optical alignment (Hz)

This equation applies to a case where drift varies linearly with respect to time (i.e., it always goes in the same direction). After a long period of time, Eq. (10.32) will overestimate the radiometric error. The three rates of change cited earlier are governed principally by the change of temperature of the instrument.

Unlike other detectors, such as MCT units, InSb detectors exhibit much less variation of responsivity with respect to temperature. DR_D/dt is then negligible compared with other sources of calibration drift, whether the detector is cooled with a liquid-nitrogen pour-filled Dewar or with a sterling cooler. This assumes, however, that the temperature of the detector remains below a certain temperature limit of about 120 K. In a mobile environment, where all directions are permitted, special care must be taken to ensure that none of the nitrogen is spilled if the Dewar is turned upside down.

dG_E/dt is affected by temperature drift as well as by electromagnetic disturbances, but these can be kept at a minimum through

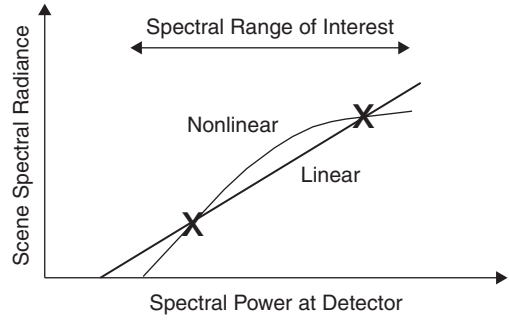


Figure 10.20 Relative radiometric error due to intrinsic linearity errors.

careful design and the use of high-quality components. DG_O/dt is very difficult to estimate since it is influenced by a great number of design parameters. The ultimate test is, of course, to actually measure the drift experimentally.

One way to estimate the radiometric accuracy budget assigned to calibration drift is to examine the typical temporal drift of the SpectRx 100, the FTIR spectrometer on which the design of the SpectRx system is based. The relative drift of the system response is 2 percent over a 16-hour period.[3] These results have been obtained with a system subjected to typical room-temperature fluctuations. However, it is probably possible to recalibrate more often than every 16 hours using a calibration source or other calibration references. For the SpectRx system, we typically use 2 percent as the relative radiometric accuracy due to drift, and assume that the corresponding time interval between characterization and object view measurement is long enough to allow measurement.

10.10.3 Intrinsic Linearity

The intrinsic linearity of the spectroradiometer pertains to errors other than calibration source errors and calibration drift errors. Typical nonlinearity is illustrated in Fig. 10.20. In this case, the measured spectral radiance from the object view is accurate at the two calibration points and deviates from the true values when we move away from these points. For this type of error, it is advantageous to choose a calibration source temperature, so the spectral radiance of the calibration measurements is included in the range of the spectral radiance to be measured. The exact location depends on the actual nonlinear response.

[3]A good part of this drift is believed to be due to the variation of the internal source.

10.10.4 Detector Nonlinearity Errors

Nonlinearity is usually thought to be caused by detector responsivity nonlinearity. Photoconductive MCT detectors are known for their nonlinear response. On the other hand, photovoltaic InSb detectors, are usually operated in a regime where they are extremely linear. They start to saturate very quickly at an illumination level corresponding to a photogenerated current of about 200 μA per detector mm^2. In the SpectRx systems, the maximum state of saturation level is not reached because the feedback resistors used are high enough so that preamplifier saturation occurs before there is detector saturation. We therefore do not anticipate detector nonlinearity problems in InSb detectors.

10.10.5 Channel Spectrum Error

Channel spectrum is spectral pollution that leads to a radiometric error. It manifests itself as a series of interference fringes appearing on top of an uncalibrated spectrum. Channel spectrum occurs when transmissive optical components with near-parallel flat faces are used. Interference fringes arise from constructive and destructive interference of internal reflected waves in the optical components.

Channel spectrum generates an oscillatory radiometric error of period $1/2tn$, where t is the plate thickness in centimeters and n is the index of refraction proportional to the spectral radiance. The relative radiometric accuracy due to channel spectrum is given by:

$$RA^{Channel} = 2R\,M_S M_T \tag{10.52}$$

where R = single plate reflectivity of the plate
$M_s \le 1$ = attenuation factor caused by the shearing of wave-fronts
$M_t \le 1$ = attenuation factor caused by the tilting of wave-fronts

Shearing wave-fronts with respect to one another attenuates the interference. The shearing can be achieved by tilting the optical component. By doing so, the main beam will shift by a certain amount, which has to be compensated for by proper optomechanical realignment. As illustrated in Fig. 10.21, the secondary beam will shift from the main one by a quantity δ given in cm by:

$$\delta = 2t\,\mathrm{Cos}\phi\,\mathrm{Tan}\left[\mathrm{Arc\,Sin}\left(\frac{\mathrm{Sin}\phi}{n}\right)\right] \tag{10.53}$$

where t = thickness of the plate (cm)
ϕ = angle of incidence of the incident beam (radians)
n = index of refraction of the plate

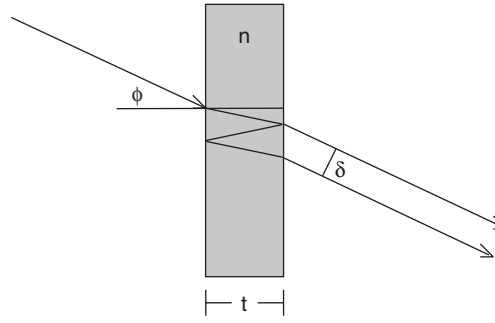


Figure 10.21 Shearing of the secondary beam in a tilted plate with parallel faces.

The attenuation factor caused by the shear is:

$$M_S = \frac{\mathrm{Sin}(2\pi\sigma\theta\delta)}{2\pi\sigma\theta\delta} \tag{10.54}$$

where σ = wave-number of interest (cm^{-1})
θ = beam divergence half angle (radians)
δ = shear given by Eq. (10.34) (cm)

Parallel-face protective windows are included in SpectRx systems right after the telescope and before the output optics. The level of channel spectrum generated by such a window is calculated using Eq. (10.33) with $M_T = 0$, and Eqs. (10.34) and (10.35). A certain degree of wedging of this window would further reduce the channel spectrum but at the cost of introducing the phenomenon of *ghost images*. Ghost images are generated when reflections from the wedge's surfaces (of an off-axis source, which can be many times more intense than the source of interest (on-axis)), reach the detector and give an erroneous signal. This is illustrated schematically in Fig. 10.22.

Ghost images are especially destructive and must be avoided if possible.

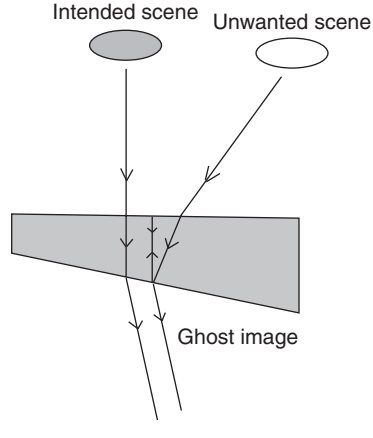


Figure 10.22 Ghost image phenomenon encountered with a wedged substrate.

10.10.6 Spectral Aliasing Error

Spectral aliasing is the folding over of all signal and noise within the bandpass of the system into the spectral free range, which spans from 0 to the Nyquist frequency (half the sampling frequency). To avoid the overlapping of spectral features, it is important that any spectral information outside the spectral range of interest be removed. This can be done electrically with a low-pass filter, but this method does not provide a sharp cut-off over the upper limit of the frequency range or attenuation below the lower limit. A more appropriate method is to use an optical filter, which provides both a sharp cut-off and very high attenuation outside the spectral range of interest. This method also reduces photon noise.

Further Reading

1. B. K. Yap, W.A.M. Blumberg, and R.E. Murphy, Applied Optics, Vol. 21, 4176 (1982).
2. J. B. Johnson, *Nature* 119, 50 (1927).
3. J. B. Johnson, *Phys. Rev.* 32, 97 (1928).
4. D. R. White et al., *Metrologia*, Vol. 33, pages 325–335 (1996).
5. D. D. Laporte, William L. Smith, and L.A. Sromosky, *Applied Optics*, Vol. 27, 3210 (1988).
6. H. Nyquist, *Phys. Rev.* 32, 110 (1928).
7. Hugh Lippincott, Notes on Johnson Noise, given on 19 February (2007).
8. L. B. Kish, Phys. Lett. A 352, 178–182 (2006) /physics/0509136; A Cho. *Science* 309, 5744, and 2148 (2005).
9. Peter R. Griffiths and James A. de Haseth, *Fourier Transform Infrared Spectrometry*, Wiley-Inter-science.
10. Sabrie Soloman, *Sensors Handbook*, McGraw-Hill Publishing Company, New York, 2009.
11. Sabrie Soloman, *Sensors and Control Systems in Manufacturing*, McGraw-Hill Publishing Company, New York, 2009.
12. Sabrie Soloman, Non-destructive Identification of Tablet and Tablet Dissolution by Means of Infrared Spectroscopy. U.S. Patent Number 5,679,954.
13. Sabrie Soloman, Real Time—On Line Analysis of Organic Compounds for Food, Fertilizers, and Pharmaceuticals Products. U.S. Patent Number 5,900,634.
14. Sabrie Soloman, Sabrie's Index of NIR-Light Energy, Sensors Symposia, Seoul, Korea, 2007.

CHAPTER 11

Economic and Social Interests in the Workplace

This chapter was written most especially for those who possess the talents of innovation but need economic know-how to implement their dreams. It is written for all my students, peers, and those entrepreneurs who wish to grasp the simple yet practical understanding needed to build their new manufacturing business with a sound financial strength.

11.1 Manufacturing Operation Control Through Financial Planning

"Engineering in the absence of economics is a degenerate science."

The small manufacturing firm, just as surely as the larger manufacturer, must have yardsticks for measuring the results of every activity of its business operation. For this purpose, management should take the following steps:

1. Define its objectives precisely.
2. Prepare a written statement of the steps to be taken to accomplish these objectives.
3. Refer continually to its statements of objectives to see that its operation is proceeding according to schedule.

For some manufacturers, the budget may be simply an informal sales forecast, production schedule, and profit forecast. If only one person is responsible for seeing that activities proceed in accordance with plans, the budget becomes a simple basis for action by a top executive. It helps an executive make on-the-spot decisions that are in

line with an overall plan. It charts the course a top manager should take in carrying responsibility for many different functions.

If the business grows to the point where selling, producing, and perhaps financing responsibilities are delegated to different persons, a more specific and formal plan becomes important. The budget must still direct the decisions of the owner-manager and key executive of the firm, but it must also coordinate the actions of the different members of the group.

11.1.1 Developing a Plan

A manufacturer's budgeting procedure must be simple. It must be focused on a limited number of functions that are important to the special circumstances. All individuals who are to have any responsibility for seeing that it works should take part in planning it. The overall budget for most manufacturers should include the following parts:

- Sales budget
- Production budget
- Selling and administrative budget
- Capital budget
- Financial budget

The first four of these budgets taken together provide a basis for projecting profits, since the sales budget provides an estimate of revenues and the other three of costs. The financial budget provides a plan for financing the first four plans. Each of the five budgets should be made in advance for periods of six months or a year.

11.1.2 Sales Budget

The accuracy of any budget or plan for profitable activity depends first on the accuracy of the sales forecast. In many small firms, the volume of sales may be affected substantially by the condition of a single new customer or the loss of a single old one. The sales forecast in these cases must include a good deal of guesswork.

Nevertheless, a forecast should be made. It will usually be based mostly on past sales figures, but sales tendencies and economic conditions that are likely to increase or decrease sales volume in the coming period should be studied.

A sales forecast or budget that might be suitable for a typical small plant appears in Table 11.1. Forecasted sales are stated in both product units and dollars and are broken down by sales territories.

11.1.3 Product Budget

Once the sales forecast has been made, the production budget can be prepared. Enough units of products 1 and 2 must be produced to

	Product 1		Product 2		
Sales territory	**Product units**	**Dollar sales**	**Products units**	**Dollar sales**	**Total dollar sales**
A	1850	14,800	2200	48,400	63,200
B	575	4,600	500	11,000	15,600
C	1400	11,200	1000	22,000	33,200
Total	3825	$30,600	3700	$81,400	$112,000

Table 11.1 Sales Forecast

supply the estimated sales volume and maintain reasonable stock levels. In this example, it has been decided that about a two-month supply should be kept in stock. If the inventories at the beginning of the period are below that level, production will have to exceed the sales forecast. Table 11.2 shows how the quantity to be produced can be figured.

On the basis of the number of production units required, budgets or standards for material, labor, and overhead costs are prepared. These production costs should be figured in detail so that cash requirements, material purchasing schedules, and labor requirements can be set up. Detailed plans are then made for monthly production levels.

11.1.4 Selling and Administrative Budget

The sales volume of a small plant usually does not provide a large enough margin over production cost to overcome high selling and administrative costs. These nonmanufacturing overhead items must be watched constantly to see that the commitment to fixed costs are kept low.

This can be done best by using a selling and administrative budget. All cost items expected in these areas should be listed and classified according to their fixed and variable tendencies. Account

	Product 1	Product 2
Sales forecast	3825	3700
Desired inventories ($^1/_6$ annual sales)	637	617
Product units required	4462	4317
Less: beginning inventories	200	400
Total production required	4262	3917

Table 11.2 Required Production Based on Sales Forecast

classifications in the accounting system should correspond to the classifications used in the budget.

At the end of each month, the fixed, variable, and total selling and administrative costs budgeted for the expected level of operation should be adjusted for the level actually reached during that month Actual costs should be matched against these adjusted figures and any sizable differences investigated. If possible, action should be taken to prevent repetition of any unfavorable variances.

11.1.5 Capital Budget

Plans for the acquisition of new buildings, machinery, or equipment should be developed annually. A list of amounts to be spent at specified times during the year will provide information for use in a firm's financial budget, as well as cost and depreciation figures to be shown on its projected income statement and balance sheet.

11.1.6 Financial Budget

Any business owner must be certain that funds are available when needed for plant or equipment replacements or additions. In addition, enough working capital must be available to take care of current needs. This requires a financial plan or budget.

Funds for major replacement or addition can ordinarily be planned for on the basis of each expenditure individually. Management must know in each case whether surplus working capital will be available or new long-term financing will be required. If new financing is the answer, a source of funds must be found. Plans for a loan or an additional investment by owners should be made well before the time when the funds will be needed.

To have enough working capital at all times without a large oversupply at any time requires detailed planning.

Planning the receipts and expenditures for each month must take into consideration expected levels of activity, expected turnover of accounts receivable and accounts payable, seasonal tendencies, and any other tendencies or circumstances that might affect the situation at any time during the budget period. Relating the inflow plus beginning balance to the outflow for each month shows whether a need for outside funds is likely to arise during the budgeted period.

11.1.7 Income Statement and Balance Sheet Projections

From information provided by the four basic budget summaries—sales, production, selling and administrative, and financial—estimated financial statements can be drawn up. These projected statements bring the details together. They serve to check the expected results of operations and the estimated financial position of the business at the end of the budget period.

11.1.8 Planning for Profit

Projection for income statement and balance sheet figures will indicate the profit, return on investment (ROI), and return on assets (ROA) management can expect. For example, these statements in highly abbreviated form may be as follows:

Net profit before interest and taxes	$430,000
Interest cost	80,000
Net profit after interest	350,000
Income taxes (40%)	140,000
Net profit	210,000
Debt	1,000,000
Capital or stock investment	2,000,000
Assets	$3,000,000

The return of investment (i.e., the profit earned in relation to the value of the capital required to produce the profit), is expected to be

$$\$210,000/\$2,000,000 = 10.5\%$$

Further, the return on assets is expected to be:

$$\$258,000/\$3,000,000 = 8.6\%$$

Note: In this last calculation, the return is calculated after taxes and before the deduction of interest costs. The $258,000 is determined either by adding the after-tax interest cost of 60 percent of $80,000, or $48,000, to the net profit, $210,000, or by subtracting the 40 percent tax on the net profit before interest or taxes of $430,000 from this amount. The 8.6 percent return on assets is increased to a 10.5 percent return on investment as the result of the leverage provided by borrowed funds at an interest of less than 8.6 percent.

These ROI and ROA figures may or may not be considered adequate. If they are thought to be inadequate, various parts of the budget should be reviewed to identify possible areas of improvement. In any case, the profit goal should be high enough to provide a reasonable return and at the same time be realistically attainable.

11.1.9 Controlling Operation

Budgeting and profit planning are necessary for control of operations, but the forecasts and budgets do not by themselves control operating events. A budget is useless without comparisons with actual operating results, and the comparisons are useless unless the manager takes action on deviations from the budget.

Suppose, for example, that the sales of product 1 in territory A are budgeted at 155 units per month (1850 divided by 12), and actual sales for January and February are 120 and 110 units. (Assume that sales of product 1 in territories B and C are in line with the budget.) On investigating, the manager finds that sales are being lost to a new competitor in territory A.

Actions the manager might take include:

- Putting more money into the selling effort and try to win back the old customers
- Lowering the price of product 1 to obtain more sales
- Recognizing that the sales are lost, and so lowering the production budget accordingly

Each of these actions will have an impact on the budget-planning operations of the company. An increase in the selling effort will affect the selling and administrative budget and the financial budget—more funds will be needed. Lowering the selling price, if the unit sales level originally forecast is just regained, will result in lower cash receipts than in the original financial budget. A decision to reduce production to avoid piling up inventories will affect purchase orders, labor costs, variable overhead costs, income taxes, and other budgeted items. The actions have been signaled by a variation from the sales forecast, but a change in plans affects all the budgets and forecasts.

Any sizable variance from budgeted figures, favorable or unfavorable, should be looked into. Management should find out why the variance resulted so that when it takes action, it will be reasonably sure it is acting wisely. For example, a favorable variance in direct labor costs might be due to a special incentive program. In such a case, management should put off any plans to use the surplus funds until it is sure that a reduction in labor costs will be permanent.

Unfavorable variances from the budget should not be studied separately from related budget items. For example, suppose the actual selling expense is much higher than the amount budgeted. It may turn out that the extra expense produced a large volume of new orders. In that case, the added cost may be entirely justifiable. Hasty action to reduce the expense without investigation might result in loss of sales and a net reduction in profits.

Great care must be given to seeking out the causes of variances, and the consequences of possible control action should be analyzed carefully in order to avoid unintended and unwanted results.

11.1.10 Cost and Profit Analysis

Costs are classified as variable (product costs) and fixed (period costs). With these classifications, management can determine which of its products contributes the most and which the least to covering

fixed cost and providing a profit. Direct costing provides information for decisions to expand, reduce, or continue production of a given product line or product.

In a standard-cost system, product costs for materials, labor, and overhead are budgeted or planned before the actual production takes place. The pre-established costs then provide a point of reference for analyzing the costs of actual production. By analyzing the variances from these standards, the manufacturer can identify the sources of excessive cost and investigate the causes. This kind of process—the identification of a problem area and the discovery of its cause—can point the way to actions that will reduce costs and improve profit.

11.2 Information Assimilation and Decision Making

Similarly, a business manager should plan ahead for all the activities of the business—that is, the manager should develop a set of budgets for sales, production, selling and administrative expense, and financial requirements. As indicated in the standard-cost system, the actual results of various activities can be compared to the budgeted or planned results. Analysis of the variances found will then help the manager make reasonable and thoughtful decisions.

These analytical methods—direct cost, standard-cost variance analysis, and budget planning and control—are ways of putting cost-accounting information to good use. The manufacturer can make decisions on the basis of information provided by these analytical tools. There are, however, several other important ways of analyzing costs and profits. Two of these—break-even analysis and incremental analysis—are discussed in the following sections.

11.2.1 Breakeven Analysis

Sales forecasts are uncertain at best. For this reason, it is important for management to know approximately what cost changes can be expected to go along with volume changes. It must know what levels of production and sales volume are necessary for profitable operation. It must know how much effort and cost are justified to keep volume at a high level. Break-even analysis provides this sort of information.

Break-even analysis is so called because the focal point of the analysis is the break-even point—the level of sales volume at which revenues and costs are just equal. At this level, there is neither profit nor loss.

11.2.2 Breakeven Chart and Formula

A break-even chart is illustrated in Fig. 11.1. A break-even chart can be prepared from budget figures and knowledge of capacity levels.

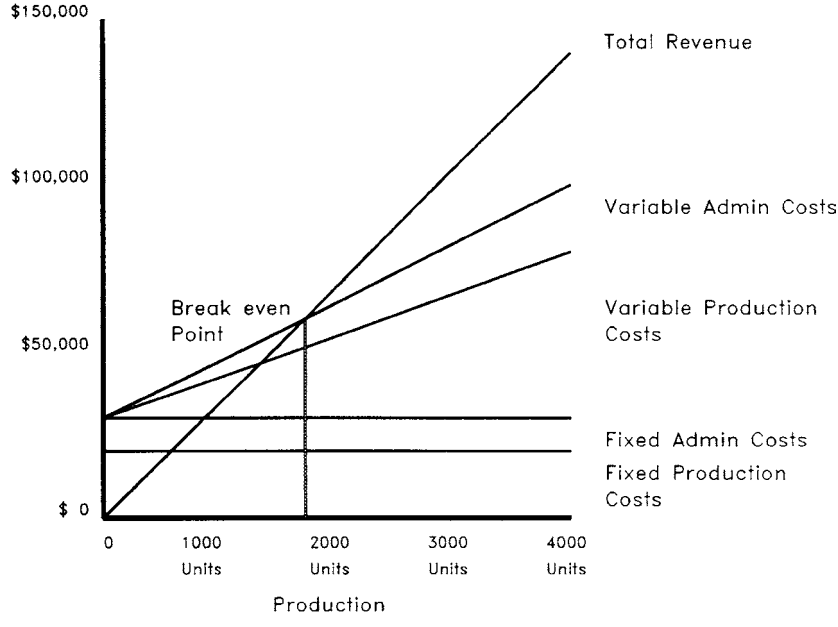


Figure 11.1 A break-even chart.

In Fig. 11.1, it is assumed that fixed costs for production are $20,000 and for selling and administrative activities, $10,000. Variable production costs are $12.50 per unit, or a total of $50,000 at the maximum capacity of 4000 units. Variable selling and administrative costs are $5.00 per unit, or $20,000 at the 4000 level. This gives minimum costs of $30,000 (the total fixed costs) at zero production and maximum total costs of $100,000 at 4000 units. The selling price is $35 per unit.

A total cost line is drawn from the $30,000 cost level for no production to the $100,000 level for maximum production and sales. A total revenue line is drawn from the zero line for no revenues to $140,000 for the maximum sales of 4000 units. The point where these two lines cross is the break-even point—the level of operation at which there will be neither profit nor loss.

The area enclosed by the two lines below the break-even point represents loss, and the enclosed area above the break-even point represents profit, Fig. 11.1.

Figure 11.1 shows that the volume at which the business can be expected to break even is slightly below one-half the maximum point, or between 1500 and 2000 product units. The exact level can be calculated as follows:

$$\frac{\text{Fixed costs}}{\text{Selling price} - \text{variable costs/units}} = \frac{\$30{,}000}{\$17.50} = 1714.29\,\text{units}$$

Verification	
Gross revenue (1714.29 units at $35)	$60,000
Less: variable costs (1714.29 units at $17.5)	30,000
Contribution margin	30,000
Less: fixed costs	30,000
Net profit or loss	$0

11.2.3 Utilization of Breakeven Analysis

A break-even chart can give approximate answers to many questions. For example, if the variable production costs are expected to increase 10 percent in the coming year without any change in the selling price of the product, the total cost line will be drawn steeper, and a new higher break-even point will result. This will show what increase in sales volume is needed to offset the increased costs.

Or suppose a change in selling price of the finished product is being considered. The total revenue line will now be steeper for an increase or less steep for a decrease in selling price. The chart will then show the effect on profit if the sales volume remains the same or if, for example, it drops—as it might if the price is increased.

Suppose the manufacturing firm whose figures are used in Fig. 11.1 is producing and selling 2000 units, thus making a profit of $5000. Now suppose it wants to give its employees a general wage increase. As planned, the increase has the effect of adding $1 to the variable costs of each unit (thereby decreasing the contribution by the same amount) and $3000 to fixed costs. The owner wants answers to these questions:

How many units must the owner produce and sell at $35 each to realize the same profit—$5000? Can the owner produce this volume without investing any more in the plant and equipment?

The required units are calculated by using the following formula. (The calculation can be verified as described in the preceding formula.)

$$\frac{\text{Fixed costs + decreased profit}}{\text{Selling price} - \text{variable costs/unit}} = \frac{\$33{,}000 + \$5000}{\$16.50} = 2303 \text{ units}$$

Thus, the manufacturing firm will have to produce and sell 2303 product units instead of 2000 in order to realize the same $5000 profit. Since the plant capacity is 4000 units, management can increase production to 2303 units without any further investment in plant and equipment.

Break-even charts give quick approximate answers. They should not take the place of detailed calculations of the results of anticipated changes, but they do encourage careful consideration of the effects of any increases in either fixed or variable costs. They can also be a constant reminder of the importance of maintaining a high sales volume.

11.2.4 Incremental Analysis

Incremental costs and revenues are those that change with increases or decreases in the production level. Fixed costs as well as variable costs may change under certain conditions. Incremental analysis is an examination of the changes in the costs and revenues related to some proposed course of action, some decision that will alter the production level or change production activity. Two kinds of incremental analyses are explained here.

11.2.5 The Large New Production Order

Sometimes it happens that a manufacturer has an opportunity to land a new customer and a very large order for goods. Usually, such a large order cannot be landed without adding new production employees and supervisors and increasing overhead costs. The new order thus creates a decision-making problem in which basic information differs somewhat from the current cost data. A study of the incremental costs and revenues—those that arise from the new contract—provides information on which the manufacturer can base the decisions to accept or reject the new order.

Suppose that a manufacturer currently produces 10,000 product units a year. Fixed costs are $20,000 and variable costs, $5 per unit—$2 for materials, $1 for labor, $1 for variable production overhead, and $1 for variable selling expenses. The unit selling price is $10. A new customer wishes to buy 10,000 units a year at $9 per unit. A new night shift would be required to double the plant's production as required by the new customer's order.

Costs at the original level of production (10,000 units) are not necessarily accurate for the new level (20,000 units). Assume that the costs of producing the 10,000 additional units required by the new order are estimated as follows:

Variable costs per unit	
Materials	$2.00
Labor (including night shift premium)	2.00
Variable production overhead	1.25
Variable selling costs	0
Total variable costs	$5.25
Incremental fixed costs	
Additional supervisors	$10,000
Power, light, heat	5,000
Office expenses	3,000
Total incremental fixed costs	$18,000

	Original production	New order	Total
Gross revenues	$100,000	$90,000	$190,000
Variable costs	50,000	52,000	102,000
Contribution margin	50,000	37,000	87,000
Fixed costs	20,000	18,000	38,000
Profit before taxes	$30,000	$19,500	$49,500
Return on sales	30.0%	21.7%	26.1%

TABLE 11.3 Financial Results

Note: For accuracy, it is important to use current or projected prices of materials, labor, and overhead in this type of incremental analysis.

Average prices or allocated costs derived from historical cost accounting records (for example, material costs based on the first-in, first-out method) may not produce current figures that can be relied on in making a decision.

The financial results of accepting the new contract would therefore be as shown in Table 11.3.

The incremental analysis shows that the new contract will be profitable, but the rate of return will not be as high as for the original level of production. The computation would have given different results if the original variable costs of $5 per unit and the fixed cost of $20,000 had been used. Incremental analysis makes it possible for a manufacturer to examine the probable results of making major changes in the production level and in the arrangement and scheduling of production activity.

11.2.6 Make versus Buy

Most manufacturing firms buy some parts from outside suppliers for use in assembling the finished products. Often, the manufacturing firm could make the part in its own plant instead of buying it from an outside source. The question is whether the cost of making the product itself would be less than the cost of purchasing it from an outside supplier. An analysis of these two alternatives is known as a *make-or-buy analysis*.

Take, for example, a manufacturing firm that now purchases 10,000 small electric motors a year at $4.80 each after deducting quantity discounts. The question it faces is this: Can it make the motor itself for $4.80 or less?

Assume that the costs of producing 10,000 motors a year in its own plant would be as follows:

Variable costs	
Materials per unit	$2.00
Overhead costs per unit for the new activity (power, suppliers, etc.)	0.75
Total variable cost per unit	$2.75
Total variable cost per 10,000 units	$27,500
Fixed costs	
Salaries of 3 technicians (capable of producing 15,000 units a year)	$21,000
Depreciation on special machinery ($40,000 spread over a 5-year life)	8,000
Total fixed costs for the new activity	$29,000
Total costs	
Total cost of producing 10,000 motors in plant	$56,500
Total cost per unit	$5.65

Thus, with a production level of 10,000 units a year, the manufacturing firm would be better off buying the motor from a supplier at $4.80 each. However, with the three technicians and the $40,000 in special equipment, it has the capacity to produce 15,000 motors a year. At this level, its total cost would be $70,250 and its unit cost $4.68. It would save at least $0.12 per unit, or $8750 over the five-year useful life of the special equipment. So the decision to make or buy the motor would depend on the volume needed.

On the other hand, the owner might decide that the $1750 a year would not be enough to offset the extra burden that policing the motor activity would place on key people. This assumes, of course, that the present supplier delivers quality motors on time.

Another analysis is also involved in this decision, however—the determination of whether the $8750 savings is enough to justify investing $40,000 in special equipment, or whether the money could be invested to better advantage elsewhere. This decision is known as a *capital budgeting decision*.

11.2.7 The Importance of the Basic Data

The techniques of cost and profit analysis discussed earlier are intended as guides. The validity of the method depends largely on the validity of the basic data—the fixed and variable cost classifications, cost estimates, and volume estimates. Errors in any one of these factors will cause mistakes in the computations and may result in erroneous decisions.

Every manufacturing firm should continually review the way its costs are classified and accumulated, and it should be very careful in

making cost estimates. Its general accounting system and cost-accounting system should be reviewed periodically to ensure that basic cost data are collected and classified according to an established accounting plan.

11.3 Communication

In a business organization, gathering and using information about the operation of the business may be a fairly simple matter. A purchasing agent, production manager, or sales manager who has only a few employees can observe day-to-day activities and exercise direct supervision and control. When a business becomes larger, however—when more employees and several levels of supervision are necessary—the processes of obtaining information and directing the business become more complex.

The larger business requires a communication system for collecting information about operations and distributing it to managers in the company. Such a system includes four processes:

- Accumulating
- Analyzing
- Reporting
- Communicating

Thus far, discussion has emphasized techniques for accumulating, analyzing, and reporting cost-accounting data. Attention must be given to methods of communicating the information to those in the firm who need it in order to accomplish their tasks.

Communication involves both a sender and a receiver of information. A manager sets up an accounting system as a sender. The manager assures that a system produces certain reports and analyses of the basic counting data.

If the information is to be useful, the needs of those who will receive it must be known and considered. Thus, communication depends on a two-way exchange between the senders and the receivers. Figure 11.2 illustrates a communication network in a business organization; the arrows show the two-way flows of information between senders and receivers.

Communication also requires that the information sent and received be useful. To be useful, information must meet the following standards:

- It must be suited to the needs of the receivers in form and content.
- It must be free from clerical errors, incorrect classification, and unrealistic estimates.

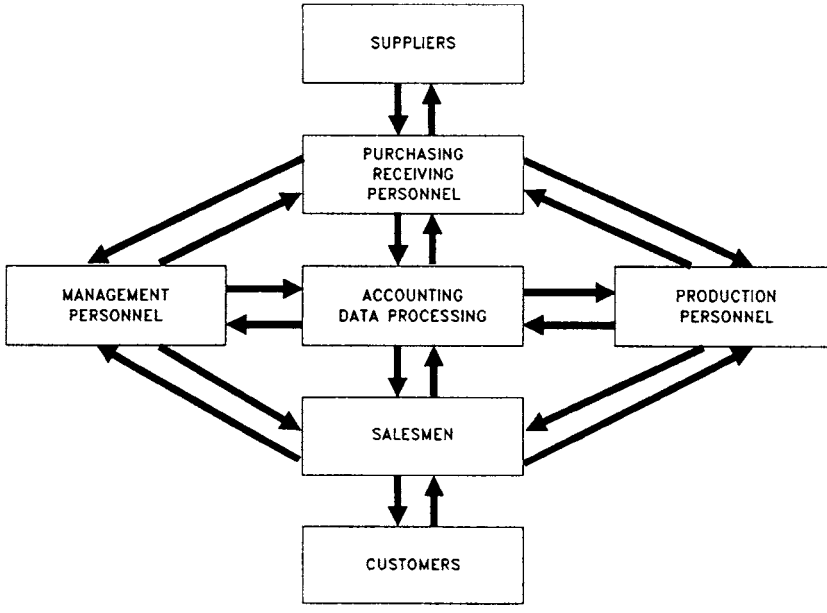


Figure 11.2 Communication network in a business organization.

- It must be presented in clear language and not buried under a mass of unnecessary details.
- It must be received soon after the actual events, in time for suitable action to be taken.

11.3.1 The Need for Timeliness

There is bound to be some lag from the time an activity is completed to the time a report of that activity can be made available. It takes time for the clerical staff to sort and total materials requisitions and time tickets for posting to job-cost sheets or for use in cost-system calculations. It takes time to prepare data for electronic data processing (EDP), deliver it to a service bureau, and obtain the printed output. If a report is delayed too long, it may be useless by the time it finally reaches the people who need it.

Each manufacturing firm must decide in which areas of operation quick reports are needed. For example, if a machining department is overloaded with work, this should be known promptly so that workers can be shifted to prevent a production bottleneck. Or perhaps work can be routed to another department where machine time is available. For this kind of scheduling, daily or even hourly performance may be needed. Activity reports must be communicated quickly where the situation is likely to change often.

Other situations may require reports on different schedules. Since labor efficiency usually changes slowly, reports on labor hours and labor costs may be prepared weekly or monthly. Material-price variance reports may be prepared at the end of a periodic buying season or after the second week in a month when vendors' invoices are received. Materials usage and efficiency reports may be prepared daily, weekly, or monthly according to the characteristics of the manufacturing process and the needs of the production manager.

In general, the amount of clerical time, effort, and cost put into a communication system is related to the importance of the activities being reported. If the manufacturing process uses a great deal of raw materials and parts, automated equipment, and a minimum of labor time, the accounting system should produce frequent reports on material prices and usage and on equipment maintenance and repair costs. Labor costs can be reported less often because they are not a large part of total production cost.

On the other hand, if labor costs do make up a large proportion of the total costs, the communication system should emphasize labor efficiency and cost. Materials cost can be reported less frequently. The manufacturing firm must decide how much it can afford to spend on its communication system and how valuable the information is to its managers.

11.3.2 The Reporting Strategy

It takes time to accumulate, calculate, and record cost data. Therefore, cost information is generally communicated in monthly, quarterly, and annual reports. The reports usually cannot be delivered in time to be useful on an hourly, daily, or weekly basis. Consequently, the more frequently issued reports contain only physical data—material units, labor units, machine hours, number of employees, machine downtime, work-in-process backlog, material spoilage, and so on.

These reports in terms of physical units (without dollar figures) are useful for managers, who can take direct action to affect use of materials, labor hours, and certain types of overhead spending. Cost reports in dollar terms are more useful for managers who can take action that affects prices, wage rates, overhead spending, and so on.

The form in which information is communicated can also be important. It varies according to the circumstances and the one who is to receive it. Sometimes an oral report is enough to get the necessary action started, but more often reports are written or printed. Reports issued frequently may be handwritten on standard forms in order to get them out quickly. Other reports may be typed or may be printed out by EDP equipment.

In some cases, visual displays, in the form of large graphs or drawings are used. For instance, displays showing employee output,

employee safety performance, and other information of interest to workers can be shown in plant areas. These displays when projected with the name and picture of an employee help keep the employees informed and interested in their job performance.

All but the smallest of manufacturing businesses are usually organized so that different managerial duties are assigned to different persons. Along with each manager's authority goes a responsibility for managing the segment of business efficiently. Figure 11.3 shows a simple organization chart for a typical manufacturing plant.

In responsibility reporting, information is reported on the basis of areas and responsibility. For example, reports containing information useful to foremen for performing their jobs will be directed to departmental foremen.

This means that the responsibility accounting and reporting system will be only as detailed as the company's organization plan. In a plant with the organization shown in Fig. 11.3, reports could be prepared for the foremen, the supervisors, the purchasing agent, the production manager, the sales manager, the salespeople, and the president. A company that has other organization units must work these into the plan if those units are to be served by responsibility-oriented reports.

A responsibility-oriented accounting system should generate reports that provide information on which each manager can base decisions. The system should also provide reports showing the results of these decisions.

11.3.3 Responsibility Centers

The responsibility centers may be cost centers, revenue (sales) centers, or profit centers. In Fig.11.3, each producing department and each of the two service departments (purchasing and accounting) can be considered a cost center if the costs of operations are analyzed along those lines. Each product can be considered a profit center, with the responsibility held jointly by the sales manager (of revenues) and the product supervisors (for product costs). The sales office can be considered a revenue center that produces the gross sales revenues of the company.

11.3.4 Controllable and Uncontrollable Costs

Costs and revenues are reported to managers according to their controllable and uncontrollable characteristics. A cost item is controllable by a manager if action directly affects the amount of cost incurred.

Assume, for example, that the product managers have sole authority to purchase equipment, thus incurring depreciation costs. The depreciation cost is then controllable at the manager's level but uncontrollable at the department foreman level. Consequently, the department foremen cannot be held responsible for depreciation

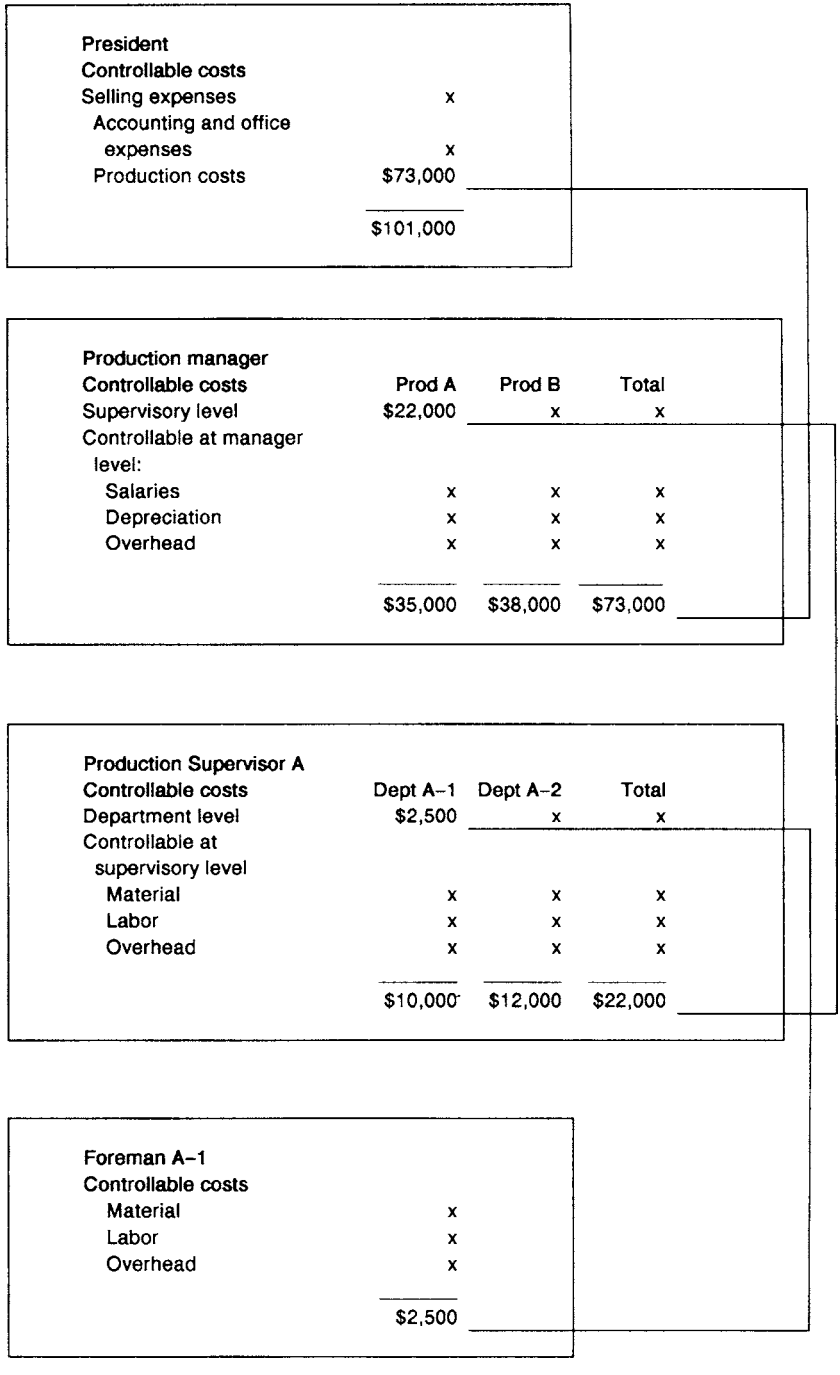


Figure 11.3 A simple organization chart for a manufacturing plant.

charges, and that cost is not reported as part of the responsibility. The principle of controllable-uncontrollable classification should be treated consistently at all management levels.

An illustration of responsibility cost reporting and the pyramiding construction of reports is shown in Fig. 11.3.

11.3.5 Construction of the Report

Each report delivered to a manager should be constructed according to the usefulness to the manager of information provided by the accounting system. The actual costs can be compared with the budgeted figures, standard-cost variances can be shown, and unit-cost calculations can be made. Any presentation of the information that promotes for sound decision making is suitable.

The classification of costs as controllable or uncontrollable can be difficult. One of the most troublesome problems is obtaining the agreement and cooperation of the foremen, supervisors, and managers who will be assigned certain responsibilities. It is important to make certain that the reports will be used for making decisions and directing activities. To this end, the supervisory employees involved should take part in planning the reporting system. Preparation of format and content of responsibility reports can serve to involve all the managers in the communication-information system, and the whole process can help bring managers together into a smooth-working team.

11.3.6 Analyzing Cost and Profit Data

The responsibility-center accounting scheme is just one way to organize cost and profit data for analyzing the results of operations. For a firm that produces and sells more than one product in more than one geographical area to customers that have a variety of characteristics, many other possible types of analysis are available. Several are listed next:

- *Product lines.* Analysis by product lines can show which products are most profitable. It can point up the need for cost-saving programs or identify products for maximum sales promotion.
- *Special orders.* The results of the production of special orders should always be analyzed to determine whether producing to custom-design specifications is really profitable. Such analyses could lead to special pricing formulas, better cost-estimating procedures, and more efficient use of labor and equipment.
- *Sales analysis.* The costs of selling products can be analyzed according to various customer characteristics, taking into account the cost of various selling efforts, as follows:

- *Customer groups.* Sales costs and profits can be classified by customer groups. This is especially valuable when a few customers account for a large part of the sales.
- *Industry.* Analysis of sales, costs, and profits of sales to certain industries may be useful in planning efforts to sell to growing industries.
- *Geographical areas.* The cost of shipping goods to faraway customers may consume a large part of the gross or contribution margin. Analysis by areas can help the sales manager assign salespeople where they can obtain the most profitable results.
- *Order size.* Quantity discounts must be watched. Also, classification of sales, costs, and profits by order size may bring out opportunities for new selling categories.
- *Distribution channels.* Classification of sales by distribution channels—wholesalers, retailers, and retail customers—may open up profit opportunities and areas for further investigation.
- *Combinations.* Combinations of the analyses outlined in items *a*, *b*, *c*, *d*, and *e* may give useful information. For example, sales, costs, and profits classified by both order size and geographical area might lead to the development of a new selling strategy.

These costs and profitability analyses could absorb a great deal of clerical time and perhaps cost more than they would be worth. Management of each firm must decide for itself which ones are worthwhile for its operation. Electronic data processing can play an important part. With proper programming, coding, and processing of the data, a manufacturer can have a vast amount of information and have it more promptly than by manual processing. The task is then that of using the information to plan production and sales strategies.

11.3.7 Communication of Business Financial Status

The primary service of accumulating and analyzing cost-accounting information is to have a sound basis for management decisions. But the information is also important for financial reports and presentation to outsiders.

The financial reports most often issued to outsiders are the balance sheet and the income statement. These statements of financial conditions and the results of operations are of interest to creditors (bankers and suppliers) and to investors and prospective investors.

If the manufacturer takes government contracts, governmental agencies will ask to audit the cost of production for various reasons—renegotiation of prices, establishing cost-plus prices, and so on.

Because of the close interaction between internal cost accounting and external financial reports, careful attention should be given to the accounting systems and the supporting cost-accounting records.

11.4 Mathematical Methods for Planning and Control

The cost records described in the preceding sections provide the data for more advanced mathematical methods of analyzing and planning business operations. This section outlines briefly some that can be used by many manufacturers:

- Dealing with uncertainty (probability)
- Capital budgeting analysis
- Inventory analysis (economical order quantities)
- Linear programming
- Project management
- Queuing
- Simulation

11.4.1 Dealing with Uncertainty

Business management always involves uncertainty. The manager is never entirely sure what will happen in the future. This uncertainty is of special concern in preparing budgets, establishing standard costs, and analyzing budget variances, as well as other decision situations.

Uncertainty means that the actual events a manager must try to predict or evaluate may take on any value within a reasonable range of estimated values.

The most important concept in dealing with uncertainty is probability. Two kinds of probability exist and are discussed in the following sections.

11.4.2 Objective Probability

Objective probability is a measure of the relative frequency of occurrence of some past event. It can be illustrated by the example in Table 11.4.

Assume that a manager observes production, counts the number of units produced, and tabulates the count according to the number of direct-labor hours used for each unit. This has been accomplished in columns (1) and (2) of Table 11.4. Column (3) indicates the expected probability for each time classification—that is, the odds that each unit will require the number of direct-labor hours. It is computed by dividing the number of observations for each classification by the total number of observations. For example, of the 1000 units whose production was observed, 400 required 2.5 direct-labor hours per unit. Thus, 2.5 direct-labor hours were needed for 4 out of 10 or 0.40 of the units (400 divided by 1000).

(1) Hours per unit	(2) Units observed	(3) Expected probability	(4) Expected hours	(5) Expected cost at $5
1	100	0.1	0.1	$0.50
2	200	0.2	0.4	2.00
2 ½	400	0.4	1.0	5.00
3	300	0.3	0.9	4.50
Total	1000	1.00	2.4	$12.00

Table 11.4 Objective Probability and Expected Values

The objective probabilities in column (3) are used to accumulate the expected direct-labor hours in column (4) [column (3) times column (1)]. The expected past figures in column (5) are found by multiplying the labor rate of $5.00 an hour by the expected hours in column (4). The totals in column (4) and (5) are expected values—the average direct-labor hours per unit and direct-labor cost per unit that can be expected on the basis of past production experience.

11.4.3 Subjective Probability

Subjective probability is not based on observation of past events. It is a manager's estimate of the likelihood that certain events will occur in the future.

Assume that a sales budget is being prepared. The manager estimates that chances are even (0.5) that sales will be the same as last year—1000 units. The probability of selling 1100 units is estimated to be 0.25 and the probability of selling 1200 units, 0.15. The manager estimates there is a 1-in-10 chance of selling 1300 units.

The expected sales, based on these estimates of the future, are calculated in Table 11.5. The resulting figure for expected sales, 1085, can be used in the sales budget.

(1) Sales in units	(2) Subject probability	(3) Expected sales, (2) × (1)
1000	0.5	500
1100	0.25	275
1200	0.15	180
1300	0.10	130
	1.00	1085

Table 11.5 Expected Sales

Notice that use of probabilities enables the manager to reduce a range of values to a single value. The manager can still retain for later reference the original data used to make the estimate. The use of probability measures, together with the data a manager must accumulate, helps deal with the uncertainty that is characteristic of every business operation.

11.4.4 Capital Budget

Capital is invested in a fixed asset only if the asset is expected to bring in enough profit to (1) recover the cost of the equipment and (2) provide a reasonable return on the investment.

The purpose of a capital budget analysis is to provide a sound basis for deciding whether the asset under consideration will in fact do this if it is purchased.

The following elements must be considered in managing a capital budget decision:

- The cost of the fixed asset
- The expected net cash flow provided by use of the asset
- The opportunity of investing funds in capital equipment
- The present value

11.4.5 Cost of the Fixed Asset

The cost of the fixed asset is usually comprised of the purchase price, transportation, and installation cost of the asset. However, buying the asset may mean that other investments will have to be made—additional inventories, for example. If so, these too should be included as part of the total investment outlay.

11.4.6 The Expected Net Cash Flow

The net cash flow may be either (1) the cash cost savings or (2) the increased sales revenue minus the increased cash costs due to using the new asset. No deduction for depreciation is made in calculating cash flow.

11.4.7 Opportunity Cost

Opportunity cost is the return that the business could realize by investing the money in the best alternative investment. For example, if the best alternative is to place the money in a savings account that pays 5 percent annually, the opportunity cost is 5 percent on the amount of the investment.

11.4.8 The Present Value

Scientific capital budgeting is based on the concept of present value—for example, $1.00 put in a savings account at 5 percent interest will

be worth $1.05 at the end of one year, $1.1025 at the end of the second year, and $1.1576 at the end of the third year.* Suppose an opportunity to invest some money with the expectation of receiving $1.1576 (including the amount invested and the interest or other income) at the end of the third year. Investment must not be more than $1.00, because the present value of $1.1576 to be received three years hence is $1.00 (based on the best alternative investment at 5 percent).

*First year: $\$1.00 + 0.05\ (\$1.00) = 1.05 \times \$1.00 = \$1.05$
Second year: $1.05\ (1.05 \times \$1.00) = 1.05^2 \times \$1.00 = \$1.1025$
Third: $1.05\ (1.05^2 \times \$1.00) = 1.05^3 \times \$1.00 = \$1.1576$

This present value is found by discounting the expected future value by the opportunity cost rate. This is accomplished by reversing the process for finding a future value, as follows:

$$\$1.1576 \div 1.05^3 = \$1.00$$

or, in general terms,

$$\text{Present value} = \text{value } n \text{ years hence} \div (1 + \text{discount rate})^n$$

Table 11.6 shows another way of applying the present-value concept to capital expenditure decisions. Consider an investment of $6.50 in an asset that would yield the amounts shown in Table 11.6 if placed in a savings account at 5 percent. At the end of five years, the amounts received from the asset investment would total $7.00 (including the salvage value of the asset), whereas the savings account would be worth $\$6.50 \times 1.05^3$, or $7.52. Therefore, the $6.50 should not be invested in the asset.

The present-value analysis in column (4) of Table 11.6 shows that the present value of the asset investment is $6.30, which is 20 cents less than the amount that would be invested. In other words, the value of the investment opportunity is minus 20 cents. If the amount

(1)	(2)	(3)	(4)
End of year	Amount to be received	Discount factor	Present value
1	$2.00	1.05	$1.90
2	2.00	$(1.05)^2$	1.81
3	3.00	$(1.05)^3$	2.59
Total	$7.00		$6.30

Table 11.6 Application of Present Values to Capital Investment

	Year 1	Year 2	Year 3
Sales revenue	$10,000	$11,000	$12,000
Less: variable costs	6,000	7,000	8,000
Less: fixed costs*	2,000	2,000	2,000
Cash flow from production	$2,000	$2,000	$2,000
Sales of equipment	0	0	1,000
Net cash flow	$2,000	$2,000	$3,000

*Not including depreciation charges.

TABLE 11.7 Revenues and Cost Characteristics

of the investment were $6.29 or less and brought in the same amount, the present value of the investment opportunity would be positive. The general rule is that investment opportunities with a positive present value should be taken.

Example. Assume you have an opportunity to buy a piece of equipment for $5000. With this equipment, products can be manufactured with the revenue and cost characteristics described in Table 11.7.

The present value of the cash flows in Table 11.7, computed as in Table 11.6, is $6300. The value of the investment in the opportunity is therefore $6300 – $5000, or $1300. This positive investment opportunity value shows that the equipment investment will have a higher rate of return than the 5 percent available on the next best investment (assumed to be a savings account paying 5 percent). The depreciation is not considered a cash flow item and thus should not be included as a cash cost. The example illustrated does not include depreciation charges. In addition, these figures do not include income tax payments (which are a cash flow item).

However, the cash flow figures in Table 11.8 do take taxes into account. Once taxes are included, depreciation should be considered as well, since depreciation is tax deductible.

This calculation may also be set up as follows:

Net cash flow (ignoring taxes)	$2000	$2000	$3000
Less: Income taxes—as above	80	80	80
Net cash flow	$1920	$1920	$2920

The present value is $6092, rather than the $6300 calculated from the earlier figure, which did not include taxes. The value of the investment opportunity, after income taxes, is $1092, which is still higher

	Year 1	Year 2	Year 3
Sales revenue	$10,000	$11,000	$12,000
Less: variable production costs	$6,000	$7,000	$8,000
Less: fixed production costs (including depreciation)	3,833	3,833	3,834
Net profit before taxes	167	167	166
Less: income taxes @ 48%	80	80	80
Net profit after taxes	87	87	86
Cash flow from sales of equipment (add back depreciation)	$1,833	$1,833	$1,834
Net cash flow	$1,920	$1,920	$1,920

TABLE 11.8 Cash Flow

than what could be earned from a savings account carrying a 5 percent interest rate.

The present-value method of analyzing a capital investment opportunity makes it possible to take into account the timing of the expected cash returns and to compare them with those of other investment opportunities. Note that the cash returns are expected values. The probability method discussed earlier is used in calculating them.

11.4.9 Inventory Analysis

Maintaining adequate inventories of raw materials and parts can tie up capital for a manufacturer. Careful management of inventory levels offers possibilities for large cost savings and the release of funds from inventory investment. Two major kinds of costs are associated with maintaining inventories: holding costs and ordering costs.

Holding costs are costs associated with inventory investment. They include insurance, taxes, rent on warehouse space, and the opportunity cost of alternative uses of the funds invested in inventory.

Ordering costs are the costs of processing purchase requisitions and vendors' invoices and of receiving the goods in the warehouse (receiving department salaries, etc.).

The *economic order quantity* (EOQ) is the quantity of goods that should be ordered at one time to ensure the lowest total inventory costs (holding cost plus ordering cost). There is a tradeoff between holding costs and ordering costs. When the order quantity is small and the inventory is kept low, the holding costs are also low. The ordering costs, however, are high, because orders must be placed often and more clerical and receiving department time is needed. When the order quantities are large and inventory is kept high, the holding costs are high (more insurance, taxes, rent), but the ordering

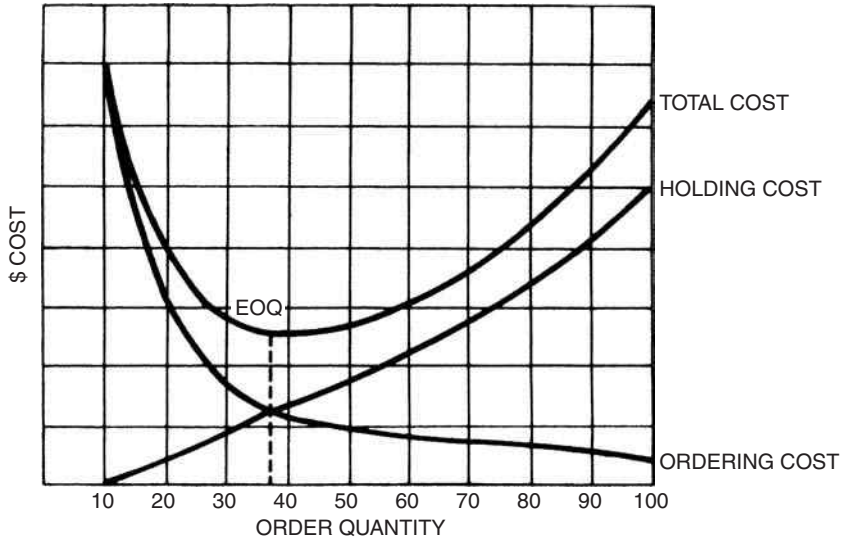


Figure 11.4 Economical order quantity.

cost is low—fewer requisitions must be processed, and receiving operations are less frequent. This cost tradeoff is shown graphically in Fig. 11.4.

The basic data for the inventory problem are the holding costs, the ordering costs, and the expected demand for raw materials and purchased parts. Other factors that may enter into purchasing decisions are the lead time (the length of time between purchase order and receipt of the materials or parts) and the availability of quantity discounts.

11.4.10 Linear Programming

Linear programming is a mathematical method for making the best possible allocation of limited resources (labor hours, machine hours, materials, etc.). It can help management decide how to use its production facilities most profitably. Suppose, for instance, that the firm produces more than one product, each with a different contribution rate. Management needs to know what combination of quantities produced, given the limitations of its facilities, will bring the highest profit.

Three principal concepts must be considered in seeking a solution to the problem:

- The profit function
- The constraints of the problem
- The production characteristics

11.4.11 The Profit Function

The profit function is a mathematical expression used to show how the profit will vary when different quantities of product 1 and product 2 are produced. It requires knowledge of the contribution that each product makes toward overhead and profit.

If the symbols x_1 and x_2 stand for the number of each product to be produced, and if product 1 makes a profit contribution of \$2 per unit and product 2's profit contribution is \$5 per unit, the profit function is $2x_1 + 5x_2$. The problem is to find the values of x_1 and x_2 that will yield the highest total value (profit), given the constraints of the problem and the production characteristics.

11.4.12 The Constraints

Sometimes the profit-maximizing solution is obvious—simply produce as many product units as possible with the resources available. However, the manufacturing process may have characteristics that prevent the use of resources in this way.

The constraints of this problem are the limitations imposed by the scarcity of production resources. Suppose that a manufacturer uses lathes, drilling machines, and polishing machines in the production process. The lathes can be used for a maximum of 400 h a month, the drilling machine for 300 h, and polishing machine for 500 h. The machines cannot be used more than the hours indicated, but they need not be used to the limit of this capacity in order to maximize the profit function. The constraints are applicable only as long as the production capacity remains constant. The manufacturer could purchase new machines or add work shifts and expand the machine time available. But both of these actions will probably change the statement of the problem. The profit contribution of the product might be changed (by higher night-shift direct-labor costs, for example). Also, a capital budgeting analysis might be required. The linear programming problem is applicable for a time period during which constraints are fixed.

11.4.12.1 Production Characteristics

The term *production characteristics* refer to the machine times used in producing the product. Suppose that product 1 requires 1 h on a lathe, no time on a drilling machine, and 1 h on a polishing machine. Product 2 requires no lathe time, 1 h on a drilling machine, and 1 h on a polishing machine. As stated earlier, the lathe can be used for a maximum of 400 h a month, the drilling machine for 300 h, and the polishing machine for 500 h. The production characteristics would then be expressed as follows:

$$x_1 \leq 400$$

No more than 400 lathe hours are available for product 1, and 1 h is required for each unit. (If 2 h per unit were required, the expression would be $2x_1 \leq 400$.)

$x_2 \leq 300$ No more than 300 drill hours are available for product 2, and 1 h per unit is required.

$x_1 + x_2 \leq 500$ No more than 500 polishing hours are available for both products, with 1 hour per unit required in both cases.

Solving the problem. The problem for the values given previously is as follows:

Many solutions would satisfy all three constraints, but there is only one maximizing solution. The possible solutions are shown graphically in Fig. 11.5. The axes represent the number of product units; the remaining solid lines, the production characteristics. Any point in the shaded area will satisfy the constraints, but the maximizing solution will be at one of the four corners of this area—point *A*, *B*, *C*, or *D*.

This point is the exact one which can be found by drawing a line that represents the profit function $2x_1 + 5x_2$ (the lower dotted line in Fig. 11.5). When this line is shifted upward, always parallel to its original position, it finally reaches the point where it touches the shaded solution space at only one point, B. This point, where $x_1 = 200$ and $x_2 = 300$, is the solution that gives the highest possible total contribution within the limited resource constraints. The solution gives the following results:

Total contribution	\$ 1900 ($200 \times \$2 + 300 \times \$5$)
Lathe time used	200 h (200×1 h)
Drill time used	300 h (300×1 h)
Polishing time used	500 h (200×1 h + 300×1 h)

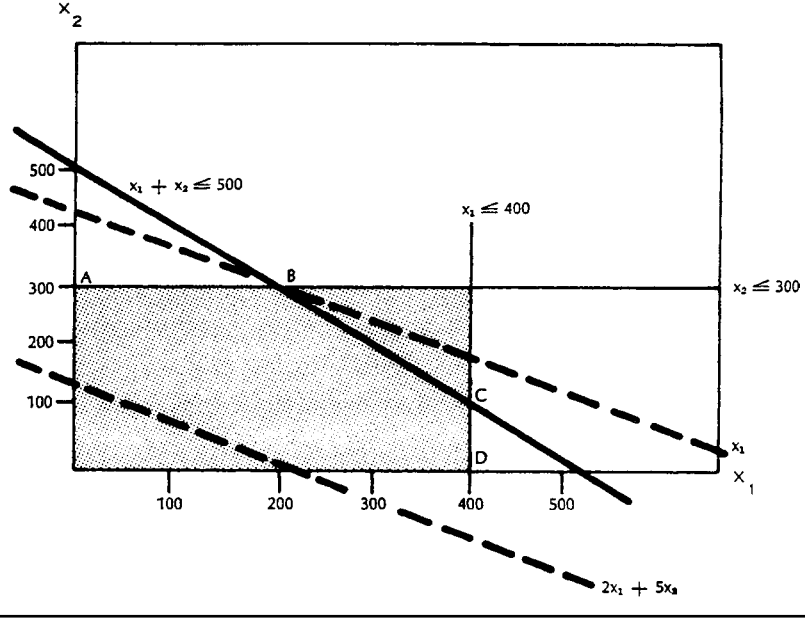


Figure 11.5 Profit function.

There are 200 idle lathe hours, but these cannot be used under existing conditions without reducing the output of product 2. If more lathe time is used for product 1, then more polishing time must also be used; and the polishing time must be taken away from product 2, thus reducing product 2 output. Any reduction in product 2 output represents a loss of $5 per unit, whereas producing additional units of product 1 will add only $2 per unit. The net loss from such a substitution is therefore $3 per unit. On the other hand, the output of product 2 cannot be increased, since there is no more drill time available.

The graphic solution described here is a rough approach to a fairly simple problem. More complex problems can be programmed into a computerized automated system.

11.4.13 Project Management

Manufacturing and construction firms may undertake one-time-only jobs, such as the construction of a building or the design and manufacture of a special piece of a large assembly system. In such projects, at any one time much of the company's time, resources, and prospects may be tied to a single large job. And whether the job will result in profit or loss for the company may depend on its being completed on time.

If the completion date is to be met, the various segments of the job must be coordinated. Subunits must be scheduled, and they must be finished according to the schedule date.

Project management techniques have been developed for planning the subunits that make up such a job. One of the most important of these methods is the project evaluation and review technique (PERT). Another one, very similar to PERT, is known as the critical path method (CPM).

11.4.13.1 PERT

PERT is based on what is called a *network plan*. A flowchart—the network—is used to illustrate how the individual parts of a project (the activities and the starting and completion events) depend on one another and which task must be finished before others can be started. It is particularly concerned with dovetailing individual parts of a project into the schedule of the project as a whole.

PERT analysis makes the following estimates from data about the activities and events that make up the project:

- Earliest expected time for completion of each activity, or task
- The latest allowable time for each completion event
- The critical path of the network—that is, the path through a flowchart that has the least total amount of slack (*Slack* is the difference between the latest allowable time and the expected time of completion.)
- The probabilities that events will occur on a schedule

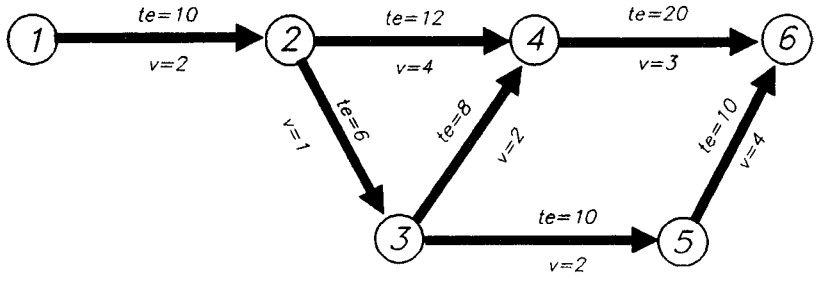


Figure 11.6 A simplified PERT network.

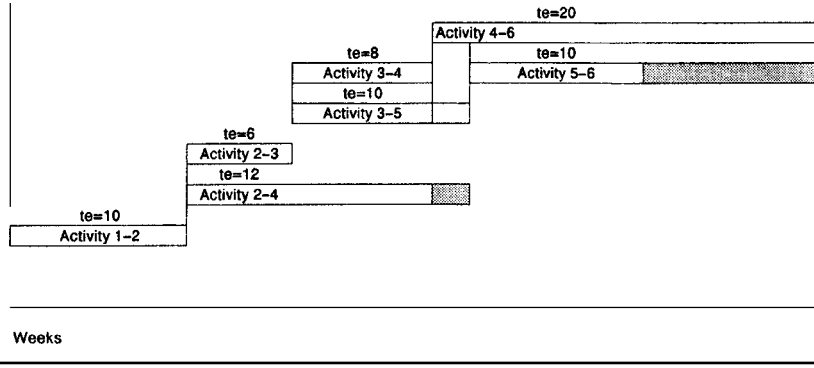


Figure 11.7 A PERT network as a bar chart.

A simplified PERT network is shown in Fig. 11.6. The same information is shown as a bar chart in Fig. 11.7. (The bar chart brings out the pattern of activities and slack time clearly, but would be unwieldy for use in a more complex network.)

The circled numbers in Fig. 11.6 represent events—the start or completion of an activity such as excavation for a building or erection of the steelwork. The lines represent work activities extending over time and leading from one event to another. Other symbols used in the figure and discussion have the following meanings:

t_e = expected time
T_E = sum of the expected times up to a given event
T_L = latest allowable time (cumulative)
v = variance

The values for t_e and v assumed in the following paragraph are those in the illustration.

If o1 represents the beginning of excavation for a building, o2 completion of excavation, and o3 completion of pouring a foundation, then the line from o1 to o2 represents the activity of excavation and the line from o2 to o3 represents the activity of pouring a foundation.

The activity of steelwork erection leading to completion event o4 cannot begin until after both the preceding completions, o2 and o3, have occurred. Other intermediate completion events must take place before any entire project is finally completed, event o6.

Time estimates must be made for each of the activities. Since actual activity time is uncertain, probability calculations are used in computing the expected time t_e and the variance v associated with the expected time.

The problem for project management is to determine the expected time for completion of the entire project and to identify the activities that could cause bottlenecks. PERT enables the management to accomplish this and thus schedule work activities for the earliest possible completion.

For example, in Fig. 11.6, work leading to completion event o2 is expected to be completed ten weeks after the start. From this point, two network paths lead to event o4: the path 2–4, with an expected completion time of 12 weeks, and the path 2–3–4, with a completion time of 14 weeks. The total expected completion time for event o4 is the larger of the total completion times on the two network paths that lead to the events. The path 1–2–4 has a total expected time of $T_E = 22$, and the path 1–2–3–4 has a total expected time of $T_E = 24$. Thus, the T_E for event o4 is 24. The variance associated with $T_E = 24$ is $v = 5$, the sum of the variances along the 1–2–3–4 path.

Thus, the manager can expect completion of event o2 in ten weeks, event o4 in 24 weeks, and event o6 in 44 weeks after work is started. One other time variable must be taken into account, however: the latest time T_L that can be allowed for an event without disturbing the T_E of the final event in the network (event o6).

If event o6 must be completed in 44 weeks, then T_L for event o6 is 44. Since T_E for event o6 is also 44, the allowable lag time is zero (T_L–T_E). In order to complete event o6 in 44 weeks, event o4 must be completed in 24 weeks. Thus, T_L for event o4 is 24 and allowable slack time is also zero. T_L and slack time can be calculated for each event by working backward from event o6.

The critical path through the network—the path in which there is least amount of slack—is identified by noting the events where allowable slack time is zero. In Fig. 11.6, the critical path lies through events 1–2–4–6. Any construction delay in these events will increase the total completion time for the project. Event o5, however, is not on the critical path. The tasks on the path segment 3–5–6 could be delayed for as much as eight weeks in all without affecting the final completion time for event o6. The activities represented by the path segment 2–4 could be delayed two weeks without delaying the final completion of the project. Resources could, if necessary, be shifted from these activities to other tasks to ensure completion of the entire project on schedule.

Note that T_E and T_L are expected values. Statistical variances are associated with each of these measures. The expected time estimates

and variance estimates are used in calculating the probabilities that events will occur on schedule.

11.4.13.2 CPM

The critical path method of project management is closely related to PERT. Both PERT and CPM involve determining expected times of completion for individual events and for the entire project. PERT goes further to include variance analysis; CPM does not. CPM, on the other hand, goes beyond PERT in another direction. It uses cost data to assess the financial effects of setting up crash programs in the network's critical path segments to ensure completion on schedule.

11.4.13.3 PERT/Cost

PERT can be augmented with cost data to facilitate financial planning for large projects. The cost data are budgeted according to work activity and completion-event classifications. Actual costs should be accumulated according to the same classifications so they can be compared with the budget figures.

11.4.13.4 Queuing or Waiting Line Analysis

A queuing or waiting-line situation exists when there is a service center of some kind that receives customers, processes their work, and sends them on their way. Three types of waiting lines often operate in business. One is seen in a doctor's office where arriving patients form one line and are served through only one station—the doctor. Another is seen in a barber shop where people form one line and are served through several stations—any of the barbers. A third type of queuing is in a supermarket where customers form many lines and are served through many stations—the checkout counters.

A waiting line can occur in a manufacturing plant when product units—"customers"—arrive at a processing department or work area for completion and transfer to the next department. A manufacturing queuing situation is illustrated in Fig. 11.8.

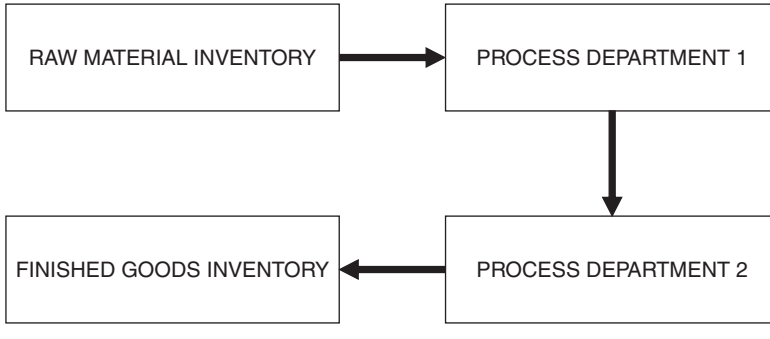


Figure 11.8 A manufacturing queuing situation.

The arrival of raw materials at processing department 1 can be controlled by the timing of materials requisitions, so it is unlikely that a waiting line of materials units will form at department 1. However, the processing or service rate in department 1 determines the timing of the release of the semifinished units to department 2.

As long as department 2 can accept the units for processing, there will be no waiting line formed. Units will be processed as they arrive at the work area. However, if department 2 is busy, a bottleneck will occur, and semifinished units will form a backlog. The time spent in line carries a cost to the manufacturer similar to the holding cost of inventory. (Note that the backlog of semifinished products is a part of the work-in-process inventory.) The queue, and therefore the waiting cost, can be reduced by speeding up the processing or service rate in department 2.

The processing or servicing of additional units, however, adds to the cost of the manufacturing process. A move to accelerate processing in department 2 and shorten the queue may add costs beyond normal production costs for materials, labor, and overhead. This would be the case, for instance, if it were necessary to add a night shift or new equipment. Thus, there is a tradeoff between waiting costs and some processing costs similar to the EOQ tradeoff between holding and ordering costs. Queuing analysis takes into consideration the following factors:

- Arrival rate
- Service rate
- Queue lengths
- Utilization of the service facility
- Total service time for produced units
- The cost of waiting
- The cost of servicing

These variables are combined in such a way as to provide data on the efficiency of the present service and the economic facility of adding service units.

11.4.14 Simulation

Simulation is a method used to represent a real situation in an analytical framework. Verbal descriptions, diagrams, flowcharts, and systems of equations are all forms of simulation. A budget is the result of simulating the financial results of operations, based on assumptions about volume, cost relationships, and planned production activities.

Simulation allows management to try out various actions without risking its resources in the real marketplace. With simulation,

management can calculate events on the basis of alternative assumptions on paper and then attempt to select the best course of action for the real application.

In the terminology of modern management science, *simulation* refers to a computerized model of a company or some segment of it. The model includes sets of mathematical equations that quantify volume-cost-profit relationships. The computer can accept changes in the relationships and calculate the effects of the changes on operations. Suppose, for example, that management is contemplating substituting a new raw material that will reduce materials requirements but increase labor. Management can program the changes into a computerized simulation and learn what the overall effect will be on manufacturing operations.

Simulation is a powerful management tool, but it is complicated by the fact that the operating characteristics have to be specified in great detail. Nevertheless, it can be used to great advantage to test and examine alternative courses of action without having to put them into actual practice. The volume, cost, profit relationships, and physical production characteristics described in this chapter can provide much of the data required for simulation models.

11.5 Where Do Sensors and Control Systems Take Us?

Currently, most industries employ equipment with limited intelligence, but in the near future, the advancing level of adaptive sensors and control systems and of computer technology will rapidly increase the use of "smart" equipment. On the other hand, as computer integrated manufacturing strategies are introduced on a wide scale and as technology moves forward, the size of the work force involved in engineering will be reduced dramatically, while the pattern of working seems destined to change completely (Fig. 11.9).

Already sensors and control systems play a significant role in manufacturing. An example is tool wear and breakage monitored by sensors using the torque or power at the spindle. Macotech Machine Tool Co. Inc., Seattle, Washington, has taken that concept a step further to adjust the feed rate and torque to obtain the optimum cutting rate whatever the conditions. As a result, cutting times have in many cases been reduced by 50 percent, while the need to program the tool in detail is eliminated; sensors and control systems no longer need to be told the material, cutting speed, and feed rate.

In one case, the concept of employing sensors and control systems in manufacturing has been applied to deep drilling, where a constant load is applied. The feed rate changes automatically to maintain the constant load on the cutting tool, but if as a result of the buildup of swarf the feed rate is reduced to a preset level, the drill is withdrawn to clear the swarf, and then starts again. Not only is the

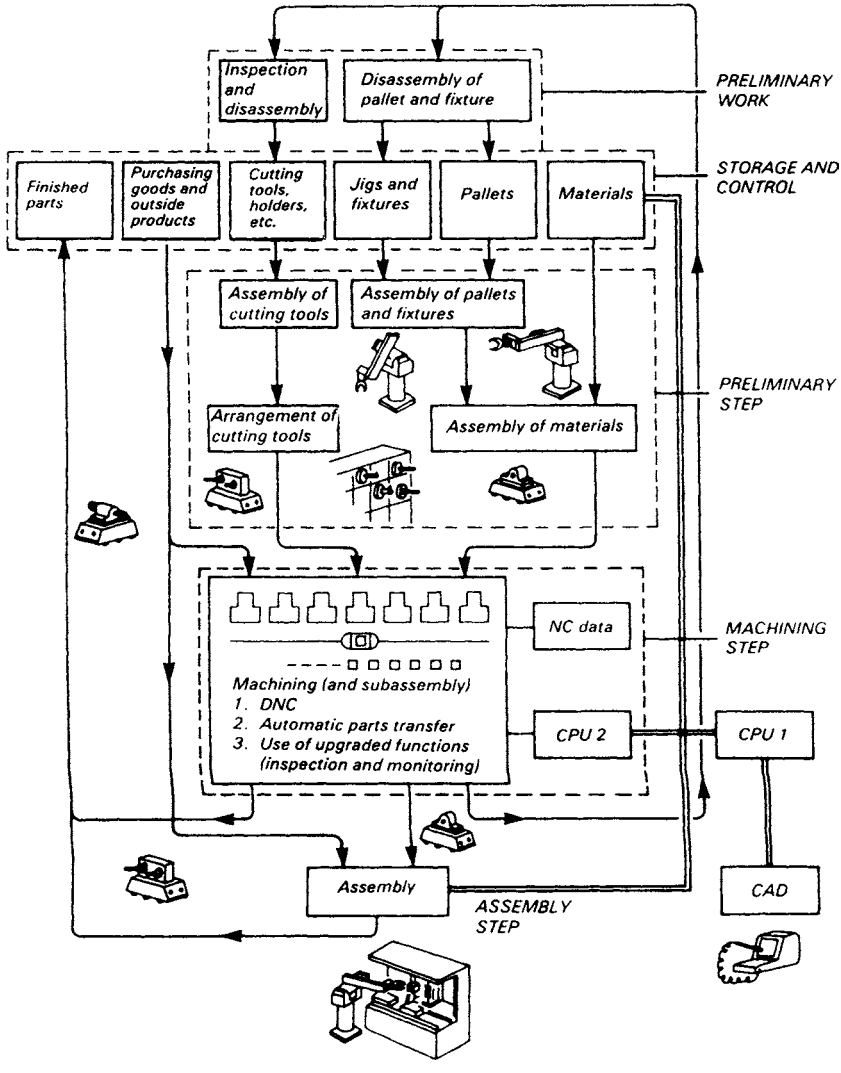


FIGURE 11.9 The current vision of CIM.

speed optimized through sensors and control systems, but as a result, drill wear is more consistent, and tools do not break.

The aim in increasing the level of intelligence of machine or assembly controls is to reduce the need for attention and programming. The extra "senses" needed are vision sensors, force sensors, optical sensors, and metrological sensors. With a vision system, the controller of the robot or machine will be able to recognize the workpiece, determine its precise position, and instruct the machine to carry out the necessary operations. Thus, it will not be necessary to

load a workpiece precisely on a pallet or install limit switches and other sensors to identify the workpiece and check its position. Since location need not be precise, sensors in robot loading will be easier to arrange. The force sensors will optimize the cutting speed whatever the material, and for a robot, will ensure that a component is installed with the correct amount of force. The use of such sensors and control systems will allow planning of manufacturing operations to be automated, with such programming as is needed done off-line.

A lathe or milling machine would be equipped with adaptive sensors and control systems on the spindle, to adjust the speed for turning, drilling, tapping, and milling. In that case, the vision sensors would be used to identify the workpiece for the robot. Lathes and milling machines can already be programmed off-line, and with vision sensors, the robot could be programmed in the same way, with the sensors adjusting the position and size of the gripper or changing the gripper to suit the workpiece. Also, the controller could assess whether a new set of chuck jaws is needed, and if so, instruct the robot to change it. Thus, without the need for in-line programming and adaptive control the lathe or milling machine would be able to machine a wide variety of components completely unmanned.

Just as the productivity of lathes and milling machines has improved, quantitatively and qualitatively, so other machines can be improved through sensors and control systems. For example, several machining centers have an extra attachment to allow machining of the fifth face, whereas some machines made by Mandelli Machine Tool, Italy, can machine the fifth face without the need for an attachment. This is a worthwhile development, since it reduces downtime. With the introduction of modular universal machining centers for turning, milling, boring, and grinding operations, manufacturers will encourage more moves in this direction. In some cases, handling equipment to permit the fifth face to be machined without the need for a full resetting operation is also likely to come into use, as are the laser welding and laser sensor measurement systems on the machine tool itself.

But the major developments will continue to come in sensors and control systems, computers, and software. In the next stage, the computer at each flexible manufacturing system might have access to an "expert" system, which is in effect a huge database. Included in the database are data concerning all the components to be produced—input from the CAD system—such as machining methods for different speeds, surface finishes, and tolerances. In fact, the database will be a complete encyclopedia of machining. This will be updated continually and directly from the CAD system and from a remote database on machining (Fig. 11.10).

Once sensors and control systems are perfected, this will allow machining to be performed without programming. When the workpiece is fed into the system, a vision sensor will identify it in detail.

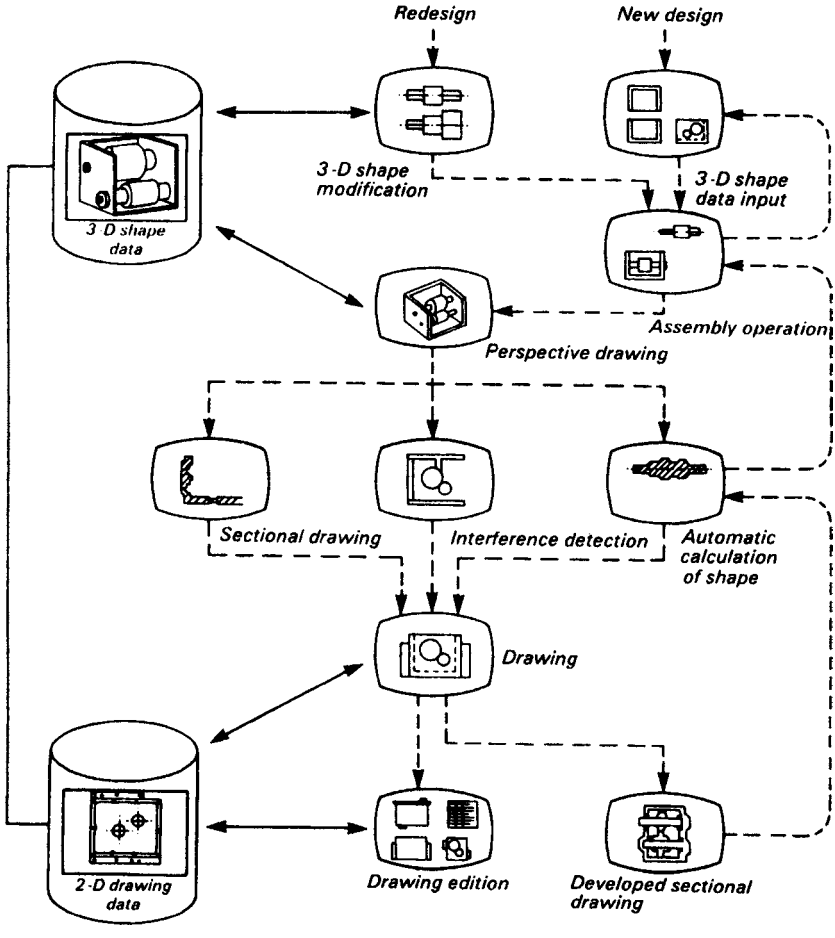


FIGURE 11.10 A dynamic manufacturing operation.

Then the data will be fed back to the database and compared with data for the workpiece that are to be machined, and full dimensional data on the workpiece will be accessed. The vision sensor will then scan the workpiece, and its dimensions will be compared with those of the finished part in the parts library.

The data will then be fed back to the controller, which will instruct the machine how to start machining, by reference to the database and not to a specific program. When necessary, the workpiece will be transferred to other places for automatic heat treatment or grinding, and later to assembly. In this way, machining without programming and without manual intervention will be practical. Lathes are unlikely to need such a control system, and instead, the database would contain a vast number of programs, the best one being selected

automatically for the shape of the component to be produced. Of course, these advanced systems in controls and sensors would not be needed to produce all components, nor would they be installed in all flexible manufacturing systems. Lower-level machines would still continue in use for more routine, medium-volume parts. However, in planning for flexible manufacturing systems with computer-integrated manufacturing strategies, managers need to take this sort of development into account.

In assembly, a similar sensor and control system could come into use, except that there will be more stages, because of the need to provide the right group of components in suitable positions to be assembled. One handling robot, on a trolley, might be used to supply components to the assembly cell, where some robots and standard units, such as presses and CNC nut runners, would carry out assembly. With the combination of these expert systems and sensors and control systems, flexible manufacturing systems will be able to machine or assemble in batches of one with the efficiency expected in high-volume production runs. Thus, the real aims of flexible manufacturing systems—high productivity in small batches, produced within a very short time after the order is placed—will be achieved.

The other major development is in computer systems for functions from purchasing through delivery. All the systems will be connected together, and will perhaps be incorporated in one computer (which will have voice recognition and voice response), so that for most jobs the keyboard will be unnecessary. Indeed, by the end of this decade, computers with conventional languages and vocal human/machine interfaces should be available. Anyone will be able to program and make the most complex transactions using normal conversation, so the problem of computer illiteracy will disappear. People will also have access to databases of products, materials, and stock and to expert systems on costing and other relevant subjects.

In the case of production control and order processing, this means that, once an order is accepted, the data will be fed to the CAD system, where the designer will produce a new design, modify a design, or merely call up an existing design. Planning will be automatic; from the data generated, the schedule will be produced and components and materials ordered. On arrival, they will be fed through the system to be packed and delivered with hardly any manual intervention. Needless to say, the materials will not wait around in a storeroom, but will come straight off the delivery truck into the system to be processed.

Overall, therefore, we are at the beginning of an era which will lead to the end of the factory as we know it. Some machines will be able to carry out simple routine maintenance themselves through advanced sensory and control systems. Very complex robots with a number of sensors will be available to carry out repairs, so that, in theory, operators will be needed only to work with the software—in

designing and updating the databases and expert systems. Both the machines and robots will be extremely complex, so the operators, who are responsible for their installation, let alone maintenance, will probably not be able to understand the complete system; that is already the situation with large-scale computers.

Some engineers expect to continue to rely on people to carry out some jobs. They argue that there is little point in taking everyone out of the plants just for the sake of doing so, and in any case, the devices that can do these jobs are likely to be very expensive. In addition, even with all these elaborate sensors, there is the possibility of some catastrophic failure, and it will be a very long time before managers are prepared to go home at night and leave a plant completely unattended. In any case, working conditions will be very good, and working hours will be short, so there is every reason to employ some people in the plants.

However, other experts say there is no point in thinking in terms of staffing the plant at all, except in an external supervisory role. They say that, because the machines will be quite complex, maintenance fitters will not know how to rectify faults. Then, because the sensors and control systems will be so advanced, the machine will in any case be better able to trace faults and put them right, so rectification by humans may not be helpful at all. Clearly, that stage is still a long way away, but managers need to be ready for it.

But just how many people will be needed in manufacturing by the year 2020? Some experts estimate that the work force in manufacturing will be cut by 55 to 65 percent in the next ten years, while the number of people actually employed on the shop floor is likely to drop by about 90 percent. The effect of a 70 percent reduction in the work forces of the major industrialized nations is shown in Table 11.9.

Although these figures appear alarming, they reflect similar trends that have taken place in agriculture, process industries, automotive industries, and computer industries. However, those changes

Country	Current industrial work force	Industrial work force reduced by 70%
United States	30,000,000	9,000,000
Japan	19,500,000	5,800,000
West Germany	11,500,000	3,400,000
Italy	7,700,000	2,300,000
France	7,500,000	2,250,000
United Kingdom	5,300,000	1,600,000
Sweden	1,350,000	400,000

Table 11.9 Worldwide Work Force Reduction

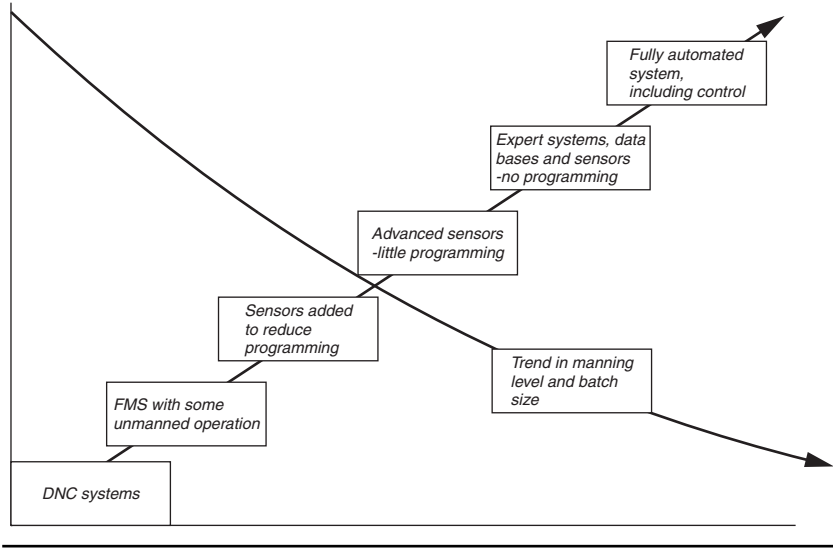


Figure 11.11 The role of sensors in moving manufacturing technology forward to the year 2020.

took place far more slowly, and the rate of change is bound to cause considerable problems to politicians and industrialists alike. With the prospect of 22,000,000 jobs being lost in the United States, 14,000,000 in Japan, 9,000,000 each in France and Italy, and 5,500,000 in the United Kingdom over a ten-year period, people are bound to be alarmed. Figure 11.11 shows the expected relationship between manufacturing technology advances and manufacturing employment.

In this period, though, many new products and systems will go into production, and, in theory, with the reduction in prices that the increased productivity will bring, demand worldwide should increase. Equally, it is clear that the development and maintenance of all this software will provide a large number of skilled jobs. In fact, there is likely to be a shortage of skills in many areas, which will have the effect of retarding development of new systems.

Clearly, the nations that train people in the relevant skills will have a head start. In any case, the changes in the number of employees are based on the assumption that each nation maintains its current share of the market for manufactured goods. In practice, this is unlikely, since some countries are more competitive than others, while the newly industrialized nations will also take some of that business.

Although it is imperative that people be educated in the right skills, this is not enough. The whole way in which people are educated and how they think of work must be changed. The introduction of computer-integrated manufacturing strategies and computerized sensors and control systems in manufacturing, transport, and perhaps

some other services will lead to the need for people to work fewer hours. It is likely that a three-day week will appear quite soon, both to spread work around and to ensure that people are prepared to work the unsocial hours that are occasionally needed.

But what will people do when they are not working? Another worry is that, as we move from active to supervisory roles, we may tend to lose skills in dexterity and become less active. The education level needs to be raised so that people can cope with the jobs available in this new society, while education also needs to be expanded so people are better equipped to face life with more time on their hands and perhaps do more for themselves in maintaining their homes and their possessions.

Perhaps many people will have more than one job, one in one company for three days a week, and another outside it for two days. Of course, others will not want much work at all, and they will probably be able to fulfill that wish. But none of those futuristic notions will be possible unless a great deal of value is being added in manufacturing or business.

We are entering an exciting era of significant opportunities, opened up by sensors, microelectronics control systems, and computer-integrated manufacturing strategies. But the competition will be intense, and only those that take big chances now will reap the benefits. As a psalmist once said, "They that sow in tears shall reap in joy. He that goeth forth and weepeth, bearing precious seed, shall doubtless come again rejoicing, bringing his sheaves with him." (Psalms 126, verses 5–6, *The Bible*)

Further Reading

Ansoff, H. I., "Planning as a Practical Management Tool," *Financial Executive*, 32: 34–37 (June 1984).

Bacon, J., "Management of the Budget Function," Studies in Business Policy, report no. 131, National Industrial Conference Board, New York, 1990.

Chambers, J. C., S. K. Mullick, and D. D. Smith, "How to Choose the Right Forecasting Technique," *Harvard Business Review*, 49:45–74 (July–August 1991).

Gentry, J. A., and S. A. Pyherr, "Stimulating an EPS Growth Model," *Financial Management*, 2:68–75 (Summer 1983).

Gershefski, G. W., "Building a Corporate Financial Model," *Harvard Business Review*, 47:61–72 (July–August 1989).

Gordon, M. J., and G. Shillinglaw, *Accounting: A Management Approach*, 4th ed., Chap. 16, Homewood, Ill., 1969.

Hamermesh, R. G., "Responding to Divisional Profit Crises," *Harvard Business Review*, 55:124–130 (March–April 1977).

Meyer, S. C., and G. A. Pogue, "A Programming Approach to Corporate Financial Management," *Journal of Finance*, 29:579–599 (May 1974).

Pappas, J. L., and G. P. Huber, "Probabilistic Short-Term Financial Planning," *Financial Management*, 2:36–44 (Autumn 1987).

Parker, G. C., and E. L. Segura, "How to Set a Better Forecast," *Harvard Business Review*, 49:99–109 (March–April 1991).

Smith, G., and W. Brainard, "The Value of a priori Information in Estimating a Financial Model," *Journal of Finance*, 31:1299–1322 (December 1976).
Soloman, S., *Modern Welding Technology*, TAB Books, Blue Ridge, Pa., 1982.
Wagle, B., "The Use of Models for Environmental Forecasting and Corporate Planning," *Operational Research Quarterly*, 22 (3):327–336.
Warren, J. M., and J. P. Shelton, "A Simultaneous Equation Approach to Financial Planning," *Journal of Finance*, 26: 1123–1142 (December 1991).
Weston, J. F., "Forecasting Financial Requirements," *Accounting Review*, 33: 427–440 (July 1958).

Index

A

A posteriori information, 196
A priori information, 196, 393, 413
AbNET protocol, 221–223
Absolute algorithm, 288
Absolute control systems, 456–457
Absolute optical encoder sensor, 280
Absorbed light, optical spectrometer, 357
Absorption curve, 168
AC power cycle, 83
Acceleration
  defined, 422
  process control sensors for, 371–372
  versus time for various inserts, 206
Accounts payable, 244–245
Accounts receivable, 245
Accuracy, 422. *See also* Radiometric accuracy
Aceh Province, Indonesia, 3
Acidity of solutions, quantifying with optical sensors, 317–318
Acoustic sensors, 381–382
Acoustics, 198
Acoustooptical synthetic-aperture radar, 305–307
Active monitor node, 222–223
Active star coupler, 164–165
Active star networks, 181–182
Active-type sensors, 328
Actual values, 50, 462
Actuators, 257–259, 325
ADC (analog-to-digital converter), 259, 328, 503
ADC (automatic defect classification), 113
Adjacent sensors, 74, 79
Administrative budget, 525–526
Adobe Wide Gamut RGB, 11
Advance interactive executive (AIX), 269
Advanced multipurpose industrial robot installations, 435
Advanced Programming Tool (APT), 463
Agriculture applications, 310–312
AGV (automated guided vehicle), 140–141, 250, 427
AI (artificial intelligence), 246
Air bubbles, flow sensors detecting, 351–352
AIX (advance interactive executive), 269
AL (application layer), 208–209
AlGaAs (aluminum gallium arsenide), 330
Aliasing
  defined, 490
  spectral, 522
Alignment techniques, sensor, 152–155
Allen Bradley Company, 446
Allocation list, PLC program, 473
Alternate frequency heads, 74
Aluminum gallium arsenide (AlGaAs), 330
Ambient light, 283
Ambient operating temperature, 30
American National Standard for Industrial Robots and Robot Systems—Safety Requirements, 426
American SensoRx SpectRx software, 494
Ames Microfluor detector, 319
Ammeters, 189
Amplitude probability density (APD), 205, 207
Analog input, 259–260
Analog links, 161
Analog output, 260
Analog values, liquid mixing operation, 485
Analog-to-digital converter (ADC), 259, 328, 503

AND function, 91–93
Angiopoietin-like 4 (ANGPTL4), 356
Angular velocity measurement, 103
Antennas, 100
APD (amplitude probability density), 205, 207
APDs. *See* Avalanche photodiode detectors
Aperture stop (AS), 316, 490, 492
Apodization, 508–509
Application layer (AL), 208–209
APT (Advanced Programming Tool), 463
Array cameras, 382
Arrays, LED, 19
Artificial intelligence (AI), 246
AS (aperture stop), 316, 490, 492
AS/RS (automated storage and retrieval systems), 248, 250
Assembly, multisensor-controlled robot, 412–417
Associative recall loop, 308
Asynchronous data transmission, 481
Attenuators, 177
Autocorrelation function, 501
Automated guided vehicle (AGV), 140–141, 250, 427
Automated method, structure vibration measurement, 300–301
Automated storage and retrieval systems (AS/RS), 248, 250
Automatic defect classification (ADC), 113
Automatic machinery, laser sensors for, 128–131
Automation program, establishing, 5–7
Automotive manufacturing industries, laser sensors for, 122–125
Autonomous-vehicle speed monitoring, 103
Auxiliary NC commands, 466
Available light sensing options, 288–289
Avalanche photodiode detectors (APDs)
  critical performance parameters, 362–363
  dark current and noise current, 365–367
  Geiger mode, 367
  overview, 359–360
  selecting
    gain, 364–365
    overview, 363–364
    responsivity, 364–365
    types of, 364
  structures, 360–362
Axes, defined, 421
Axial approach, 71

B

Background suppression, 147–148
Balance sheet projections, 526
Bandpass spectral filter, 322
Bar-code identification systems, 275–277
Bars, LED, 19
Beam angles, calculating, 37–38
Beam break mode (direct scanning), 145
Beam splitter, 303
Bearing fault sensors, 298–300
Beer-Lambert relationship, 356
Bent-pipe satellite, 225
BER (bit error rate), 159
Bifurcate, 166
Bill of materials (BOM) file, 250
Billing, 245
Binary logic chart–parallel connection, 90
Binary signals, 439
Biological material sensors, 319–320
Biomedical technology sensors
  detecting biological materials, 319–320
  early detection and treatment of lung tumors, 320–321
  overview, 319
  ultrasensitive single-molecule detection, 321–323
Bio-sensors, mapping RNA protein through, 8–9
Bit error rate (BER), 159
Blackbody error, 516
Blackbody radiation, 334–335
Black-light (long-wave), 21, 23
Blocking diodes, 91
Blood velocity, measuring, 370
Blue (Gallium Nitride) chip, 27
Boiler-drum water, liquid level sensor for, 354
BOM (bill of materials) file, 250
Box detection, 146
Boxcar apodization, 508
Break even analysis, 529–531
Breast cancer, and sensory MEMS, 355–356
Bridges, 219–220, 270
Brightness (luminance), 34–35
Broadband system, for MAP, 217–218
Buckling, 344
Budget, financial, 526
Building blocks, CIM, 263–264
Bulk material level control, 64
Bulk optical systems, 370
Bundles, fiber-optic
  design considerations, 167–168
  overview, 165–167

Burden current, 81
Burn-in philosophy, 256
Bus system, 440
Business management, 231, 244–245. *See also* Financial planning

C

Cable influences, eliminating by special design, 343
Cable/wire stub terminations, 88
CAD (computer-aided design), 236, 248, 250–251, 558
Cadmium sulfide (CdS) sensors, 292
CAE (computer-aided engineering), 236, 248, 250–251
Calibration
  defined, 490
  factory, 348
  industrial sensors and control, 344–345
  (*See also* Radiometric calibration)
Calibration drift, 515, 517–519
Calibration source errors, 515–517
CAM (computer-aided manufacturing), 248, 250–251
Canadian Standards Association (CSA), 151
Candela, 32
Candlepower (luminous intensity), 32–39
Cantilever-type acceleration sensor, 372
Capacitance, 75
Capacitive proximity sensors
  features of, 77
  versus inductive sensors, 62
  in manufacturing
    circuit accessories, 87–88
    control circuits, 85–87
    output circuits, 85–87
    overview, 80–81
    relays, 81–82
    response time, 93–97
    switching logic, 88–93
    transistor DC switches, 84–85
    triac devices, 82–84
  nano-positioning, 340–343
  overview, 75
  principles of operation, 75–77
  range, 77–78
  surrounding conditions, 79
  target material, 78–79
  target size, 78–79
  typical applications of, 64
Capek, Josef, 418
Capek, Karel, 417–418
Capital budget, 526, 534, 544
Capital investment, application of present values to, 545
CAPP (computer-aided process planning) programs, 237
Carbon-fiber-reinforced plastic (CFRP), 367
Carrier-band system, for MAP, 219
Carrier-sense multiple-access with collision detection (CSMA/CD), 216
Carrying capacity (payload), 422
Cartesian coordinate system, 425–426, 458–459
Cash flow, 547
CCDs (charge coupled devices), 302–303, 307
CDF (consolidated design file), 266–267
CdS (cadmium sulfide) sensors, 292
Central control units, 448–449
Centralized control system, 483
CEO (chief executive officer), 247
CFRP (carbon-fiber-reinforced plastic), 367
Chain-termination method, 7
Channel spectrum error, 520–521
Characterisation, 511
Charge coupled devices (CCDs), 302–303, 307
Chemical excitation, 117–118
Chemical plants, liquid level sensors in, 353–355
Chief executive officer (CEO), 247
Chromaticity, 10–15, 18, 284
CIE (Commission Internationale de l'Éclairage), 10, 13, 285
CIM. *See* Computer-integrated manufacturing; Networking, manufacturing
CIM DB (computer-integrated manufacturing database), 252–253
CIM-OSA (computer-integrated manufacturing open systems architecture), 271
Circuits, accessories for sensor, 87–88
Circular variable filter radiometers, 492
Closed-loop control systems, 49–51, 326
Clusters, LED, 19
CNC (computer numerical control) systems, 466–468
CNC/DNC (computer numerical control and direct numerical control), 248
Coaddition, 490
Coaxial cable, 210
Coding, 461
Color comparators, 286–287
Color gamut, 15
Color interpretation, 284
Color measurement, units of, 285–286
Color rendering index (CRI), 10–15, 18–19
Color sensor algorithms, 287–288

Color spectrophotometers (color meters), 285
Color temperature (chromaticity), 10–15, 18, 284
Color theory, 283–285
Commercial crack detection sensors, 367–368
Commission Internationale de l'Éclairage (CIE), 10, 13, 285
Communication, workplace
  business financial status, 541–542
  controllable and uncontrollable costs, 538–540
  cost and profit data, analyzing, 540–541
  overview, 535–536
  report construction, 540
  reporting strategy, 537–538
  responsibility centers, 538
  timeliness, 536–537
Communication interfaces, CIM, 261–262
Communications
  CIM, 264
  computer, 207–212
  data transmission
    parallel, 478–480
    serial, 480–482
  decentralized manufacturing systems, 483–486
  fiber optics principles, 155–157
  indexing for manufacturing applications
    linear, 475–477
    synchronous, 477–478
  industrial handling, 469–473
  input control sensors, 440–444
  microcomputer interactive development system, 444–445
  NC controllers
    absolute control, 457–458
    auxiliary commands, 466
    CNC system, 466–468
    geometric commands, 464–465
    machining program, 452–457
    manufacturing procedure and control, 451
    motion commands, 465–466
    operation of system, 462–464
    overview, 450–451
    software, 458–462
  overview, 439
  packaging technology, 473–475
  personal computers as single-board computers, 445–450
  plant floor, 264–266
  single-board computers, 439–440
Complementary output sensor, 86
Compliance, defined, 423
Computer communications
  application layer, 209
  data link layer, 210
  network layer, 210
  networks based on OSI, 211–212
  overview, 207–209
  physical layer, 210–211
  presentation layer, 209
  session layer, 210
  transport layer, 210
Computer networks, based on OSI, 211–212
Computer numerical control and direct numerical control (CNC/DNC), 248
Computer numerical control (CNC) systems, 466–468
Computer systems, CIM, 261
Computer-aided design (CAD), 236, 248, 250–251, 558
Computer-aided engineering (CAE), 236, 248, 250–251
Computer-aided manufacturing (CAM), 248, 250–251
Computer-aided process planning (CAPP) programs, 237
Computer-controlled manufacturing system, 7
Computer-controlled multisensor assembly station, 412
Computer-integrated manufacturing (CIM)
  analysis and design, 254–256
  data acquisition, 257–262
  decision support system, 252–254
  design with sensors and control systems, 246–252
  manufacturing enterprise model
    business management, 244–245
    engineering and research, 235–238
    marketing, 234–235
    overview, 232–234
    physical distribution, 242–244
    plant operations, 239–242
    predictive monitoring sensors, 317–318
    production planning, 238–239
      modeling plan, 230–232
      overview, 5–6, 229–230
  role of sensors and control technology in, 229–230
  strategy with emphasis on role of sensors
    building blocks, 263–264
    communications, 264
    integration requirements, 269–273
    managing data, 266–267
    overview, 262–263
    plant floor communications, 264–266
  presentation, 267–269. *See also* Networking, manufacturing

Computer-integrated manufacturing database (CIM DB), 252–253
Computer-integrated manufacturing open systems architecture (CIM-OSA), 271
Configurations, fiber optics
  analog links, 161
  bundles, 165–168
  digital links
    carrier-based, 160–161
    pulsed, 159–160
  fiber pairs for remote sensing, 168–169
  liquid level sensing, 169–170
  optical power budget, 158–159
  overview, 157–158, 165
  video links, 161–162
Confocal Microscopy sensors, 110–113
Confocal profiling characterization systems, 110–111
Connectivity analysis, 417
Consolidated design file (CDF), 266–267
Contact life, relay, 81
Contact sensors (limit switches), 79–80, 133, 411
Continuous web fabrication, controlling input/output speed of, 369–370
Contrast, 148
Control circuits, 85–87
Control computer, multisensor-controlled robot assembly, 415
Control processes, classification of, 48–49
Control technology, personal computers, 446
Controllable costs, 538–540
Controllers, 325, 480
Convergent sensing mode, 154
Conveyer system control, 63
Cool white, 25
Copper, 112
Corner-cube reflectors, 138, 139
Corporate manufacturing strategy, 46
Correction factors, target, 71
Cosine law, 36
Cost accounting, 245
Costs
  analyses, 528–529, 540–541
  fixed, 528
  holding, 547
  ordering, 547
  sensor evaluation criteria, 256
  variable, 528
Counters, 261, 449
CPU card, 261
Crack detection sensors, 367–368
CRI (color rendering index), 10–15, 18–19
Critical path method (CPM), 554
Critical performance parameters, 362–363
Cross-sectioning, 110
Crosstalk, 141
Cryogenic manufacturing applications, LVDT, 410
CSA (Canadian Standards Association), 151
CSMA/CD (carrier-sense multiple-access with collision detection), 216
CTGA technique, 8
Cup-shaped coil/core assembly, 67
Current monitor, fiber-optic, 188
Current-to-voltage conversion, 293
Customer groups analysis, 541
Customizable solutions, CIM, 272
Cytokines, 355–356

D

D/A (digital-to-analog) converter, 260
Da Vinci, Leonardo, 418
DAE (Distributed Automation Edition), 264–265
Daisy chain networks, 162–163, 181
Damped condition, 66
Dark current, 365–367
Dark noise, 503–504
Dark-to-light energizing, 52–53
Data acquisition, CIM
  actuators, 257–259
  analog input, 259–260
  analog output, 260
  communication interfaces, 261–262
  computer systems, 261
  counter/timer, 261
  digital input and output, 260
  overview, 257
  real-world phenomena, 257
  sensors, 257–259
  for sensors and control hardware, 259–261
  signal conditioning, 259
  software, 262
Data bus networks, 162–165
Data Communication Service (DCS), 266–267
Data integration, CIM, 270
Data link layer (DL), 208–210
Data rate, 159
Data transmission
  parallel, 478–480
  serial, 480–482
Database management systems (DBMSs), 246
Daylight, 18–19
DBm measurement, 173
DC (direct current) switches, transistor, 84–85

DCS (Data Communication Service), 266–267
De Vaucanson, Jacques, 418
Decentralized manufacturing systems, 483–486
Decision making, 529–535
Decision support systems (DSS), 246
Decoding genome, 8
Deflection, pressure sensors and, 337
Degrees of freedom, 422
Density measurements, robot, 434–436
Detailed balance of power, 502
Detection cycle, 66
Detection methods, 58–61. *See also specific detection methods by name*
Detector circuit, 66, 149
Detector current, 498
Detector nonlinearity errors, 520
Devol, George, 419
Diagnostic systems, 196–199
Dielectrics, 75
Diffuse-reflection. *See* Proximity detection
Digital encoder sensors, 278–280
Digital input and output, 260
Digital links
  carrier-based, 160–161
  pulsed, 159–160
Digital-to-analog (D/A) converter, 260
Direct costing, 529
Direct current (DC) switches, transistor, 84–85
Direct numerical control (DNC) systems, 379
Direct reflection. *See* Proximity detection
Direct scanning (beam break mode), 145
Direction of motion, detecting, 98, 105–106, 279
Directional encoders, 279
Displacement sensors, 348–350, 412
Distributed Automation Edition (DAE), 264–265
Distribution channels, 541
Distribution of service robots, 435–436
Divergence, 490, 492
Divergent sensing mode, 153
DL (data link layer), 208–210
DNC (direct numerical control) systems, 379
Door-opening sensors, 105
Doppler frequency, 100
Double pole, double throw (DPDT) relay, 52
Drilling machine, 450–451
Drive, robot, 422–423
Driver-level programming, 262
Drivers, LED, 20
Driving inspection requirements, Confocal Microscopy sensors, 111–113
DSS (decision support systems), 246
Dual laser test sets, 175–179
Dual theory of light, 113
Dual-level control, 441
Dwell range, 95
Dwell time, 94
Dynamic machine part (bearing) fault sensors, 298–300
Dynamic manufacturing operation, 559

E

E (irradiance), 491–492
EB (electron beam) welding, 410
Economic interests
  communication, 535–542
  financial planning, 523–529
  future of sensors and control systems, 556–563
  information assimilation and decision making, 529–535
  mathematical methods for planning and control
    capital budget, 544
    constraints, 549–551
    cost of fixed asset, 544
    expected net cash flow, 544
    inventory analysis, 547–548
    linear programming, 548
    objective probability, 542–543
    opportunity cost, 544
    present value, 544–547
    profit function, 549
    project management, 551–555
    simulation, 555–556
    subjective probability, 543–544
    uncertainty, 542
  overview, 523
Economic order quantity (EOQ), 547–548
Edge detection and extraction, 391–395
EDI (electronic data interchange), 235
EDP (electronic data processing), 536
Effectors, 325
EIA (Electronic Industries Association), 213
Eight-port active star coupler, 165
8000XG fiber identifier, 178–179
Electrical excitation, 116–117
Electrical handshaking, 160
Electrically programmable ROM (EPROM), 444
Electrode geometry, 343
Electromagnetic identification of manufactured components, 277–278

Electromagnetic interference (EMI), 30, 151, 364
Electromotive force (emf), 197
Electron beam (EB) welding, 410
Electronic assemblies, sensor, 293
Electronic data interchange (EDI), 235
Electronic data processing (EDP), 536
Electronic Industries Association (EIA), 213
Electronic SAR sensors utilizing vision technology, 305–307
Electronics industry, 119–122
Electronics support, 345–348
Electrooptic links, networking with, 179–184
Elektro, 418
Embeddable sensors, 73, 79
Emf (electromotive force), 197
EMI (electromagnetic interference), 30, 151, 364
Encapsulation, LED, 28–29
End effector sensors
  detecting partially visible objects, 395–400
  for edge detection and extraction, 391–395
  linear variable-displacement transformer, 407–411
  overview, 424–425
  sound-vision recognition, 400–407
  ultrasonic, 400
Endoscopes, as image transmission sensor, 372–373
Engelberger, Joseph F., 419
Engineering
  computer-integrated manufacturing, 231
  facilities, 237
  integrated system, 48
  management, 237–238
  overview, 235–236
  process development, 237
  product development, 236
  release control, 237
Enterprise model
  business management, 244–245
  engineering, 235–238
  marketing, 234–235
  overview, 230, 232–234
  physical distribution, 242–244
  plant operations, 239–242
  production planning, 238–239
  research, 235–238
Enterprise planning, 232, 245
Enterprise services, 245
Environment presentation, CIM, 267–269
Environments, excess gain and, 61
EOQ (economic order quantity), 547–548
Epitaxial silicon APD (EPI-APD), 361
Epoxy resin enclosure, LED, 25
EPROM (electrically programmable ROM), 444
Error voltage recognition facility, 448
Etch defects, 112
Ethernet, 215–216
Excess gain, 60–61, 145–147
Excess noise factor, 365–367
Excessive marking, probe, 112
Excitation mechanism, laser, 115–118
Executors, 416–417
Expandability, 256
Expected net cash flow, 544
Expenses, 46
Experimental techniques, 204
Extended range core and coil, 68
Extended source, 490
Extended straight-line control, 454
Extreme environments, 409–410
Extrinsic sensors, 328
Eye sensitivity, maximum, 37

F

Facilities engineering, 237
Fact tables, 441
Factory calibration, 348
Failure modes, probe, 111–112
FAMULUS robot, 420
Far output, 140–141
Fast food drive-thru sensors, 104–105
Fast Fourier transform (FFT) spectrometer, 492–493
Fault finders, 177–178
Fault sensors. *See* Machinery fault sensor networks
Fault-detection systems, power line, 375–376
Federal Communications Commission (FCC), 99
Female pin connector receptacle, 88–89
Ferdig, Mark, 4
Ferrous targets, 69
FFT (fast Fourier transform) spectrometer, 492–493
Fiber identifiers, 178–179
Fiber nervous surveillance system, 368
Fiber optics
  bifurcated, 143–145
  communications principles, 155–157
  configurations, 157–162, 165–170
  data bus networks, 162–165
  data link layer communication, 210
  dual laser test sets, 175–179
  flexibility of, 170–172
  individual, 143
  inductive proximity sensors, 148–150
  in industrial communication and control, 155

Fiber optics (*Cont.*):
information link, 157
laser Doppler velocimeter, 369
LEDs, wavelengths of, 152
light sources, testing, 172–174
limit switches, 150–151
liquid level sensing, 169–170
networking with electrooptic links, 179–184
noncontact metal detection, 148–150
overview, 133, 141–143
parameters, 145–148
photoelectric sensors
automated guided vehicle system, 140–141
LEDs, 134–136
overview, 133–134
polarized reflex detection, 138–139
proximity detection, 139–140
reflex photoelectric controls, 137–138
through-beam, 136–137
position sensors, selecting, 151–152
power meters, 173–174
pressure sensors, 338–340
sensor alignment techniques, 152–155
sensor network architecture, 373
sensory links for minicell controllers, 184
terminations, 170–172
testing, 172–174
versatility in industrial applications, 184–189
video links, 162
Fiber pairs, for remote sensing, 168–169
Field of view, 490
Field stop (FS), 490
Financial budget, 526
Financial management, 244
Financial planning
balance sheet projections, 526
capital budget, 526
controlling operation, 527–528
cost and profit analysis, 528–529
financial budget, 526
income statement projections, 526
overview, 244, 523–524
plan development, 524
planning for profit, 527
product budget, 524–525
sales budget, 524
selling and administrative budget, 525–526
Finish, sensor surface, 343
5200 series optical fault finders, 177–178
Fixed asset costs, 544
Fixed costs, 528
Fixed-focus proximity (convergent sensing) mode, 154
Flags, 449
Flash memory, 10
Flatness, sensor surface, 343
Flexibility, of fiber optics, 170–172
Flexible manufacturing cells
detecting tool failure, 204–207
overview, 196–197, 199–201
software problems, 201–204
Flexible manufacturing systems (FMSs)
current technology, 381–383
end effector sensors
detecting partially visible objects, 395–400
for edge detection and extraction, 391–395
linear variable-displacement transformer, 407–411
sound-vision recognition, 400–407
ultrasonic, 400
evaluation of, 199
number of products in, 193
overview, 379–380
robots
control through sensors, 411–412
density measurements, 434–436
guidance with vision system, 389–391
history of, 417–419
invention of, 419–426
machine-vision sensing, 383–388
market projections, 436–437
multisensor-controlled assembly, 412–417
programming, 426–428
vision locating position, 388–389
world market, 428–434
role of sensors in, 380–383
Flexure-guided nano-positioning systems, 345
Flow sensors, 351–352
Fluorescent dye, 321
Fluorescent light, 18–19, 357
Flush-mounted sensors, 73
Flying metal chips, 73, 79
FMSs. *See* Flexible manufacturing systems
Focus feature, 399
Focused proximity sensors, 139–140
Food industries, 127–128
Foot-candle, 18, 34–35
Foot-lambert, 35
Force sensors, 310–312, 557
Force-control loop, 312
Fourier transformation, 299
Frame buffer (image buffer), 390

Frequency-shift keying (FSK), 160
FS (field stop), 490
Full spectrum, 18
Full-duplex fiber-optic link, 304
Function chart, PLC program, 471–472
Function faults, software, 202
Fuzzy logic color sensors
  color comparators, 286–287
  color sensor algorithms, 287–288
  color theory, 283–285
  controller flowcharts, 296–298
  design considerations in, 295–298
  overview, 282
  sensing principles, 282–283
  true color measuring instruments, 286–287
  units of color measurement, 285–286

G

Gain, APD, 364–365
Gakutensoku robot, 418–419
Gallium Nitride (blue) chip, 27
Gamma, 15
Gamut, 10–11
Gas, on-line measuring and monitoring, 356–359
Gateway nodes, 222
Ge. *See* Germanium
GeAPD (germanium avalanche photodiode), 331
Geiger mode, 367
General Motors Corp., 216
General-purpose interface bus (GPIB), 261
Generated ultrasonic beam, 371
Generation of process signals, in decentralized manufacturing systems, 483–486
Genome, 7–8
Geographical areas analysis, 541
Geometric NC commands, 464–465
Germanium (Ge)
  APD, 364
  photodiodes, 291
Germanium avalanche photodiode (GeAPD), 331
Germicidal (short wave) light, 23
Ghost images, 521
Glass optical fibers, 142
Glass-fiber-reinforced plastic (GFRP), 367
Global network data flow, 226
GPIB (general-purpose interface bus), 261
GRIN lens, 331
Grippers, 310–311
Ground speed measurement, tractor, 104
Group technology (GT) applications, 236
Guidance, of robots with vision system, 389–391
Gunn diode, 99

H

H (hysteresis), 66, 77
Hand-eye coordination sensors, 309–310
Handling, industrial, 469–473
Hardware considerations, CIM computer systems, 261
Hardware-level programming, 262
Headend remodulator, 217
Heavy-duty duplex fiber-optic cables, 186–187
Helium-neon laser, 115
Hemispherical concave reflector, 316
Hermetic sealing, 411
Heterogeneous systems, 271
Hexadecimal counting system, 442
High-acidity sensor, 317–318
High-clad fiber-optic cables, 185–188
High-precision image display and measurement, 307–308
High-speed press machines, 48
High-temperature sensors, 330
Holding costs, 547
Holographic scanners, 277
Horizontal synchronous rotary indexing table, 478–479
Hue, saturation, and intensity (HSI), 283
Human eye sensitivity, 37
Human Genome Project, 7, 322
Human vision, 384
Hybrid fiber networks, 182–184
Hybrid method, genome sequencing, 8
Hybrid wire/fiber networks, 180
Hybrids
  photoconductive sensors, 293
  of photomultipliers options, 295
Hyphenated Systems, 110
Hysteresis (H), 66, 77

I

I (intensity), 19, 27, 490
IBM DCS/CDF, 266–267
Ideal equipment units, 192
Identification of manufactured components, 275–278
IEC (International Electrotechnical Commission), 151
IEEE (Institute of Electrical and Electronics Engineers), 216
IGES (initial graphics exchange specifications), 271

Illuminant, 285
Illumination, 34–35
ILS (instrument line shape), 506–509
ILS (integrated linearization system), 341–342, 348
Image buffer (frame buffer), 390
Image fibers, 373
Image transformation, 383–384
Image transmission sensor, endoscope, 372–373
Imaging camera sensors, reflective strip, 315–317
Imbedded information, 230
Improved business process, 273
Incandescent lights, 18–19
Income statements, 526
Incremental analysis, 532
Incremental control systems, 454–456
Incremental optical encoder sensor, 280
Incremental positioning data, 452
Independent safety monitoring equipment, 103–104
Indexing for manufacturing applications, 475–478
Indian Ocean Tsunami, 3
Indium gallium arsenide (InGaAs), 174, 291, 330
Induced eddy current, 65
Inductive proximity sensors
  versus capacitive sensors, 62
  in manufacturing
    circuit accessories, 87–88
    control circuits, 85–87
    output circuits, 85–87
    overview, 80–97
    relays, 81–82
    response time, 93–97
    switching logic, 88–93
    transistor DC switches, 84–85
    triac devices, 82–84
  overview, 64, 133
  principles of operation, 64–66
  range, 67–68
  sensing distance, 68–69
  surrounding conditions, 73–75
  target material, 69–71
  target shape, 71–72
  target size, 69–71
  typical applications of, 62–64
  variation between devices, 72–73
Industrial handling, 469–473
Industrial PCs (IPCs), 251–252, 261
Industrial robots. *See* Robots
Industrial sensors and control
  cable influences, eliminating, 343
  calibrating, 344–345
  crack detection sensors, 367–368
  displacement sensors for robotic applications, 348–350
Industrial sensors and control (*Cont.*):
  electrode geometry, 343
  electronics support, 345–348
  endoscope, as image transmission sensor, 372–373
  fiber optics, 184–189
  installation, 348
  integrated linearization system, 348
  laser
    automatic machinery, 128–131
    automotive manufacturing, 122–125
    electrical and electronics, 119–122
    food processing and packaging, 127–128
    metal/steel/nonferrous, 125–127
  LDV sensors, 369–370
  manufacturing
    network architectures, 373–375
    overview, 328–329
  materials achieving accuracy, 343–344
  measuring ranges, 344–345
  measuring surfaces, parallelism of, 345
  mounting, 344–345
  nano-positioning capacitive metrology sensors, 340–343
  overview, 325–328
  power line fault-detection systems, 375–376
  pressure sensors, 337–340
  process control sensors
    for acceleration, 371–372
    liquid flow, 350–355
    temperature sensors, 329–336
  semiconductor displacement, 119
  sensory MEMS, 355–356
  spectroscopy, 356–359
  surface, sensor, 343
  ultrasonic/laser nondestructive evaluation sensor, 370–371. *See also* Avalanche photodiode detectors
  Industry analysis, 541
Industry standards, 271–272
Information assimilation, 529–535
Infrared (IR)
  gallium arsenide LED, 135
  light domain, 294
  scanners, 277
InGaAs (indium gallium arsenide), 174, 291, 330
Initial graphics exchange specifications (IGES), 271
Input control sensors, 440–444
Input modules, 447
Input/output (I/O) board, 257
Input/output (I/O) devices, 439
Input/output (I/O) ports, 164–165

Input/output speed of continuous web fabrication, 369–370
InSb detectors (Johnson noise), 499–503
Insertion machine, 475
Inspection, 241, 389–390
Installed base investment, CIM, 270–271
Institute of Electrical and Electronics Engineers (IEEE), 216
Instrument line shape (ILS), 506–509
Integrability, sensor, 256
Integrated circuits, photoconductive sensor, 293
Integrated linearization system (ILS), 341–342, 348
Integration
  computer networks, 214
  computer-controlled devices, 258
  computer-integrated manufacturing, 269–273
Intensity (I), 19, 27, 490
Interface faults, 203–204
International Electrotechnical Commission (IEC), 151
International Standards Organization (ISO), 207–208
Interoperability, MAP, 217
Intrinsic linearity, 515, 519
Intrinsic sensors, 328, 359
Inventory analysis, 547–548
Inventory management, JIT, 246
Inverse square law, 36
I/O (input/output) board, 257
I/O (input/output) devices, 439
I/O (input/output) ports, 164–165
Ionization ration, 367
IPCs (industrial PCs), 251–252, 261
IR. *See* Infrared
Irradiance (E), 491–492
ISO (International Standards Organization), 207–208
Iv (luminous intensity value), 28

J

Jacketed glass fibers, 142
Jitter noise, 504
Johnson, J. B., 499
Johnson noise (InSb detectors), 499–503
Just-in-time (JIT) inventory management, 246

K

Kelvin (K) color temperature, 19
Keratosis, 298
Kinematics, 422
Knowledge systems (KS), 254

L

L (radiance), 491, 492
L.a.b. numbering system, 285
Lambert, 35
Language system (LS), 253–254
LANL (Los Alamos National Laboratory), 317
LANs (local area networks), 215, 246
Laparoscopic robotic surgery machine, 421
Laser Doppler velocimeter (LDV), 369–370
Laser light, properties of, 114
Laser scanning applications, 278
Laser sensors
  active medium, 114–115
  excitation mechanism, 115–118
  feedback mechanism, 118
  flexible manufacturing systems, 382
  industrial applications
    automatic machinery, 128–131
    automotive manufacturing, 122–125
    electrical and electronics, 119–122
    food processing and packaging, 127–128
    metal/steel/nonferrous, 125–127
  nondestructive evaluation, 370–371
  output coupler, 118–119
  overview, 113–114
  properties of laser light, 114
  range-finding, 382
  semiconductor displacement, 119
Laser-diode sources, 158
Latching current, APD, 367
Lateral (side by) approach, 71–72
LDV (laser Doppler velocimeter), 369–370
Lead time, 46
Lead-by-the-nose technique, 424
LEDs. *See* Light-emitting diodes
LEL (lower explosion limit) concentration, 358
LFF (local-feature-focus) method, 399
Lidové noviny journal, 418
Light amplification, by stimulated emission of radiation, 10
Light energy, 10
Light pipes, 142
Light sources, fiber optics, 172–174
Light-emitting diodes (LEDs)
  color chart, 15–17
  design currents, 30
  EMI countermeasures, 30
  heat dissipation, 30
  indicators, 87
  intensity of, 27
  monochromatic, 26
  non-phosphor white, 30–33
  operating life, 29–30

Light-emitting diodes (LEDs) (*Cont.*):
overview, 19–23, 25–26
photoelectric sensors, 134–136
task lighting-lamps, 23–25
visibility, 28–29
voltage, 30
wavelengths, commonly used, 152
white light, 26–27
Limit switches, 79–80, 133, 411
Limited-capability control loops, 325
Line of purples, 13
Line powered control circuits, 80
Linear arrays, 278, 382
Linear compression, 338
Linear indexing, for manufacturing applications, 475–477
Linear instruments, 510
Linear multiple point calibrations, 512–513
Linear programming, 548
Linear variable-displacement transformer (LVDT) sensors, 407–411
Linearity, improving with factory calibration, 348
Line-powered burden current, 81
Liquid flow sensors, 350–355
Liquid level sensors
capacitive proximity sensors, 64
fiber-optic, 169–170
process control for petroleum and chemical plants, 353–355
Listeners, 480
Load powered control circuits, 80
Load release time, 94
Load response time, 94
Load-powered residual current, 80
Local area networks (LANs), 215, 246
Local-feature-focus (LFF) method, 399
Logic ring, 221
Long-wave (black-light) light, 21, 23
Los Alamos National Laboratory (LANL), 317
Low voltage LEDs, 20
Lower explosion limit (LEL) concentration, 358
Low-pass filter, 522
Low-temperature sensors, 330
L_{pk} (peak wavelength), LED, 26
LS (language system), 253–254
Lumens, 19–20, 33, 37–38
Luminance (brightness), 34–35
Luminous flux, 33–34, 38
Luminous intensity (candlepower), 32–39
Luminous intensity value (Iv), 28
Lung tumors, sensors for, 320–321
Lux, 20, 35
LVDT (linear variable-displacement transformer) sensors, 407–411

M

MA (milliamps), 21
Machine control, with inductive proximity sensors, 63
Machinery fault sensor networks
acoustics, 198
diagnostic systems, 196–199
dynamic machine part fault sensors, 298–300
flexible manufacturing cell evaluation sensor networks, 199–207
motor current signature analysis, 197–198
overview, 196
quantifying workpiece quality, 199
resonance analysis, 197
temperature, 198
vibration analysis, 197
Machine-vision sensing, 383
Machining programs, 452–457
Macotech Machine Tool Co. Inc., 556
Magellan spacecraft, 1–2
Magnetic fields, 74, 79
Main article–robotics suite, 423
Main processor unit (MPU), 415
Maintenance
plant, 242
sensor, 256
Make-or-buy-analysis, 533
Male pin connector receptacle, 88–89
Manufacturing automation protocol (MAP)
bridges, 219–220
broadband system, 217–218
carrier-band system, 219
overview, 216–217
token systems, 220–221
Marketing, 234–235
Mass flow, regulation of, 50
Master industrial terminal, 170–171
Master production planning, 238
Masuoka, Fujio, 10
Matching faults, software, 201–202
Material receiving function, 240
Material transfer, 242
Materials
accuracy, 343–344
measuring dynamic changes, 312–314
Mathematical planning and control methods
capital budget, 544
constraints, 549–551
cost of fixed asset, 544
expected net cash flow, 544
inventory analysis, 547–548
linear programming, 548
objective probability, 542–543

Mathematical planning and control methods (*Cont.*):
opportunity cost, 544
present value, 544–547
profit function, 549
project management, 551–555
simulation, 555–556
subjective probability, 543–544
uncertainty, 542
Max rated temperature (operating temperature), 21
Mcd (millicandela), 21
MCPP (microchannel plate photomultiplier), 322
Mean spherical candlepower (MSCP) values, 28
Mean time between failures (MTBFs), 80, 256
Mean time between operator interventions (MTOI), 192, 194
Mean time of intervention (MTI), 192, 194
Mean time of processing, 194–196
Mean time to repair (MTTR), 261
Measured surface, 400
Measurements
at high temperatures, 411
multichannel, 348
Measuring
dynamic changes in materials, 312–314
ranges, industrial sensors and control, 344–345
velocity, with microwave sensors, 105
Measuring surfaces, parallelism of, 345
Mechanical convergence, 154–155
Mechanical limit switches, 61
MEMS (microelectromechanical systems), 5, 355–356
Mercy Corps, 4
Message processing, computer network, 208
Metal industries, 125–127
Metallic targets, 69
Metrological sensors, 557
Metrology Laboratories, 343
Microbend sensor, 350
Microchannel plate photomultiplier (MCPP), 322
Microcomputer interactive development system, 444–445
Microcomputers, 248, 251–252
Microelectromechanical systems (MEMS), 5, 355–356
Microfluor detector, 319
Micro-radian-range nano-positioning systems, 346
Microrobotic motion, sensors for hand-eye coordination of, 309–310
Microscopic fiber flaws, 167
Microscopy, 5, 112
Microsupercomputers, 252
Microwave integrated circuit technology (MICT), 109
Microwave sensors
characteristics of, 98–99
direction of motion, 105–106
measuring velocity with, 105
motion, 100–104
overview, 98
presence, 104–105
principles of operation, 99–100
range, 106–109
technology advancement, 109
Military crack detection sensors, 367–368
Military Microwave Integrated Circuit (MIMIC) program, 109
Milliamps (mA), 21
Millicandela (mcd), 21
Milling machine, 450
Minicell controllers, fiber-optic sensory links for, 184
Mirrors, for feedback mechanism, 118
MIT arm, 420
Modulation concept, 160
Monitor program, 439
Monochromatic IR stimulus, 506
Monochromatic LEDs, 26
Motion, sensors for hand-eye coordination of microrobotic, 309–310
Motion control
defined, 422
inductive proximity sensors, 62
Motion detection
inductive proximity sensors, 62
with microwave sensors, 98, 100–104
Motion NC commands, 465–466
Motor current signature analysis, 197–198
Mounting industrial sensors and control, 344–345
Movement, industrial robot, 425–426
MPU (main processor unit), 415
MSCP (mean spherical candlepower) values, 28
MTBFs (mean time between failures), 80, 256
MTI (mean time of intervention), 192, 194
MTOI (mean time between operator interventions), 192, 194
MTTR (mean time to repair), 261
Multichannel fiber-optic cables, 186
Multichannel measurements, ILS, 348
Multi-chip packages, Confocal Microscopy sensors, 113
Multiobjective sensor selection, 256
Multiobjective structure, decision support system, 253–254

Multiple point calibrations, 512
Multiple-pole switch, 80
Multiple-ring digital communication network, 221–223
Multiplexing, FFT spectrometry, 493
Multipurpose advanced industrial robot installations, 435
Multisensor-controlled robot assembly, 412–417

N

NA (numerical aperture), 156
Nano-capacitive positioning sensors, 340–343
Nano-crystals, 9–10
Nanometers (nm), 22
Nano-positioning capacitive metrology sensors, 340–343
Nano-sensors, 9
Nanowatts power, 295
National Cancer Institute, 321
National Center for Manufacturing Sciences (NCMS), 9
National Electrical Manufacturers Association (NEMA), 151
National Science Foundation (NSF), 9
NC. *See* Numerical control
NC (normally closed) switches, 79–80
NCMS (National Center for Manufacturing Sciences), 9
Near infrared (NIR). *See* SpectRx NIR technology
Near output, 140–141
Negative volume changes, 406
NEMA (National Electrical Manufacturers Association), 151
NEP (noise equivalent power), 363
NESR (noise equivalent spectral radiance), 505–506
Network architectures for manufacturing sensors, 373–375
Network layer (NL), 208–210
Network load, 215
Network plans, 551
Network topologies, 374
Networking, manufacturing
  AbNET, 221–223
  computer communications, 207–211
  with electrooptic links, 179–184
  flexible systems, number of products in, 193
  manufacturing automation protocol, 216–221
  network types
    Ethernet, 215–216
    overview, 213
    RS-232-based, 213–215
    TCP/IP, 216
  overview, 191–193
Networking, manufacturing (*Cont.*):
  Satellite Sensor Data Service, 225–226
  tracking sensors, 194–196
  universal memory network, 223–225.
  (*See also* Machinery fault sensor networks)
9XT optical source driver, 172–173
19XT optical attenuator, 177
NIR (near infrared). *See* SpectRx NIR technology
Nishimura, Makoto, 419
NL (network layer), 208–210
Nm (nanometers), 22
NO (normally open) switches, 79–80
Noise, sources of, 498
Noise current, 365–367
Noise equivalent power (NEP), 363
Noise equivalent spectral radiance (NESR), 505–506
Noise factor, excess, 365–367
Nominal sensing range, 68
Noncontact sensors, 109, 334–336, 411
Nondestructive evaluation sensors, 370–371
Nonembeddable sensors, 73
Nonferrous industries, 125–127
Nonferrous targets, 69
Nonlinear multiple point calibrations, 513–515
Nonlinearity, 346
Non-phosphor white LEDs, 29–33
Nonshielded inductive proximity sensors, 149
Normally closed (NC) switches, 79–80
Normally open (NO) switches, 79–80
NPN (sinking) parallel sensor arrangement, 90–91
NPN transistor, 84
NSF (National Science Foundation), 9
Numerical aperture (NA), 156
Numerical control (NC)
  absolute control, 457–458
  auxiliary commands, 466
  computer numerical control system, 466–468
  geometric commands, 464–465
  machining program, 452–457
  manufacturing procedure and control, 451
  motion commands, 465–466
  operation, 462–464
  overview, 450–451
  programming, 237
  software, 458–462
Nyquist, H., 499

O

Object view spectral radiation, 495–497
Objective probability, 542–543
OCR (optical character recognition), 278
OEMs (original equipment manufacturers), 192
OFF delay logic, 96–97
ON delay logic, 96–97
One-shot delay, 97
On-line measuring and monitoring of gas, 356–359
ON/OFF delay logic, 97
Open coil without core, 67
Open interfaces, CIM, 271–272
Open system interconnect (OSI) model, 207, 211–212
Open-loop control systems, 49–51
Operating life, LED, 22, 29–30
Operating temperature (max rated temperature), 21
Operational level, CIM, 247
Opportunity cost, 544
Opposed sensing mode, 144, 152
Optic ground wire (OPGW), 375–376
Optical character recognition (OCR), 278
Optical encoder, 279
Optical fibers, 145–148, 305
Optical filter, 522
Optical power budget, 158–159
Optical power meter, hand-held, 174
Optical sensors, 310–312, 317–318, 382, 557
Optical spectrometer, 356
Optical time-domain reflectometry (OTDR), 172, 376
Optical-fiber flow sensor, 351–352
Optical-fiber pyrometer, 336
Optical-fiber sensors, 328–329
Optical-fiber thermometer, 333
Optimum detectors in light sensors, 288
Optimum range, proximity and reflex sensors, 60
Optoelectronic systems
  feedback signals for servomotors, 303–305
  fuzzy logic for, 282–288
  sensor tracking targets, 302–303
  vision associative memory, 307–308
OR function, 89–91
Oral reports, 537
Order size analysis, 541
Ordering costs, 547
Organization theory, 6
Original equipment manufacturers (OEMs), 192
Oscillator circuit, 65, 76
OSI (open system interconnect) model, 207, 211–212
OTDR (optical time-domain reflectometry), 172, 376
Output logic, 86
Output modules, 448
Output resolution of incremental encoder, 282

P

Package-level programming, 262
Packaging machines, 48
Packets, 216
Parallel data transmission, 478–480
Parallel metrology, 342
Parallel sensor arrangement, 90
Parallel-connection logic, 89–91
Parallel-face protective windows, 521
Parallel-type data transmission, 479
Parallelism of measuring surfaces, 345
Parameter values, 195
Part mating function, 413
Partially visible objects, detecting, 395–400
Particle theory, 113
Parts, manufacturing status of, 204–207
Passive backplane and CPU card, 261
Passive optical coupler, 170–171
Passive sensors, 328, 381
Passive star coupler, 164
Path controls, 465–466
Payload (carrying capacity), 422
Payroll, 245
PCBs (printed circuit boards), 22
PCS (plastic clad silica) fiber, 169
PDES (product definition exchange specification), 271
Peak wavelength (L_{pk}), LED, 26
Performance parameters, critical, 362–363
Personal computers, as single-board computers
  central control units, 448–449
  overview, 445–446
  process computer, 450
  programmable logic controllers, 446–448
Personal service robots, 435–436
PERT (project evaluation and review technique), 551–554
Petroleum plants, liquid level sensors in, 353–355
Phased implementation, CIM, 272
PHIGS (programmer's hierarchical interactive graphics standard), 271
PhL (physical layer), 208–211
Phosphors, 15
Photo sensing fluorescence, 7–8

Photoconductive sensors
hybrids, 293
integrated circuits, 293
MCT detectors, 520
overview, 292–293
visible light region, 295
Photodarlingtons, 291–292
Photodetector, 133
Photodiodes, 189, 289–291, 295
Photodyne 9XT optical source driver, 172–173
Photodyne 19XT optical attenuator, 177
Photodyne 21XTL fiber-optic test set/talk set, 176
Photodyne 2260XF switchable dual laser test set, 175
Photodyne 2260XFA laser test set, 175
Photodyne 2285XQ fiber-optic power meter, 173–174
Photodyne 5200 series optical fault finders, 177–178
Photodyne 8000XG fiber identifier, 178–179
Photoelectric sensors
automated guided vehicle system, 140–141
light-emitting diodes, 134–136
manufacturing applications, 52–58
overview, 51, 133–134
polarized reflex detection, 138–139
principles of operation, 51–52
proximity detection, 139–140
reflex photoelectric controls, 137–138
through-beam, 136–137
Photoluminescence, semiconductor temperature detector using, 331–334
Photomultiplier tube (PMT), 289–290, 360
Photon noise, 498
Photons, 113, 289, 498
Phototransistors, 291–292
Photovoltaic InSb detectors, 520
Physical distribution
CIM, 231
operations, 243–244
overview, 242–243
planning, 243
Physical layer (PhL), 208–211
PID (proportional integral-differential) control, 260
Piezoelectric crystals, 337–338
Pigment, 284
Pin connectors, 88
PL (presentation layer), 209
Planck's equation, 496
Plant floor communications, 264–266
Plant level sensors and control systems
automated storage and retrieval systems, 250
computer-aided engineering/design/manufacturing, 250–251
in flexible manufacturing systems, 248–249
in material handling, 250
microcomputers, 251–252
overview, 248
Plant operations, 239–242
Plasma, 117
Plastic clad silica (PCS) fiber, 169
Plastic optical fiber, 142
PLCs (programmable logic controllers), 74, 251, 446–448
PMM (product moment matrix), 301
PMT (photomultiplier tube), 289–290, 360
PNP (sourcing) parallel sensor arrangement, 90
Point source, 491
Point-to-point (PTP) controls, 453, 465
Point-to-point data link, 163
Poissonian statistics, 366
Polarization, 148
Polarized reflex detection, 138–139
Polarized sensor, 53
Police radar, 105
Porphyrin, 320
Ports, 440
Position, locating with robot vision, 388–389
Position encoder sensors in manufacturing, 280–282
Positioning sensors, nano-capacitive, 340–343
Position-sensitive detector (PSD), 119
Position-servo loop, 312
Positive volume changes, 406
Potentiometer, 337
Power at detector, 497–498
Power line fault-detection systems, 375–376
Power meters, 173–174
Power source, 422
Power spectrum, 501
Power supply, LED, 22
PPPL (Princeton Plasma Physics Laboratory), 189
Predictive monitoring method (PREMON), 314–315
Presence, detecting with microwave sensors, 98, 104–105
Present value, 544–547
Presentation, CIM environment, 267–269
Presentation layer (PL), 209
Pressure sensors, 337–340

Primagraphics VME boards, 416
Primary, telescope, 491
Princeton Plasma Physics Laboratory (PPPL), 189
Printed circuit boards (PCBs), 22
Prisms, 353
Private service robots, 435–436
Probe assembly, UDVSS, 313
Probe-mark inspection requirements, Confocal Microscopy sensors, 113
Problem processing systems, 254
Process computer, 450
Process control sensors
  for acceleration, 371–372
  capacitive proximity, 64
  diagram, 439
  inductive proximity, 63
  liquid flow, 350–355
  noncontact, 334–336
  overview, 328–330
  semiconductor absorption, 330–331
  semiconductor temperature detector, 331–334
  temperature detectors using point-contact sensors, 334
Process development, 237
Process signals, in decentralized manufacturing systems, 483–486
Procurement, 239
Product budget, 524–525
Product definition exchange specification (PDES), 271
Product development, 236
Product lines analysis, 540
Product moment matrix (PMM), 301
Product shipping, 242
Production management, 239–240
Production planning
  CIM, 231
  master production planning, 238
  materials, 239
  overview, 238
  plant release, 239
  procurement, 239
  resources, 239
Production process, 241
Production reliability, 46
Productivity, 46
Professional service robots, 435
Profiling characterization systems, 110–111
Profit analyses, 528–529, 540–541
Profit function, 549–550
Program flowchart, 442
Program statements, 443
Programmable controller communication network, 180
Programmable logic controllers (PLCs), 74, 251, 446–448
Programmable output, 86
Programmable Universal Machine for Assembly (PUMA), 420
Programmer's hierarchical interactive graphics standard (PHIGS), 271
Programming faults, software, 203
Project evaluation and review technique (PERT), 551–554
Project management
  CPM, 554
  PERT, 551–554
  queuing analysis, 554–555
Proportional integral-differential (PID) control, 260
Proximity detection
  methods, 59–61
  photoelectric sensors, 139–140
Proximity sensing mode, 144, 153
Proximity sensors, 61–62, 412. *See also* Capacitive proximity sensors; Inductive proximity sensors
PSD (position-sensitive detector), 119
PTP (point-to-point) controls, 453, 465
Pulse phase-locked loop, 313
Pulse Width Modulation (PWM), 22
Pulse-width distortion, 159–160
PUMA (Programmable Universal Machine for Assembly), 420
Pumping the laser, 116
Punched tape, 461, 466
Punch-through, 112
Pushbroom scan, 315
PWM (Pulse Width Modulation), 22
Pyrometers, 334–336

Q

Quality
  corporate manufacturing strategy, 46
  testing, 241
  workpiece, 199
Quantization noise, 503
Queuing (waiting line) analysis, 554–555

R

Radiance (L), 491, 492
Radio-frequency (RF) devices, 277
Radio-frequency interference (RFI), 74–75, 79
Radiometers, 491, 509
Radiometric accuracy
  calibration drift, 517–519
  calibration source errors, 515–517
  channel spectrum error, 520–521
  detector nonlinearity errors, 520
  intrinsic linearity, 519
  spectral aliasing error, 522

Radiometric calibration
linear multiple point, 512–513
multiple point, 512
nonlinear multiple point, 513–515
overview, 509
theory, 510–512
Radiometric errors, 515
Radiometric gain, 511
Radiometric offset, 511
Random access memory (RAM) array, 390
Range sensing, 72–73, 98, 106–109, 412
Reach-through structure, APDs, 361, 367
Read-only memory (ROM), 439
Real-world phenomena, 257
Receivers, 157–158
Red, green, and blue (RGB), 10–11, 14, 22–23, 283
Reference values, 50, 462, 485
Reference-level sensing, 293
Reflected energy, 100
Reflected light, 357
Reflection center, 223
Reflection-type sensor, 328
Reflective mode bifurcate fiber optics, 167
Reflective strip imaging camera sensors, 315–317
Reflex detection methods, 58–59
Reflex photoelectric controls, 137–138
Reflex photoelectric sensors, 52
Refractive index, 353
Relative algorithm, 288
Relative radiometric errors, 517–519
Relative sensor gain, 346
Relays, 81–82
Release, plant, 239
Release control, engineering, 237
Reliability, sensor, 256
Remote sensing, fiber pairs, 168–169
Repeatability, 422
Reporting. *See* Communication, workplace
Research
computer-integrated manufacturing, 231
engineering management, 237–238
engineering release control, 237
facilities engineering, 237
overview, 235–236
process development, 237
product development, 236
Research-Acquire routine, 493–494
Residual current, 80–81
Resistance temperature detectors (RTDs), 257
Resonance and vibration analysis, 197
Resources, 239
Response time, inductive and capacitive sensor
high-speed operation, 94
maximum target length, 94–95
OFF delay logic, 96–97
ON delay logic, 96
one-shot delay, 97
ON/OFF delay logic, 97
overview, 93
release time, 93–94
target duty cycle, 95–96
timing functions, 96
Responsibility centers, 538
Responsivity, APD, 364–365
Restriction faults, software, 202
Retroreflective sensing mode, 144, 152–153
Return on assets (ROA), 527
Return on investment (ROI), 527
Reverse-polarity protection, 88
RF (radio-frequency) devices, 277
RFI (radio-frequency interference), 74–75, 79
RGB (red, green, and blue), 10–11, 14, 22–23, 283
RIA (Robotics Industry Association), 432
Riboswitches, 9
Ring coupler, 163–164
RNA protein mapping, 8–9
ROA (return on assets), 527
Robotic grippers, 310–312
Robotics Industry Association (RIA), 432
Robots
control through sensors, 411–412
density measurements, 434–436
diagnostic systems, 198–199
displacement sensors for, 348–350
history of, 417–419
interfaces, 423–424
invention of, 419–426
material handling, 250
multisensor-controlled assembly, 412–417
programming, 423–424, 426–428
projections for, 436–437
vision system
control through, 383–388
guidance with, 389–391
locating position, 388–389
world market for, 428–434
Rockwell International Corporation, 446
ROI (return on investment), 527
ROM (read-only memory), 439
Roomba domestic vacuum cleaner, 430
RS-232 interface, 213–215, 262
RS-422 interface, 262

RS-485 interface, 262
RTDs (resistance temperature detectors), 257
Runtime phase, 399–400
R.U.R. (Rossum's Universal Robots) play, 417–418

S

SAA (systems application architecture), 269
Sabrie's Index (SI), 491
SADT (structured analysis and design technique), 255–256
Safety monitoring equipment, 103–104
Sales analysis, 540
Sales budget, 524
Sanger, Frederick, 7
Sanger method, 7
SAR (synthetic-aperture radar), 305–307
Satellite Sensor Data Service, 225–226
SAW (surface acoustic waves), 278
Scalable memory technologies, 9–10
Scanning electron microscope (SEM), 5, 110–111
Scanning instability noise, 504
Scheinman, Victor, 420
Schmitt trigger, 293
Selectable functions, 272–273
Self-contained photoelectric controls, 52
Selling and administrative budget, 525–526
SEM (scanning electron microscope), 5, 110–111
Semiconductor absorption sensors, 330–331
Semiconductor device fabrication, 334
Semiconductor displacement laser sensors, 119
Semiconductor temperature detector, 331–334
Sensing principles, 282–283
Sensitivity, SpectRx system
  dark noise, 503–504
  detector current, 498
  efficiency, 504–505
  jitter noise, 504
  Johnson noise, 499–503
  noise equivalent spectral radiance, 505–506
  object view spectral radiation, 495–497
  overview, 495
  photon noise, 498
  power at detector, 497–498
  quantization noise, 503
  scanning instability noise, 504
Sensitivity, SpectRx system (*Cont.*):
  signal-to-noise ratio, 506
  sources of noise, 498
  spectroradiometer transmission, 497
  throughput, 497
Sensitivity adjustment potentiometers, 78
Sensor probes, fiber-optic. *See* Bundles, fiber-optic
Sensors
  overview, 1–4
  types of, 43–48
Sensor-to-target distance, 71
Sensory MEMS, 355–356
Sequencing genome, 7–8
Serial data transmission, 480–482
Series-connection logic, 91–93
Service robots, 435–437. *See also* Robots
Servo systems, 49
Servomotors, 303–305
Session layer (SL), 208–210
SHAPES (spatial high-accuracy position-encoding sensor), 302
Shared memory interfaces (SMIs), 224
Shielded inductive proximity sensors, 73, 149
Shielded sensing coil, 67
Short wave (germicidal) light, 23
Short-circuit protection, 87–88
Shot noise, 362
Showering arc, 75, 79
SI (Sabrie's Index), 491
Side by (lateral) approach, 71–72
Signal conditioning, 259
Signal-to-noise ratio (SNR), 363, 495, 506
Silicon APDs, 364, 367
Silicon nano-crystals, 9–10
Simplex fiber-optic cables, 185
Simulation, 555–556
Single output, 86
Single pole, double throw (SPDT) relay, 52
Single reflex photoelectric sensor, 53
Single-board computers
  overview, 439–440
  personal computers as, 445–450
Single-molecule detection, 321–323
Single-pole switch, 80
Singularities, 425–426
Sinking (NPN) parallel sensor arrangement, 90–91
Site services, plant, 242
SL (session layer), 208–210
SLiK (Super-Low k) APDs, 367
SMA style connector, 171–172
Smart equipment, 556
Smart highways, 1
SMD (surface mount device), 23

SMIs (shared memory interfaces), 224
SMT (surface mount technology), 23
SNA (systems network architecture), 264
SNR (signal-to-noise ratio), 363, 495, 506
Snubber circuit, 83–84
Social interests. *See* Economic interests
Software
  computer-integrated manufacturing, 262
  multisensor-controlled robot assembly, 416–417
  problems, in machinery fault sensor networks, 201–204
  robot, 423
Solid laser, 117
Solid-state arrays, 382
Solid-state lighting (SSL), 23
Solutions, quantifying acidity of with optical sensors, 317–318
Sound-vision recognition end effector sensors
  large surface measurements, 402–404
  overview, 400–402
  positioning, 405–407
  sensitivity of measurements, 404
  small surfaces, 404–405
  standoff, 402
Sourcing (PNP) parallel sensor arrangement, 90
Space industry crack detection sensors, 367–368
Spatial high-accuracy position-encoding sensor (SHAPES), 302
SPDT (single pole, double throw) relay, 52
Special orders analysis, 540
Special processing units (SPUs), 415
Spectral aliasing error, 522
Spectral intensity, 491
Spectral irradiance, 491
Spectral locus, 15
Spectral noise, 362
Spectral radiance, 491
Spectral resolution, 507–508
Spectralon® material, 39–40
Spectroradiometers, 491, 497
Spectroscopy, 356–359
SpectRx NIR technology
  abbreviations, 487–490
  advantages of, 492–494
  apodization, 508–509
  instrument line shape, 506–509
  mathematical relations, 491–492
  overview, 487
SpectRx NIR technology (*Cont.*):
  radiometric accuracy
    calibration drift, 517–519
    calibration source errors, 515–517
    channel spectrum error, 520–521
    detector nonlinearity errors, 520
    intrinsic linearity, 519
    spectral aliasing error, 522
  radiometric calibration, 509–515
  sensitivity
   dark noise, 503–504
    detector current, 498
    efficiency, 504–505
    jitter noise, 504
    Johnson noise, 499–503
    noise equivalent spectral radiance, 505–506
    object view spectral radiation, 495–497
    overview, 495
    photon noise, 498
    power at detector, 497–498
    quantization noise, 503
    scanning instability noise, 504
    signal-to-noise ratio, 506
    sources of noise, 498
    spectroradiometer transmission, 497
    throughput, 497
  spectral resolution, 507–508
  terminology, 490–491
Speed, 422
Speed sensors, 412
Spherical reflecting surface, decentration of, 317
Splice technology, 158
SPUs (special processing units), 415
SQL (Structured Query Language), 266
SRGB color space, 11, 13
SSL (solid-state lighting), 23
Standard platforms, 269–270
Standard range core coil, 67, 68
Standard range field plot, 68
Standard white light, 114–115
Standard-cost system, 529
Star network, 375
Steel industry, 125–127
Stefan-Boltzmann law, 334–335
Stepping motors, 49
Stereoscopic cameras, 309
Straight-line control, 454
Strain gauges, 338
Strategic level, CIM, 247
Strips, LED, 20
Structure vibration measurement sensors, 300–302
Structured analysis and design technique (SADT), 255–256
Structured decisions, 253
Structured Query Language (SQL), 266

Subjective probability, 543–544
Super-Low k (SLiK) APDs, 367
Surface, sensor, 343
Surface acoustic waves (SAW), 278
Surface mount device (SMD), 23
Surface mount technology (SMT), 23
Switching logic
  overview, 88–93
  parallel-connection logic, 89–91
  series-connection logic, 91–93
Synchronous data transmission, 481
Synchronous indexing, 477–478
Synthetic-aperture radar (SAR), 305–307
System performance, improving, 359–360
System response time, 94
Systems application architecture (SAA), 269
Systems network architecture (SNA), 264

T

Tactical level, CIM, 247
Tactile sensors, 381–382, 412
Talk sets, 176
Talkers, 480
Talkers/listeners, 480
Tanaka, Hisashige, 418
Tank level sensor, 109
Target background, 79
Target duty cycles, 95
Task lighting-lamps, 23–25
TCP/IP (Transmission Control Protocol/Internet Protocol), 216
Teach pendants, 423–424
Telescope, 492
Teletypewriters (TTY), 482
Televox, 418
Temperature, 198, 496
Temperature sensors in process control
  noncontact, 334–336
  overview, 329–330
  point-contact, in process manufacturing plants, 334
  semiconductor absorption, 330–331
  semiconductor temperature detector, 331–334
Terminations, fiber-optic, 170–172
Tesla, Nikola, 418
Test sets, dual laser, 175–179
Testing
  fiber optics, 172
  light sources, 172–174
TGF (transforming growth factor), 355
Thermally agitated electrons, 499
Three-channel optical link, 160
Three-wire series-connected circuits, 92–93
Through-beam detection methods, 58
Through-beam sensors, 52–53, 55, 136–137
Throughput
  defined, 491
  FFT spectrometry, 493
  SpectRx NIR technology, 497
Time delay before availability, 93
Timeliness, 536–537
Timers, 261, 449
TL (transport layer), 210
Token systems, 220–221
Tool failure, detecting, 204–207
Torque sensors, 412
Total internal reflection, 141–142
Transceivers, 99
Transducers, 411
Transfer lines, 250
Transforming growth factor (TGF), 355
Transistor DC switches, 84–85
Translation, mechanism for, 255
Transmission Control Protocol/Internet Protocol (TCP/IP), 216
Transmission-type sensors, 329
Transmitters, 157
Transponders, 277
Transport layer (TL), 210
Triac devices, 82–84
Triangular ranging, 147
Trinitron phosphors, 15
Trio-stimulus LEDs, 285
Triple-roll wrist, 426
True color measuring instruments, 286–287
TTY (teletypewriters), 482
TV camera applications, 278
21XTL fiber-optic test set/talk set, 176
Two active star couplers, 182
Two-axis machine tool, 450
Two-dimensional tasks, 396
Two-point calibration, 511
Two-step dual-pole limit switches, 150
Two-wire switches, 80
2260XF switchable dual laser test set, 175
2260XFA laser test set, 175
2285XQ fiber-optic power meter, 173–174
Tyndahl, John, 141

U

UL (Underwriters Laboratories), 151
Ultrasensitive sensors, for single-molecule detection, 321–323
Ultrasonic dynamic vector stress sensor (UDVSS), 312
Ultrasonic sensors
  end effector, 400
  measuring dynamic changes in materials, 312–314
  nondestructive evaluation, 370–371
  overview, 101

UMN (universal memory network), 223–225
Uncertainty, 542
Uncontrollable costs, 538–540
Undamped condition, 66
Underwriters Laboratories (UL), 151
Unimation, 419–420
Universal memory network (UMN), 223–225
Universal product code (UPC), 275
Unstructured decisions, 253
User friendliness, sensor, 256
UV-A, 23
UV-B, 23
UV-C, 23

V

V-24 plug, 482
Variable costs, 528
Variable transport system (VTS), 413
Velocity sensing, 98
Verification, inductive proximity sensors, 63
Vertical synchronous rotary indexing system, 477–478
Very small aperture terminals (VSATs), 1
Vibration analysis, 197
Vibration measurement of structure, 300–302
Video links, 161–162
View angle degree, 24
Viewing angles, 28
Visibility, LED, 28–29
Visible light, 21, 295
Vision systems
  acoustooptical SAR sensors with, 305–307
  allowing robots to recognize objects, 386–388
  future of, 557
  for hand-eye coordination of microrobotic motion, 309–310
  multisensor-controlled robot assembly, 416–417
  overview, 278, 381, 412
  for robot control, 383–388
  robot guidance with, 389–391
  sound recognition and, 400–407
Vitruvian Man, 418
Voice recognition systems, 278
Voltage, task lighting-lamps, 24
Voltage clamp, 85
Voltage regulator, task lighting-lamps, 24
Voltage states, 500
Volts, 24
Vortex-shedding flow sensor, 350–351
VSATs (very small aperture terminals), 1
VTS (variable transport system), 413

W

Waiting line (queuing) analysis, 554–555
Walkie-talkies, 74
Walter, William Grey, 419
Warm white, 25
Waterproof task lighting-lamps, 24
Watts
  versus lumen, 37
  per LED, 25
  task lighting-lamps, 24
Wave number, 491
Wave theory, 113
Wavelength
  accuracy, FFT spectrometry, 493
  laser, 116
  light, 13
Wavelength-division multiplexed (WDM) fiber-optic links, 175
Weatherproof task lighting-lamps, 25
Web loop control, 55
Welding robots, 401
White LEDs, 26–27
White task lighting-lamps, 25
White-points, 15
Whole genome sequencing (WGS), 8
Wide-angle non-diffused LEDs, 27
Wide-beam diffuse (proximity) mode, 153
WIP (work-in-process) inventory, 246
Wire bonding sensor application, 111
Wire termination, 88
Word processing, 449
Work cells, 43–44
Work centers, 43–44
Work force reduction, worldwide, 561
Work-in-process (WIP) inventory, 246
Workshop programming, 468
Workstations, 43–44
Wrist flip, 426

Y

Y-guide probe, 338–339, 348–349
Yield, tracking, 194
Yttrium Aluminum Garnet, 27

Z

Zinc selenide (ZnSe), 31, 497
Zipcord fiber-optic cables, 186
Zone plate, pressure sensor, 339–340

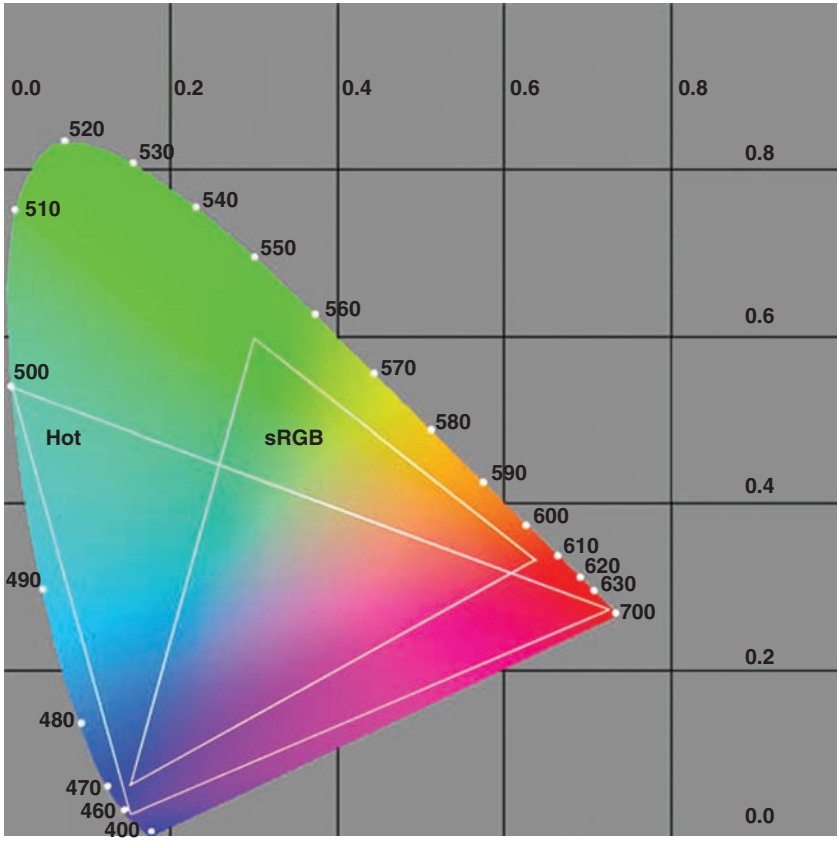


Figure 1.6 Chromaticity—color rendering index (CRI).

	Wavelength (nm)	Color Name	Fwd Voltage (V_f @ 20 ma)	Intensity 5 mm LEDs	Viewing Angle	LED Dye Material
	635	High Eff. Red	2.0	200MCD @20mA	15°	GaAsP/GaP - Gallium Arsenic Phosphide/ Gallium Phosphide
	633	Super Red	2.2	3500MCD @20mA	15°	InGaAIP - Indium Gallium Aluminum Phosphide
	623	Red-Orange	2.2	4500MCD @20mA	15°	InGaAIP - Indium Gallium Aluminum Phosphide
	612	Orange	2.1	160MCD @20mA	15°	GaAsP/GaP - Gallium Arsenic Phosphide/ Gallium Phosphide
	592	Amber Yellow	2.1	7000MCD @20mA	15°	InGaAIP - Indium Gallium Aluminum Phosphide
	585	Yellow	2.1	100MCD @20mA	15°	GaAsP/GaP - Gallium Arsenic Phosphide/ Gallium Phosphide
	3500K	"Incan-descent" White	3.6	2000MCD @20mA	20°	SiC/GaN - Silicon Carbide/Gallium Nitride
	5000K	Pale White	3.6	4000MCD @20mA	20°	SiC/GaN - Silicon Carbide/Gallium Nitride
	6500+K	Cool White	3.6	6000MCD @20mA	20°	SiC/GaN - Silicon Carbide/Gallium Nitride
	574	Super Lime Yellow	2.4	1000MCD @20mA	15°	InGaAIP - Indium Gallium Aluminum Phosphide
	570	Super Lime Green	2.0	1000MCD @20mA	15°	InGaAIP - Indium Gallium Aluminum Phosphide
	565	High Efficiency Green	2.1	200MCD @20mA	15°	GaP/GaP - Gallium Phosphide/Gallium Phosphide
	560	Super Pure Green	2.1	350MCD @20mA	15°	InGaAIP - Indium Gallium Aluminum Phosphide
	555	Pure Green	2.1	80MCD @20mA	15°	GaP/GaP - Gallium Phosphide/Gallium Phosphide
	525	Aqua Green	3.5	10,000MCD @20mA	15°	SiC/GaN - Silicon Carbide/Gallium Nitride
	505	Blue Green	3.5	2000MCD @20mA	45°	SiC/GaN - Silicon Carbide/Gallium Nitride
	470	Super Blue	3.6	3000MCD @20mA	15°	SiC/GaN - Silicon Carbide/Gallium Nitride
	430	Ultra Blue	3.8	100MCD @20mA	15°	SiC/GaN - Silicon Carbide/Gallium Nitride

TABLE 1.2 The LED Color Chart

LED Color	*Standard Brightness*				*High Brightness*			
	Chip Material	Ipk (NM)	Iv (MCD)	Viewing Angle	Chip Material	Ipk (NM)	Iv^3 (MCD)	Viewing Angle
Red	GaAsP/GaP	635	120	35	AS AlInGaP	635	900	30
Orange	GaAsP/Gap	605	90	30	AS AlInGaP	609	1,300	30
Amber	GaAsP/Gap	583	100	35	AS AlInGaP	592	1,300	30
Yellow	Gap	570	160	30	–	–	–	–
Green	Gap	565	140	24	GaN	520	1,200	45
Turquoise	–	–	–	–	GaN	495	2,000	30
Blue	–	–	–	–	GaN	465	325	45

TABLE 1.3 Comparison of Chip Technologies for Wide-Angle Non-Diffused LEDs

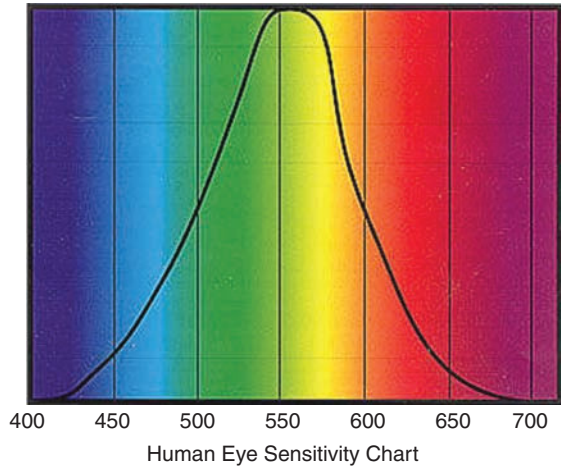


Figure 1.9 Visible light spectra.

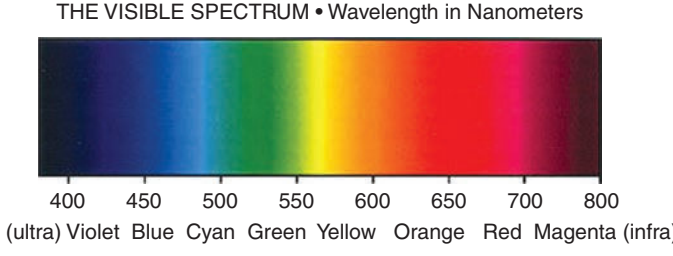


Figure 1.10 Visible ultraviolet light to red light spectra.

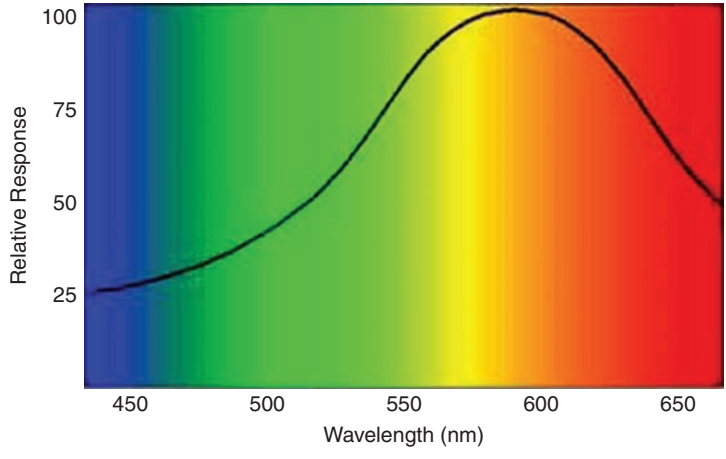


Figure 1.11 An LED peak wavelength (L_{pk}).

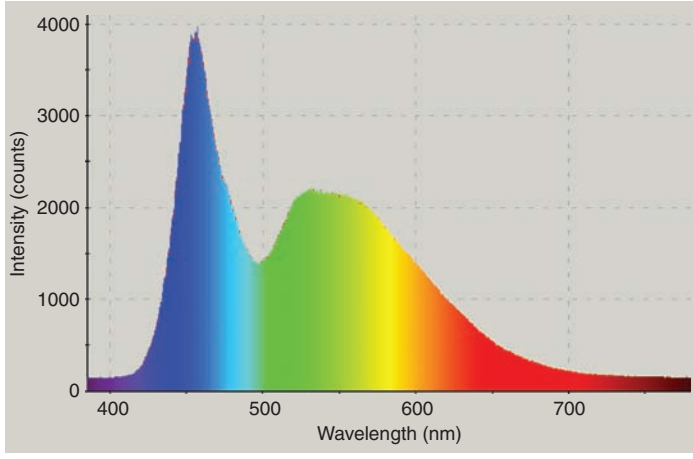


FIGURE 1.17 Spectrographic analysis of a 5-mm non-phosphor white LED.

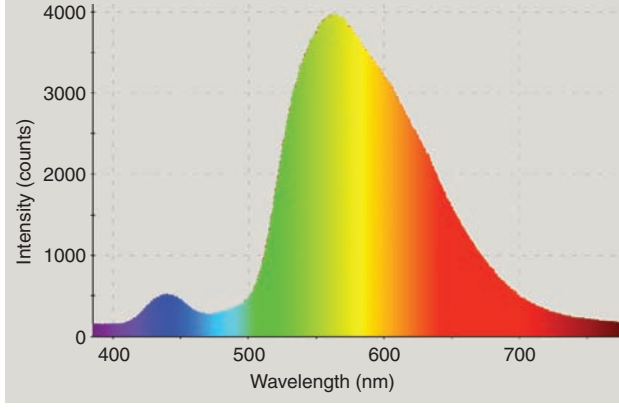


FIGURE 1.18 Spectrographic analysis of a 10-mm non-phosphor warm white LED.

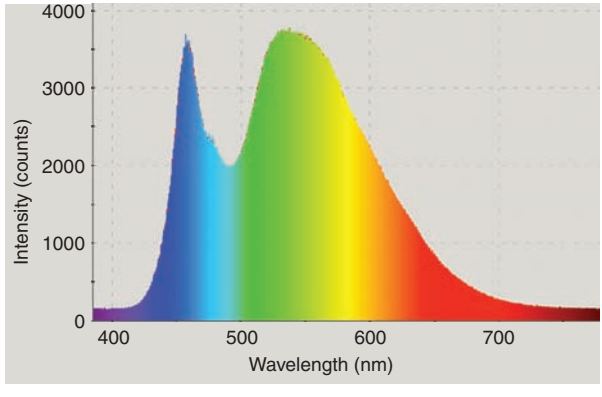


FIGURE 1.19 Spectrographic analysis of a 10-mm non-phosphor LED, when driven at 350.7 mA.