ELEMENTS OF PHOTONICS

Volume II
For Fiber and Integrated Optics

Keigo Iizuka

University of Toronto

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

Designations used by companies to distinguish their products are often claimed as trademarks. In all instances where John Wiley & Sons, Inc., is aware of a claim, the product names appear in initial capital or ALL CAPITAL LETTERS. Readers, however, should contact the appropriate companies for more complete information regarding trademarks and registration.

Copyright © 2002 by John Wiley & Sons, Inc., New York. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic or mechanical, including uploading, downloading, printing, decompiling, recording or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ@WILEY.COM.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional person should be sought.

ISBN 0-471-22137-6

This title is also available in print as ISBN 0-471-40815-8.

For more information about Wiley products, visit our web site at www.Wiley.com.

Kuro, starling dear,
nature's gentle companion
from start to finish

CONTENTS

Volume I

Preface xxv

1 Fourier Optics: Concepts and Applications 1

1.1 Plane Waves and Spatial Frequency / 1
1.1.1 Plane Waves / 1
1.1.2 Spatial Frequency / 4
1.2 Fourier Transform and Diffraction Patterns in Rectangular Coordinates / 9
1.3 Fourier Transform in Cylindrical Coordinates / 16
1.4 Special Functions in Photonics and Their Fourier Transforms / 20
1.4.1 Rectangle Function / 20
1.4.2 Triangle Function / 21
1.4.3 Sign and Step Functions / 25
1.4.4 Circle Function / 25
1.4.5 Delta Function / 28
1.4.6 Shah Function (Impulse Train Function) / 30
1.4.7 Diffraction from an Infinite Array of Similar Apertures with Regular Spacing / 32
1.4.8 Diffraction from an Infinite Array of Similar Apertures with Irregular Spacing / 36
1.4.9 Diffraction from a Finite Array / 37
1.5 The Convex Lens and Its Functions / 40
1.5.1 Phase Distribution After a Plano-Convex Lens / 41
1.5.2 Collimating Property of a Convex Lens / 42
1.5.3 Imaging Property of a Convex Lens / 43
1.5.4 Fourier Transformable Property of a Convex Lens / 46
1.5.5 How Can a Convex Lens Perform the Fourier Transform? / 50
1.5.6 Invariance of the Location of the Input Pattern to the Fourier Transform / 50
1.6 Spatial Frequency Approaches in Fourier Optics / 52
1.6.1 Solution of the Wave Equation by Means of the Fourier Transform / 52
1.6.2 Rayleigh–Sommerfeld Integral / 58
1.6.3 Identifying the Spatial Frequency Components / 60

1.7 Spatial Filters / 61
1.7.1 Image Processing Filters / 61
1.7.2 Optical Correlators / 64
1.7.2.1 Vander Lugt Correlator / 64
1.7.2.2 Detailed Analysis of the Vander Lugt Correlator / 66
1.7.2.3 Joint Transform Correlator / 70
1.7.2.4 Comparison Between VLC and JTC / 72
1.7.3 Rotation and Scaling / 73
1.7.4 Real-Time Correlation / 77
1.7.5 Cryptograph / 78
1.8 Holography / 81
1.8.1 Gabor-Type Hologram / 82
1.8.2 Off-Axis Hologram / 85
1.8.3 Pseudoscopic Image / 87
1.8.4 Volume Hologram / 87
1.8.5 Applications of Holography / 92
1.8.5.1 Three-Dimensional Displays / 92
1.8.5.2 Microfiche Recording / 93
1.8.5.3 Measurement of Displacement / 93
1.8.5.4 Measurement of Vibration / 95
1.8.5.5 Nonoptical Holographies / 95
1.8.5.6 Computer-Generated Holograms / 97
1.8.5.7 Holographic Video Display / 99
Problems / 101
References / 108

2 Boundaries, Near-Field Optics, and Near-Field Imaging 110

2.1 Boundary Conditions / 110
2.2 Snell's Law / 112
2.3 Transmission and Reflection Coefficients / 113
2.3.1 Transmission and Reflection Coefficients (at Normal Incidence) / 114
2.3.2 Transmission and Reflection Coefficients (at an Arbitrary Incident Angle) / 118
2.3.3 Impedance Approach to Calculating Transmission and Reflection Coefficients / 124
2.4 Transmittance and Reflectance (at an Arbitrary Incident Angle) / 124
2.5 Brewster's Angle / 127
2.6 Total Internal Reflection / 130
2.6.1 Bends in a Guide / 131
2.7 Wave Expressions of Light / 132
2.7.1 Fields Near the Boundary / 133
2.8 The Evanescent Wave / 134
2.8.1 Transmission and Reflection Coefficients for Total Internal Reflection / 135

2.8.2 Goos-Hänchen Shift / 141
2.8.3 Evanescent Field and Its Adjacent Fields / 142
2.8.4 k Diagrams for the Graphical Solution of the Evanescent Wave / 145
2.9 What Generates the Evanescent Waves? / 147
2.9.1 Structures for Generating Evanescent Waves / 147
2.10 Diffraction-Unlimited Images out of the Evanescent Wave / 150
2.10.1 Resolution of a Lens-Type Microscope / 150
2.10.2 Near-Field Optical Microscopes / 152
2.10.2.1 Photon Tunneling Microscope / 152
2.10.2.2 Scanning Near-Field Optical Microscope (SNOM) / 154
2.10.3 Probes to Detect the Evanescent Field / 154
2.10.4 Apertures of the SNOM Probes / 158
2.10.5 Modes of Excitation of the SNOM Probes / 158
2.10.6 SNOM Combined with AFM / 160
2.10.7 Concluding Remarks / 161
Problems / 163
References / 164

3 Fabry–Pérot Resonators, Beams, and Radiation Pressure 166

3.1 Fabry–Pérot Resonators / 166
3.1.1 Operating Principle of the Fabry–Pérot Resonator / 167
3.1.2 Transmittance and Reflectance of the Fabry–Pérot Resonator with an Arbitrary Angle of Incidence / 170
3.2 The Scanning Fabry–Pérot Spectrometer / 176
3.2.1 Scanning by the Reflector Spacing / 177
3.2.1.1 Fabry–Pérot Resonator with a Fixed Resonator Spacing (Etalon) / 179
3.2.1.2 Monochromatic Incident Light with Scanned Reflector Spacing / 179
3.2.1.3 Free Spectral Range (FSR) / 179
3.2.2 Scanning by the Angle of Incidence / 184
3.2.3 Scanning by the Index of Refraction / 187
3.2.4 Scanning by the Frequency of Incident Light / 190
3.3 Resolving Power of the Fabry–Pérot Resonator / 192
3.4 Practical Aspects of Operating the Fabry–Pérot Interferometer / 199
3.4.1 Methods for Parallel Alignment of the Reflectors / 199
3.4.2 Method for Determining the Spacing Between the Reflectors / 202
3.4.3 Spectral Measurements Without Absolute Measurement of d / 203
3.5 The Gaussian Beam as a Solution of the Wave Equation / 205
3.5.1 Fundamental Mode / 206

3.5.2 Properties of the q Parameter / 208
3.5.2.1 Beam Waist / 208
3.5.2.2 Location of the Waist / 208
3.5.2.3 Radius of Curvature of the Wavefront / 208
3.5.3 With the Origin at the Waist / 209
3.5.3.1 Focal Parameters / 209
3.5.3.2 Correction Factor / 210
3.5.4 Gaussian Beam Expressions / 211
3.5.4.1 Amplitude Distribution / 211
3.5.4.2 Intensity Distribution / 211
3.5.4.3 Angle of the Far-Field Divergence / 212
3.5.4.4 Depth of Focus / 213
3.6 Transformation of a Gaussian Beam by a Lens / 214
3.6.1 Transformation of the q Parameter by a Lens / 215
3.6.2 Size of the Waist of the Emergent Beam / 216
3.6.3 Location of the Waist of the Emergent Beam / 217
3.6.4 Rayleigh Range of the Emergent Beam / 218
3.6.5 Angle of the Far-Field Divergence of the Emergent Beam / 218
3.6.6 Comparison with Ray Optics / 218
3.6.7 Summary of the Equations of the Transformation by a Lens / 219
3.6.8 Beam Propagation Factor m^2 / 220
3.7 Hermite Gaussian Beam (Higher Order Modes) / 223
3.8 The Gaussian Beam in a Spherical Mirror Cavity / 227
3.9 Resonance Frequencies of the Cavity / 232
3.10 Practical Aspects of the Fabry–Pérot Interferometer / 234
3.10.1 Plane Mirror Cavity / 234
3.10.2 General Spherical Mirror Cavity / 235
3.10.3 Focal Cavity / 235
3.10.4 Confocal Cavity / 236
3.11 Bessel Beams / 237
3.11.1 Features of the Bessel Beam / 237
3.11.2 Practical Applications of the Bessel Beam / 239
3.11.2.1 Precision Optical Measurement / 239
3.11.2.2 Power Transport / 239
3.11.2.3 Nonlinear Optics / 239
3.11.3 One-Dimensional Model / 239
3.11.4 Mathematical Expressions for the Bessel Beam / 242
3.11.5 Methods of Generating Bessel Beams / 245
3.12 Manipulation with Light Beams / 249
3.12.1 Radiation Pressure of Laser Light / 249
3.12.2 Optical Tweezers / 251
3.13 Laser Cooling of Atoms / 254
Problems / 255
References / 260

4 Propagation of Light in Anisotropic Crystals 263

4.1 Polarization in Crystals / 264
4.2 Susceptibility of an Anisotropic Crystal / 266
4.3 The Wave Equation in an Anisotropic Medium / 268
4.4 Solving the Generalized Wave Equation in Uniaxial Crystals / 269
4.4.1 Graphical Derivation of the Condition of Propagation in a Uniaxial Crystal / 270
4.4.2 Analytical Representation of the Conditions of Propagation in a Uniaxial Crystal / 273
4.4.3 Wavenormal and Ray Direction / 275
4.4.4 Derivation of the Effective Index of Refraction / 280
4.5 Graphical Methods / 282
4.5.1 Wavevector Method / 282
4.5.2 Indicatrix Method / 285
4.6 Treatment of Boundary Problems Between Anisotropic Media by the Indicatrix Method / 292
4.6.1 Refraction of the e-Wave at the Boundary of Anisotropic Media / 292
4.6.2 Reflection of the e-Wave at the Boundary of Anisotropic Media / 294
4.6.3 Total Internal Reflection of the e-Wave at the Boundary of Anisotropic Media / 296
Problems / 298
References / 301

5 Optical Properties of Crystals Under Various External Fields 302

5.1 Expressing the Distortion of the Indicatrix / 302
5.2 Electrooptic Effects / 304
5.2.1 Pockels Electrooptic Effect / 304
5.2.2 Kerr Electrooptic Effect / 316
5.3 Elastooptic Effect / 317
5.4 Magnetooptic Effect / 326
5.4.1 Faraday Effect / 326
5.4.2 Cotton-Mouton Effect / 327
5.5 Optical Isolator / 327
5.5.1 Polarization-Dependent Optical Isolator / 328
5.5.2 Polarization-Independent Optical Isolator / 330
5.6 Photorefractive Effect / 331
5.7 Optical Amplifier Based on the Photorefractive Effect / 334
5.7.1 Enhanced Photorefractive Effect by an External Electric Field / 334
5.7.2 Energy Transfer in the Crystal / 335
5.7.3 Optical Amplifier Structure / 338
5.8 Photorefractive Beam Combiner for Coherent Homodyne Detection / 339
5.9 Optically Tunable Optical Filter / 341

5.10 Liquid Crystals / 341
5.10.1 Types of Liquid Crystals / 341
5.10.1.1 Cholesteric / 342
5.10.1.2 Smectic / 343
5.10.1.3 Nematic / 343
5.10.1.4 Discotic / 344
5.10.2 Molecular Orientations of the Nematic Liquid Crystal Without an External Field / 344
5.10.3 Molecular Reorientation of the Nematic Liquid Crystal with an External Electric Field / 345
5.10.4 Liquid Crystal Devices / 346
5.10.4.1 Liquid Crystal Fabry–Pérot Resonator / 346
5.10.4.2 Liquid Crystal Rotatable Waveplate / 346
5.10.4.3 Liquid Crystal Microlens / 347
5.10.4.4 Twisted Nematic (TN) Liquid Crystal Spatial Light Modulator (TNSLM) / 349
5.10.4.5 Electrically Addressed Spatial Light Modulator (EASLM) / 350
5.10.4.6 Optically Addressed Spatial Light Modulator (OASLM) / 351
5.10.4.7 Polymer-Dispersed Liquid Crystal (PDLC)-Type Spatial Light Modulator (SLM) / 352
5.10.5 Guest–Host Liquid Crystal Cell / 353
5.10.6 Ferroelectric Liquid Crystal / 354
5.11 Dye-Doped Liquid Crystal / 357
Problems / 358
References / 359

6 Polarization of Light 362

6.1 Introduction / 363
6.2 Circle Diagrams for Graphical Solutions / 365
6.2.1 Linearly Polarized Light Through a Retarder / 365
6.2.2 Sign Conventions / 368
6.2.3 Handedness / 374
6.2.4 Decomposition of Elliptically Polarized Light / 375
6.2.5 Transmission of an Elliptically Polarized Wave Through a $\lambda/4$ Plate / 377
6.3 Various Types of Retarders / 378
6.3.1 Waveplates / 379
6.3.2 Compensator / 380
6.3.3 Fiber-Loop Retarder / 382
6.4 How to Use Waveplates / 385
6.4.1 How to Use a Full-Waveplate / 385
6.4.2 How to Use a Half-Waveplate / 385
6.4.3 How to Use a Quarter-Waveplate / 386
6.4.3.1 Conversion from Linear to Circular Polarization by Means of a $\lambda/4$ Plate / 386

6.4.3.2 Converting Light with an Unknown State of Polarization into Linearly Polarized Light by Means of a $\lambda/4$ Plate / 389
6.4.3.3 Measuring the Retardance of a Sample / 392
6.4.3.4 Measurement of Retardance of an Incident Field / 393
6.5 Linear Polarizers / 394
6.5.1 Dichroic Polarizer / 394
6.5.2 Birefringence Polarizer or Polarizing Prism / 402
6.5.3 Birefringence Fiber Polarizer / 404
6.5.4 Polarizers Based on Brewster's Angle and Scattering / 407
6.5.5 Polarization Based on Scattering / 408
6.6 Circularly Polarizing Sheets / 409
6.6.1 Antiglare Sheet / 409
6.6.2 Monitoring the Reflected Light with Minimum Loss / 411
6.7 Rotators / 412
6.7.1 Saccharimeter / 417
6.7.2 Antiglare TV Camera / 419
6.8 The Jones Vector and the Jones Matrix / 421
6.8.1 The Jones Matrix of a Polarizer / 422
6.8.2 The Jones Matrix of a Retarder / 424
6.8.3 The Jones Matrix of a Rotator / 425
6.8.4 Eigenvectors of an Optical System / 428
6.9 States of Polarization and Their Component Waves / 431
6.9.1 Major and Minor Axes of an Elliptically Polarized Wave / 431
6.9.2 Azimuth of the Principal Axes of an Elliptically Polarized Wave / 434
6.9.3 Ellipticity of an Elliptically Polarized Wave / 438
6.9.4 Conservation of Energy / 437
6.9.5 Relating the Parameters of an Elliptically Polarized Wave to Those of Component Waves / 439
6.9.6 Summary of Essential Formulas / 439
Problems / 446
References / 449

7 How to Construct and Use the Poincaré Sphere 451

7.1 Component Field Ratio in the Complex Plane / 452
7.2 Constant Azimuth θ and Ellipticity ϵ Lines in the Component Field Ratio Complex Plane / 455
7.2.1 Lines of Constant Azimuth θ / 455
7.2.2 Lines of Constant Ellipticity ϵ / 458
7.3 Argand Diagram / 459
7.3.1 Solution Using a Ready-Made Argand Diagram / 460
7.3.2 Orthogonality Between Constant θ and ϵ Lines / 465
7.3.3 Solution Using a Custom-Made Argand Diagram / 468

7.4 From Argand Diagram to Poincaré Sphere / 469
7.4.1 Analytic Geometry of Back-Projection / 469
7.4.2 Poincaré Sphere / 474
7.5 Poincaré Sphere Solutions for Retarders / 479
7.6 Poincaré Sphere Solutions for Polarizers / 485
7.7 Poincaré Sphere Traces / 490
7.8 Movement of a Point on the Poincaré Sphere / 494
7.8.1 Movement Along a Line of Constant Longitude (or Constant θ Line) / 494
7.8.2 Movement Along a Line of Constant Latitude (or Constant β Line) / 497
Problems / 501
References / 503

8 Phase Conjugate Optics 504

8.1 The Phase Conjugate Mirror / 504
8.2 Generation of a Phase Conjugate Wave Using a Hologram / 504
8.3 Expressions for Phase Conjugate Waves / 507
8.4 Phase Conjugate Mirror for Recovering Phasefront Distortion / 508
8.5 Phase Conjugation in Real Time / 511
8.6 Picture Processing by Means of a Phase Conjugate Mirror / 512
8.7 Distortion-Free Amplification of Laser Light by Means of a Phase Conjugate Mirror / 513
8.8 Self-Tracking of a Laser Beam / 514
8.9 Picture Processing / 519
8.10 Theory of Phase Conjugate Optics / 521
8.10.1 Maxwell's Equations in a Nonlinear Medium / 521
8.10.2 Nonlinear Optical Susceptibilities $\chi^{(2)}$ and $\chi^{(3)}$ / 523
8.10.3 Coupled Wave Equations / 526
8.10.4 Solutions with Bohr's Approximation / 529
8.11 The Gain of Forward Four-Wave Mixing / 533
8.12 Pulse Broadening Compensation by Forward Four-Wave Mixing / 537
Problems / 541
References / 543

Appendix A Derivation of the Fresnel–Kirchhoff Diffraction Formula from the Rayleigh–Sommerfeld Diffraction Formula 545

Appendix B Why the Analytic Signal Method is Not Applicable to the Nonlinear System 547

Appendix C Derivation of P_{NL} 551

Answers to Problems 554

Index I-1

CONTENTS

Volume II

Preface xxv

9 Planar Optical Guides for Integrated Optics 605

9.1 Classification of the Mathematical Approaches to the Slab Optical Guide / 606
9.2 Wave Optics Approach / 607
9.3 Characteristic Equations of the TM Modes / 610
9.3.1 Solutions for K and γ / 610
9.3.2 Examples Involving TM Modes / 612
9.4 Cross-Sectional Distribution of Light and its Decomposition into Component Plane Waves / 615
9.5 Effective Index of Refraction / 619
9.6 TE Modes / 620
9.7 Other Methods for Obtaining the Characteristic Equations / 622
9.7.1 Coefficient Matrix Method / 623
9.7.2 Transmission Matrix Method (General Guides) / 625
9.7.3 Transmission Matrix Method (Symmetric Guide) / 630
9.7.4 Modified Ray Model Method / 636
9.8 Asymmetric Optical Guide / 638
9.9 Coupled Guides / 643
9.9.1 Characteristic Equations of the Coupled Slab Guide / 643
9.9.2 Amplitude Distribution in the Coupled Slab Guide / 646
9.9.3 Coupling Mechanism of the Slab Guide Coupler / 651
Problems / 652
References / 654

10 Optical Waveguides and Devices for Integrated Optics 655

10.1 Rectangular Optical Waveguide / 655
10.1.1 Assumptions / 655

10.1.2 Characteristic Equation for the Rectangular Guide / 657
10.1.3 A Practical Example / 659
10.2 Effective Index Method for Rectangular Optical Guides / 661
10.3 Coupling Between Rectangular Guides / 664
10.4 Conflection / 666
10.4.1 Conflection Lens / 667
10.5 Various Kinds of Rectangular Optical Waveguides for Integrated Optics / 670
10.5.1 Ridge Guide / 670
10.5.2 Rib Guide / 670
10.5.3 Strip-Loaded Guide / 671
10.5.4 Embedded Guide / 671
10.5.5 Immersed Guide / 672
10.5.6 Bulge Guide / 672
10.5.7 Metal Guide / 672
10.5.8 Buffered Metal Guide / 672
10.5.9 Photochromic Flexible Guide / 672
10.6 Power Dividers / 673
10.6.1 The Y Junction and Arrayed-Waveguide Grating / 673
10.6.2 Power Scrambler / 677
10.7 Optical Magic T / 678
10.8 Electrode Structures / 680
10.8.1 Laminated Electrodes / 680
10.8.2 Electrode Configurations / 681
10.8.2.1 Applying a Longitudinal Field to Bulk Waves / 681
10.8.2.2 Applying a Transverse Field to Bulk Waves / 681
10.8.2.3 Vertical Field in an Embedded Guide / 683
10.8.2.4 Vertical Field in Adjacent Embedded Guides / 683
10.8.2.5 Velocity Matched Mach–Zehnder Interferometer / 683
10.8.2.6 Horizontal Field in an Embedded Guide / 683
10.8.2.7 Horizontal Field in a Rib Guide / 684
10.8.2.8 Horizontal and Vertical Fields / 684
10.8.2.9 Periodic Vertical Field / 684
10.8.2.10 Periodic Horizontal Field / 684
10.8.2.11 Trimming Electrodes / 684
10.8.2.12 Switching Electrodes / 685
10.9 Mode Converter / 685
Problems / 688
References / 690

11 Modes and Dispersion in Optical Fibers 692

11.1 Practical Aspects of Optical Fibers / 693
11.1.1 Numerical Aperture of a Fiber / 693
11.1.2 Transmission Loss of Fibers / 694
11.1.3 Loss Increase Due to Hydrogen and Gamma-Ray Irradiation / 695
11.1.4 Dispersion / 699
11.1.5 Mode Dispersion / 699
11.1.6 Material and Waveguide Dispersions / 701
11.1.7 Various Kinds of Optical Fibers / 703
11.1.7.1 Multimode Step-Index Fiber / 703
11.1.7.2 Multimode Graded-Index Fiber / 705
11.1.7.3 Single-Mode Fiber / 705
11.1.7.4 Dispersion-Shifted Fiber / 705
11.1.7.5 Silica Core Fluorine-Added Cladding Fiber / 705
11.1.7.6 Plastic Fiber / 706
11.1.7.7 Multi-Ingredient Fiber / 706
11.1.7.8 Holey Optical Fiber (HF) / 706
11.1.7.9 Polarization-Preserving Fiber / 707
11.1.8 Optical Fibers Other Than Silica Based Fibers / 708
11.2 Theory of Step-Index Fibers / 709
11.2.1 Solutions of the Wave Equations in Cylindrical Coordinates / 709
11.2.2 Expressions for the E_z and H_z Components / 711
11.2.2.1 Solutions in the Core Region / 713
11.2.2.2 Solutions in the Cladding Region / 714
11.2.3 Expressions for the E_r, E_ϕ, H_r, and H_ϕ Components / 715
11.2.4 Characteristic Equation of an Optical Fiber / 717
11.2.5 Modes in Optical Fibers / 718
11.2.5.1 Meridional Modes: $v = 0$ / 718
11.2.5.2 Skew Modes: $v \neq 0$ / 721
11.3 Field Distributions Inside Optical Fibers / 730
11.3.1 Sketching Hybrid Mode Patterns / 732
11.3.2 Sketching Linearly Polarized Mode Patterns / 735
11.4 Dual-Mode Fiber / 739
11.5 Photoimprinted Bragg Grating Fiber / 741
11.5.1 Methods of Writing Photoinduced Bragg Gratings in an Optical Fiber / 742
11.5.1.1 Internal Writing / 743
11.5.1.2 Holographic Writing / 743
11.5.1.3 Point-by-Point Writing / 744
11.5.1.4 Writing by a Phase Mask / 744
11.5.2 Applications of the Photoinduced Bragg Gratings in an Optical Fiber / 744

11.6 Definitions Associated with Dispersion / 748
11.6.1 Definitions of Group Velocity and Group Delay / 748
11.6.2 Definition of the Dispersion Parameter / 749
11.7 Dispersion-Shifted Fiber / 749
11.7.1 Group Delay in an Optical Fiber / 749
11.7.2 Dispersion Parameter of an Optical Fiber / 751
11.8 Dispersion Compensator / 755
11.8.1 Phase Conjugation Method / 755
11.8.2 Bragg Grating Method / 755
11.8.3 Dual-Mode Fiber Method / 757
11.9 Ray Theory for Graded-Index Fibers / 759
11.9.1 Eikonal Equation / 759
11.9.2 Path of Light in a Graded-Index Fiber / 762
11.9.3 Quantization of the Propagation Constant in a Graded-Index Fiber / 766
11.9.4 Dispersion of Graded-Index Fibers / 768
11.9.5 Mode Patterns in a Graded-Index Fiber / 770
11.10 Fabrication of Optical Fibers / 775
11.10.1 Fabrication of a Fiber by the One-Stage Process / 775
11.10.2 Fabrication of a Fiber by the Two-Stage Process / 777
11.10.2.1 Fabrication of Preforms / 777
11.10.2.2 Drawing into an Optical Fiber / 782
11.11 Cabling of Optical Fibers / 783
11.12 Joining Fibers / 786
11.12.1 Splicing Fibers / 786
11.12.2 Optical Fiber Connector / 790
Problems / 790
References / 793

12 Detecting Light 796

12.1 Photomultiplier Tube / 796
12.2 Streak Camera / 798
12.3 Miscellaneous Types of Light Detectors / 800
12.4 PIN Photodiode and APD / 801
12.4.1 Physical Structures of PIN and APD Photodetectors / 801
12.4.2 Responsivity of the PIN Photodiode and APD / 803
12.5 Direct Detection Systems / 805
12.6 Coherent Detection Systems / 807
12.6.1 Heterodyne Detection / 807
12.6.2 Homodyne Detection / 809
12.6.3 Intradyne System / 812

12.7 Balanced Mixer / 814
12.8 Detection by Stimulated Effects / 815
12.8.1 Stimulated Effects / 816
12.8.2 Homodyne Detection by Stimulated Brillouin Scattering / 817
12.9 Jitter in Coherent Communication Systems / 819
12.9.1 Polarization Jitter Controls / 819
12.9.1.1 Computer-Controlled Method of Jitter Control / 820
12.9.1.2 Polarization Diversity Method / 822
12.9.2 Phase Jitter / 823
12.10 Coherent Detection Immune to Both Polarization and Phase Jitter / 826
12.11 Concluding Remarks / 830
Problems / 830
References / 831

13 Optical Amplifiers 833

13.1 Introduction / 833
13.2 Basics of Optical Amplifiers / 834
13.3 Types of Optical Amplifiers / 836
13.4 Gain of Optical Fiber Amplifiers / 838
13.4.1 Spectral Lineshape / 839
13.5 Rate Equations for the Three-Level Model Of Er^{3+} / 848
13.5.1 Normalized Steady-State Population Difference / 849
13.5.2 Gain of the Amplifier / 852
13.6 Pros and Cons of 1.48-μm and 0.98-μm Pump Light / 853
13.7 Approximate Solutions of the Time-Dependent Rate Equations / 857
13.8 Pumping Configuration / 864
13.8.1 Forward Pumping Versus Backward Pumping / 864
13.8.2 Double-Clad Fiber Pumping / 866
13.9 Optimum Length of the Fiber / 867
13.10 Electric Noise Power When the EDFA is Used as a Preamplifier / 868
13.11 Noise Figure of the Receiver Using the Optical Amplifier as a Preamplifier / 880
13.12 A Chain of Optical Amplifiers / 882
13.13 Upconversion Fiber Amplifier / 889
Problems / 889
References / 892

14 Transmitters 893

14.1 Types of Lasers / 893
14.1.1 Gas Lasers / 893

14.1.2 Solid-State Lasers / 894
14.1.3 Dye Lasers / 894
14.1.4 Chemical Lasers / 895
14.2 Semiconductor Lasers / 895
14.2.1 Gain of a Semiconductor Laser Amplifier / 895
14.2.2 Laser Cavity / 903
14.2.3 Conditions for Laser Oscillation / 904
14.2.3.1 Amplitude Condition for Laser Oscillation / 905
14.2.3.2 Phase Condition for Laser Oscillation / 907
14.2.4 Qualitative Explanation of Laser Oscillation / 908
14.3 Rate Equations of Semiconductor Lasers / 909
14.3.1 Steady-State Solutions of the Rate Equations / 911
14.3.2 Threshold Electron Density and Current / 911
14.3.3 Output Power from the Laser / 913
14.3.4 Time-Dependent Solutions of the Rate Equations / 914
14.3.4.1 Turn-On Delay / 914
14.3.4.2 Relaxation Oscillation / 916
14.3.5 Small Signal Amplitude Modulation / 916
14.3.5.1 Time Constant of the Relaxation Oscillation / 917
14.3.5.2 Amplitude Modulation Characteristics / 919
14.3.5.3 Comparisons Between Theoretical and Experimental Results / 920
14.4 Confinement / 930
14.4.1 Carrier Confinement / 930
14.4.2 Confinement of the Injection Current / 933
14.4.2.1 Narrow Stripe Electrode / 934
14.4.2.2 Raised Resistivity by Proton Bombardment / 934
14.4.2.3 Barricade by a Back-Biased p-n Junction Layer / 935
14.4.2.4 Dopant-Diffused Channel / 936
14.4.2.5 Modulation of the Layer Thickness / 937
14.4.3 Light Confinement / 937
14.4.3.1 Gain Guiding / 937
14.4.3.2 Plasma-Induced Carrier Effect / 939
14.4.3.3 Kink in the Characteristic Curve / 941
14.4.3.4 Stabilization of the Lateral Modes / 942
14.5 Wavelength Shift of the Radiation / 943

14.5.1 Continuous Wavelength Shift with Respect to Injection Current / 944
14.5.2 Mode Hopping / 945
14.6 Beam Pattern of a Laser / 946
14.7 Temperature Dependence of L–I Curves / 951
14.8 Semiconductor Laser Noise / 952
14.8.1 Noise Due to External Optical Feedback / 953
14.8.2 Noise Associated with Relaxation Oscillation / 955
14.8.3 Noise Due to Mode Hopping / 955
14.8.4 Partition Noise / 955
14.8.5 Noise Due to Spontaneous Emission / 955
14.8.6 Noise Due to Fluctuations in Temperature and Injection Current / 955
14.9 Single-Frequency Lasers / 956
14.9.1 Surface Emitting Laser / 956
14.9.2 Laser Diodes with Bragg Reflectors / 957
14.9.3 $\lambda/4$ Shift DFB Lasers / 961
14.9.4 Diode Pumped Solid-State Laser / 967
14.10 Wavelength Tunable Laser Diode / 970
14.10.1 Principle of Frequency Tuning of the DFB Laser / 970
14.10.1.1 Tuning of Wavelength by the Phase Controller Tuning Current I_p Alone / 973
14.10.1.2 Tuning of Wavelength by the Bragg Reflector Tuning Current I_b Alone / 973
14.10.1.3 Continuous Wavelength Tuning by Combining I_p and I_b / 975
14.10.2 Superstructure Grating Laser Diode (SSG-LD) / 977
14.11 Laser Diode Array / 980
14.12 Multi-Quantum-Well Lasers / 984
14.12.1 Energy States in a Bulk Semiconductor / 985
14.12.2 Energy States in a Quantum Well / 988
14.12.3 Gain Curves of the MQW Laser / 992
14.12.4 Structure and Characteristics of a MQW Laser / 994
14.12.5 Density of States of a Quantum Wire and Quantum Dot / 999
14.13 Erbium-Doped Fiber Laser / 1004
14.14 Light-Emitting Diode (LED) / 1007
14.14.1 LED Characteristics / 1007
14.14.2 LED Structure / 1008
14.15 Fiber Raman Lasers / 1009
14.16 Selection of Light Sources / 1011
Problems / 1013
References / 1014

15 Stationary and Solitary Solutions in a Nonlinear Medium 1017

15.1 Nonlinear (Kerr) Medium / 1017
15.2 Solutions in the Unbounded Kerr Nonlinear Medium / 1021
15.2.1 Method by a Trial Solution / 1024
15.2.2 Method by Integration / 1026
15.2.3 Method Using Jacobi's Elliptic Functions / 1027
15.3 Guided Nonlinear Boundary Wave / 1030
15.4 Linear Core Layer Sandwiched by Nonlinear Cladding Layers / 1037
15.4.1 General Solutions / 1038
15.4.2 Characteristic Equations from the Boundary Conditions / 1039
15.4.3 Normalized Thickness of the Nonlinear Guides / 1041
15.4.4 Fields at the Top and Bottom Boundaries / 1042
15.4.5 Modes Associated with Equal Boundary Field Intensities ($a_0 = a_2$) / 1043
15.4.6 Modes Associated with $a_0 + a_2 = \epsilon_1 - \epsilon_0$ / 1048
15.5 How the Soliton Came About / 1049
15.6 How a Soliton is Generated / 1050
15.7 Self-Phase Modulation (SPM) / 1053
15.8 Group Velocity Dispersion / 1055
15.9 Differential Equation of the Envelope Function of the Solitons in the Optical Fiber / 1059
15.10 Solving the Nonlinear Schrödinger Equation / 1067
15.11 Fundamental Soliton / 1068
15.12 Pulsewidth and Power to Generate a Fundamental Soliton / 1071
15.13 Ever-Expanding Soliton Theories / 1074
Problems / 1077
References / 1079

16 Communicating by Fiber Optics 1081

16.1 Overview of Fiber-Optic Communication Systems / 1082
16.1.1 Transmitters / 1082
16.1.2 Modulation of Light / 1082
16.1.3 Transmission Through the Optical Fiber / 1084
16.1.4 Received Signal / 1085
16.1.5 Multiplexing Hierarchies / 1085
16.2 Modulation / 1085
16.2.1 Amplitude Modulation / 1086
16.2.2 Variations of Amplitude Modulation / 1086
16.2.3 Angle Modulation / 1092

16.2.4 Pulse Modulation / 1094
16.2.5 Pulse Code Modulation / 1094
16.2.6 Binary Modulation (Two-State Modulation) / 1095
16.2.7 Amplitude Shift Keying (ASK) / 1095
16.2.8 Frequency Shift Keying (FSK) and Phase Shift Keying (PSK) / 1095
16.2.9 Representation of Bits / 1096
16.3 Multiplexing / 1097
16.3.1 Wavelength Division Multiplexing (WDM) / 1098
16.3.2 Frequency Division Multiplexing (FDM) / 1099
16.3.3 Time Division Multiplexing (TDM) / 1100
16.4 Light Detection Systems / 1102
16.4.1 Equivalent Circuit of the PIN Photodiode / 1102
16.4.2 Frequency Response of the PIN Diode / 1104
16.4.3 Coupling Circuits to a Preamplifier / 1106
16.4.3.1 Coupling Circuits to a Preamplifier at Subgigahertz / 1106
16.4.3.2 Coupling Circuits to a Preamplifier Above a Gigahertz / 1110
16.5 Noise in the Detector System / 1113
16.5.1 Shot Noise / 1113
16.5.2 Thermal Noise / 1114
16.5.3 Signal to Noise Ratio / 1115
16.5.4 Excess Noise in the APD / 1117
16.5.5 Noise Equivalent Power (NEP) / 1117
16.5.6 Signal to Noise Ratio for ASK Modulation / 1121
16.5.7 Signal to Noise Ratio of Homodyne Detection / 1122
16.5.8 Borderline Between the Quantum-Limited and Thermal-Noise-Limited S/N / 1123
16.5.9 Relationship Between Bit Error Rate (BER) and Signal to Noise Ratio / 1123
16.6 Designing Fiber-Optic Communication Systems / 1129
16.6.1 System Loss / 1130
16.6.2 Power Requirement for Analog Modulation / 1130
16.6.3 Rise-Time Requirement for Analog Modulation / 1132
16.6.4 Example of an Analog System Design / 1134
16.6.5 Required Frequency Bandwidth for Amplifying Digital Signals / 1139
16.6.6 Digital System Design / 1141
16.6.7 Example of Digital System Design / 1144
16.6.7.1 Power Requirement / 1144
16.6.7.2 Rise-Time Requirement / 1145
Problems / 1147
References / 1149

Appendix A PIN Photodiode on an Atomic Scale **1151**

A.1 PIN Photodiode / 1151
A.2 I–V Characteristics / 1156

Appendix B Mode Density **1160**

Appendix C Perturbation Theory **1164**

Answers to Problems **1167**

Index **I-1**

PREFACE

After visiting leading optics laboratories for the purpose of producing the educational video *Fiber Optic Labs from Around the World* for the Institute of Electrical and Electronics Engineers (IEEE), I soon realized there was a short supply of photonics textbooks to accommodate the growing demand for photonics engineers and evolving fiber-optic products. This textbook was written to help fill this need.

From my teaching experiences at Harvard University and the University of Toronto, I learned a great deal about what students want in a textbook. For instance, students hate messy mathematical expressions that hide the physical meaning. They want explanations that start from the very basics, yet maintain simplicity and succinctness. Most students do not have a lot of time to spend reading and looking up references, so they value a well-organized text with everything at their fingertips. Furthermore, a textbook with a generous allotment of numerical examples helps them better understand the material and gives them greater confidence in tackling challenging problem sets. This book was written with the student in mind.

The book amalgamates fundamentals with applications and is appropriate as a text for a fourth year undergraduate course or first year graduate course. Students need not have a previous knowledge of optics, but college physics and mathematics are prerequisites.

Elements of Photonics is comprised of two volumes. Even though cohesiveness between the two volumes is maintained, each volume can be used as a stand-alone textbook.

Volume I is devoted to topics that apply to propagation in free space and special media such as anisotropic crystals. Chapter 1 begins with a description of Fourier optics, which is used throughout the book, followed by applications of Fourier optics such as the properties of lenses, optical image processing, and holography.

Chapter 2 deals with evanescent waves, which are the basis of diffraction unlimited optical microscopes whose power of resolution is far shorter than a wavelength of light.

Chapter 3 covers the Gaussian beam, which is the mode of propagation in free-space optical communication. Topics include Bessel beams characterized by an unusually long focal length, optical tweezers useful for manipulating microbiological objects like DNA, and laser cooling leading to noise-free spectroscopy.

Chapter 4 explains how light propagates in anisotropic media. Such a study is important because many electrooptic and acoustooptic crystals used for integrated optics are anisotropic. Only through this knowledge can one properly design integrated optics devices.

Chapter 5 comprehensively treats external field effects, such as the electrooptic effect, elastooptic effect, magnetooptic effect, and photorefractive effect. The treatment includes solid as well as liquid crystals and explains how these effects are applied to such integrated optics devices as switches, modulators, deflectors, tunable filters, tunable resonators, optical amplifiers, spatial light modulators, and liquid crystal television.

Chapter 6 deals with the state of polarization of light. Basic optical phenomena such as reflection, refraction, and deflection all depend on the state of polarization of the light. Ways of converting light to the desired state of polarization from an arbitrary state of polarization are explained.

Chapter 7 explains methods of constructing and using the Poincaré sphere. The Poincaré sphere is an elegant tool for describing and solving polarization problems in the optics laboratory.

Chapter 8 covers the phase conjugate wave. The major application is for optical image processing. For example, the phase conjugate wave can correct the phasefront distorted during propagation through a disturbing medium such as the atmosphere. It can also be used for reshaping the light pulse distorted due to a long transmission distance inside the optical fiber.

Volume II is devoted to topics that apply to fiber and integrated optics.

Chapter 9 explains how a lightwave propagates through a planar optical guide, which is the foundation of integrated optics. The concept of propagation modes is fully explored. Cases for multilayer optical guides are also included.

Chapter 10 is an extension of Chapter 9 and describes how to design a rectangular optical guide that confines the light two dimensionally in the x and y directions. Various types of rectangular optical guides used for integrated optics are compared. Electrode configurations needed for applying the electric field in the desired direction are also summarized.

Chapter 11 presents optical fibers, which are the key components in optical communication systems. Important considerations in the choice of optical fibers are attenuation during transmission and dispersion causing distortion of the light pulse. Such special-purpose optical fibers as the dispersion-shifted fiber, polarization-preserving fiber, diffraction grating imprinted fiber, and dual-mode fiber are described. Methods of cabling, splicing, and connecting multifiber cables are also touched on.

Chapter 12 contains a description of light detectors for laboratory as well as communication uses. Mechanisms for converting the information conveyed by photons into their electronic counterparts are introduced. Various detectors, such as the photomultiplier tube, the photodiode, and the avalanche photodiode, and various detection methods, such as direct detection, coherent detection, homodyne detection, and detection by stimulated Brillouin scattering, are described and their performance is compared for the proper choice in a given situation.

Chapter 13 begins with a brief review of relevant topics in quantum electronics, followed by an in-depth look at optical amplifiers. The optical amplifier has revolutionized the process of pulse regeneration in fiber-optic communication systems. The chapter compares two types of optical amplifier: the semiconductor optical amplifier and the erbium-doped fiber amplifier. Knowledge gained from the operation of a single fiber amplifier is applied to the analysis of concatenated fiber amplifiers.

Chapter 14 is devoted to lasers, which is a natural extension of the preceding chapter on optical amplifiers. The chapter begins with an overview of different types of lasers,

followed by an in-depth treatment of semiconductor lasers, which are the preferred light sources for most fiber-optic communication systems. The basic relationship among the laser structure, materials, and operational characteristics are clarified. The ability to tune the laser wavelength, which is indispensible to the wavelength division multiplexing of the communication system, is addressed. The quantum well, quantum wire, and quantum dot laser diodes that have low threshold current and hence a high upper limit on the modulation frequency are also included. The erbium-doped or Raman fiber lasers that are simple in structure and easy to install in an optical fiber system are also explained.

In Chapter 15, an introduction to the nonlinear (Kerr) effect is presented. Optical devices based on the Kerr effect are controlled by photons and can respond much faster than those controlled by electrons. The chapter also provides the mechanism of formation of a soliton wave. A light pulse that propagates in an optical fiber spreads due to the dispersion effect of the fiber, but as the intensity of the pulse is increased, the nonlinear effect of the fiber starts to generate a movement directed toward the center of the light pulse. When these two counteracting movements are balanced, a soliton wave pulse that can propagate distortion-free over thousands of kilometers is formed. The attraction of distortion-free pulse propagation is that it can greatly reduce, or even eliminate, the need for pulse regenerators (repeaters) in long-haul fiber-optic communication systems.

Chapter 16 interweaves the design skills developed throughout the book with realistic problems in fiber-optic communication systems.

The problems at the end of each chapter are an integral part of the book and supplement the explanations in the text.

As a photonics textbook, each volume would be sufficient for a two-semester course. If time is really limited, Chapter 16 alone can serve as a crash course in fiber-optic communication systems and will give the student a good initiation to the subject.

For those who would like to specialize in optics, I highly recommend reading through each volume, carefully and repeatedly. Each chapter will widen your horizon of optics that much more. You will be amazed to discover how many new applications are born by adding a touch of imagination to a fundamental concept.

This two-volume work has been a long time in the making. I applaud Beatrice Shube, and George Telecki and Rosalyn Farkas of John Wiley & Sons for their superhuman patience. Sections of the manuscript went through several iterations of being written, erased, and then rewritten. As painstaking as this process was, the quality of the manuscript steadily improved with each rewrite.

I am very grateful to Professor Joseph W. Goodman of Stanford University who first suggested I publish my rough lecture notes in book form.

I am indebted especially to Mary Jean Giliberto, who spent countless hours proofreading the text, smoothing the grammatical glitches, and checking equations and numerical examples for completeness and accuracy. I greatly valued her comments and perspective during our many marathon discussions. This book was very much a partnership, in which she played a key role.

I would like to express my gratitude to Dr. Yi Fan Li, who provided much input to Chapter 15 on nonlinear optics, and Professor Subbarayan Pasupathy of the University of Toronto and Professor Alfred Wong of the University of California, Los Angeles, who critically read one of the appendixes. Frankie Wing Kei Cheng has double-checked the equations and calculations of the entire book.

I would also like to acknowledge the following students, who went through the manuscript very critically and helped to refine it: Claudio Aversa, Hany H. Loka, Benjamin Wai Chan, Soo Guan Teh, Rob James, Christopher K. L. Wah, and Megumi Iizuka.

Lena Wong's part in typing the entire manuscript should not be underestimated. I also owe my gratidue to Linda Espeut for retyping the original one-volume manuscript into the current two-volume manuscript. I wish to express my heartfelt thanks to my wife, Yoko, and children, Nozomi, Izumi, Megumi, and Ayumi, for their kind sacrifices. Ayumi Iizuka assisted in designing the cover of the book.

KEIGO IIZUKA

University of Toronto

ELEMENTS OF PHOTONICS

Volume II

9

PLANAR OPTICAL GUIDES FOR INTEGRATED OPTICS

The contents of Chapters 9 and 10 are closely interwoven. Both chapters deal with guided waves, but Chapter 9 deals primarily with guided waves in a medium bounded in one direction (taken as the x direction), while Chapter 10 deals with guided waves in a medium bounded in two directions (taken as the x and y directions). The slab optical guide is an example of a medium bounded in one direction; and the rectangular optical guide is an example of a medium bounded in two directions. For the analysis in both Chapters 9 and 10, the z direction is taken as the direction of the wave propagation.

The foundation of integrated optics is the planar optical guide. The light is guided by a medium whose index of refraction is higher than that of surrounding layers. An optical guide made of an electrooptic material changes its characteristics with a change in the applied electric field. This type of guide is very useful for fabricating electronically controllable optical switches, directional couplers, interferometers, and modulators.

According to geometrical optics, light will propagate by successive total internal reflections with very little loss provided that certain conditions are met. These conditions are that the layer supporting the propagation must have a higher refractive index than the surrounding media, and the light must be launched within an angle that satisfies total internal reflection at the upper and lower boundaries. This simple geometrical optics theory fails when the dimensions of the guiding medium are comparable to the wavelength of the light. In this regime, the guide supports propagation only for a discrete number of angles, called *modes of propagation*. In this chapter, the concept of modes of propagation is fully explored. We will explain what a mode looks like, how many modes there are, how to suppress some unwanted modes, and how to accentuate only one particular mode. This information is essential for designing an optical guide.

The chapter starts with the characteristic equation that primarily controls the mode configuration. Then, details of each mode are described. Due to the simplicity of the geometries studied, exact solutions will be obtained in many cases. Such knowledge is essential for designing the various optical waveguide configurations in the next chapter.

9.1 CLASSIFICATION OF THE MATHEMATICAL APPROACHES TO THE SLAB OPTICAL GUIDE

Slab optical guides consist of three planes and can broadly be classed into two types. One is the symmetric guide and the other is the asymmetric guide. "Slab optical guide" will be called simply "guide." The refractive indices of the top and bottom layers of the symmetric guide are identical, as indicated in Fig. 9.1, whereas those of an asymmetric guide are different. In integrated optics, both types are used. The core material of the symmetric guide is completely imbedded inside the substrate (cladding) material. The asymmetric guide consists of a film layer as the guiding core layer, with air or some other covering material as the top cladding layer, and substrate as the bottom cladding layer.

Since the mathematics dealing with a symmetric guide is much simpler than that of an asymmetric guide, the symmetric guide will be treated first for better physical insight.

The commonly used methods of analysis are the following:

1. The wave optics approach, which is the most rigorous but sometimes more complicated method.
2. The coefficient matrix approach whose manipulation is more or less mechanical and straightforward.
3. The transmission matrix method, which has the potential to be extended to solve multilayer problems.
4. The modified ray model method, which is simple but provides less information.

The wave optics approach is explained in Sections 9.2 to 9.6 and the other methods in Section 9.7.

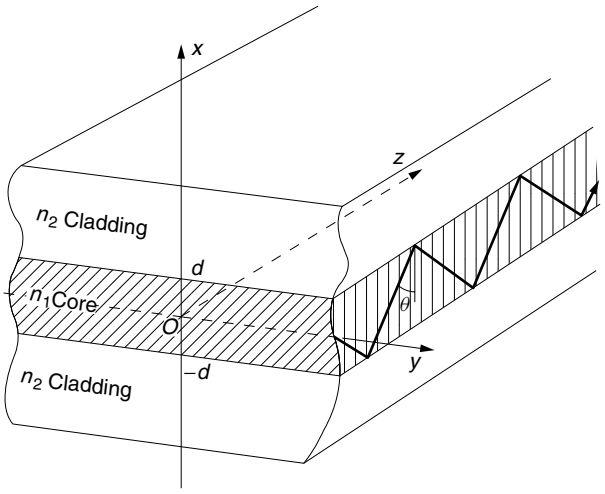


Figure 9.1 Geometry of the slab optical guide.

9.2 WAVE OPTICS APPROACH

This method starts with Maxwell's equations [1–5]. It needs no approximation and the results are rigorous. First, the field expressions are derived, followed by a derivation of the characteristic equations that are instrumental in determining the propagation modes in the guide.

Let a sinusoidally time-varying wave propagate in the z direction. The propagation constant in the z direction is β. The electric and magnetic components of the wave are expressed as

$$\begin{aligned} \mathbf{E} &= \mathbf{E}_0(x, y)e^{j(\beta z-\omega t)} \\ \mathbf{H} &= \mathbf{H}_0(x, y)e^{j(\beta z-\omega t)} \end{aligned} \tag{9.1}$$

The following two assumptions simplify the analysis:

Assumption 1. No component of the field varies in the y direction:

$$\frac{\partial}{\partial y} = 0 \tag{9.2}$$

Assumption 2. There is no magnetic field component in the z direction (TM modes),

$$H_z = 0 \tag{9.3}$$

The first assumption means that a wave is launched that does not vary in the y direction. This means that the field extends indefinitely in the y direction, but in reality there is no such wave. Note also that an infinite dimension does not guarantee Assumption 1. The infinity of the y dimension of the layer is certainly a necessary condition for $\partial/\partial y = 0$, but not a sufficient condition. For instance, a plane wave has variations even in an infinitely large medium in the direction of propagation. The manner of launching determines this condition.

The second assumption leads to a natural way of dividing the solutions, but it is not the only way to divide the solutions. The solutions are separated into two waves: one that has only transverse and no longitudinal magnetic field, that is, $H_z = 0$; and the other that has only transverse and no longitudinal electric field, that is, $E_z = 0$. The former is called *transverse magnetic* or TM mode (wave) and the latter, a *transverse electric* or TE mode. In general, a wave has both H_z and E_z components. The H_z component is accounted for by the H_z component of the TE mode and the E_z component, by the E_z component of the TM mode. The field is composed of both TM and TE modes in general. Except for Section 9.6, TM modes are assumed in this chapter.

With the assumption of Eq. (9.2), the H_y component of Eq. (9.1) is found first by inserting it into the wave equation:

$$\nabla^2\mathbf{H} + (n_{1,2}k)^2\mathbf{H} = 0 \tag{9.4}$$

$$\frac{\partial^2 H_y}{\partial x^2} + (n_{1,2}^2k^2 - \beta^2)H_y = 0 \tag{9.5}$$

Equation (9.5) is applicable for both core and cladding layers by using the respective values of n_1 or n_2 for $n_{1,2}$.

There are two kinds of solutions for the differential equation, Eq. (9.5): trigonometric solutions such as $\cos Kx$ or $\sin Kx$ for a positive value of $(n_{1,2}^2 k^2 - \beta^2)$, and exponential solutions such as $e^{\gamma x}$ or $e^{-\gamma x}$ for a negative value of $(n_{1,2}^2 k^2 - \beta^2)$. Here, only the guided wave is treated, that is, the wave whose amplitude decays with both x and $-x$. The solutions are chosen to fit the physical conditions of the guided wave. Inside the core layer, the wave is oscillatory and the trigonometric solutions are suitable. Inside the cladding layer, however, only the evanescent wave is allowed and the solution must have a decaying nature. Thus, inside the core one has

$$n_1^2 k^2 - \beta^2 = K^2 \quad |x| < d \tag{9.6}$$

and in the cladding,

$$n_2^2 k^2 - \beta^2 = -\gamma^2 \quad |x| > d \tag{9.7}$$

The range of values of β^2 that satisfy both Eqs. (9.6) and (9.7) is limited. The left-hand side of Eq. (9.6) has to be positive, while that of Eq. (9.7) has to be negative. This is especially true because the difference between n_1 and n_2 is normally a fraction of 1% of n_1. The range of β set by Eqs. (9.6) and (9.7) is

$$n_1 k > \beta > n_2 k \tag{9.8}$$

Moreover, β is allowed to take only discrete values in the above range, as will be shown later.

The solution of Eq. (9.5) with Eq. (9.6) inside the core is

$$H_y = A \cos Kx + B \sin Kx \tag{9.9}$$

and the solution of Eq. (9.5) with Eq. (9.7) inside the cladding layer is

$$H_y = Ce^{-\gamma x} + De^{\gamma x} \tag{9.10}$$

where the factor $e^{j(\beta z - \omega t)}$ was suppressed.

The next step is to find the constants A, B, C, and D using the boundary conditions. In the upper cladding layer, D has to be zero (note that zero is also a legitimate constant) so as to prevent H_y from becoming infinitely large as x approaches $+\infty$. Using the same reasoning, C has to be zero in the lower cladding layer.

$$H_y = \begin{cases} Ce^{-\gamma x}, & x > d \\ De^{\gamma x}, & x < -d \end{cases} \tag{9.11}$$

Matters are simplified if Eq. (9.9) is separated into two parts:

$$H_y = A \cos Kx \tag{9.12}$$

$$H_y = B \sin Kx \tag{9.13}$$

In the end, the two solutions are combined to reach the final solution. Equation (9.12) is called the *even-mode solution*, and Eq. (9.13) is the *odd-mode solution*, simply because $\cos Kx$ is an even function of x (i.e., $\cos(-Kx) = \cos Kx$), and $\sin Kx$ is an odd function of x (i.e., $\sin(-Kx) = -\sin Kx$). This way of separating the solutions into

two is quite natural. If the slab optical guide is excited with an incident wave whose amplitude distribution is symmetric with respect to x, A is nonzero and B is zero. B is nonzero and A is zero for a perfectly antisymmetric incident amplitude distribution.

In order to determine the values of the constants, the boundary condition of the continuity of the tangential **H** field is used at $x = d$, and from Eqs. (9.11) and (9.12), this boundary condition for the even TM modes gives

$$A \cos Kd = Ce^{-\gamma d} \tag{9.14}$$

Putting this equation back into Eq. (9.11) gives

$$H_y = A(\cos Kd)e^{-\gamma(x-d)} \tag{9.15}$$

An expression for the lower cladding layer is obtained using the boundary condition at $x = -d$. The results for the even TM modes are summarized as

$$H_y = \begin{cases} A(\cos Kd)e^{-\gamma(x-d)} & \text{in the upper cladding} \\ A \cos Kx & \text{in the core} \\ A(\cos Kd)e^{\gamma(x+d)} & \text{in the lower cladding} \end{cases} \tag{9.16}$$

Expressions for the odd modes are obtained by starting with Eq. (9.13) instead of Eq. (9.12) and following the same procedure. The results for the odd TM modes are summarized as

$$H_y = \begin{cases} B(\sin Kd)e^{-\gamma(x-d)} & \text{in the upper cladding} \\ B \sin Kx & \text{in the core} \\ -B(\sin Kd)e^{\gamma(x+d)} & \text{in the lower cladding} \end{cases} \tag{9.17}$$

Next, E_x and E_z are obtained from H_y using Maxwell's equations:

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \tag{9.18}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{9.19}$$

With Eqs. (9.1), (9.2), and (9.18), E_x and E_z are

$$E_x = \frac{\beta}{\omega \epsilon_r \epsilon_0} H_y \tag{9.20}$$

$$E_z = \frac{j}{\omega \epsilon_r \epsilon_0} \frac{\partial H_y}{\partial x} \tag{9.21}$$

The value of ϵ_r depends on the medium and

$$\epsilon_r = \begin{cases} n_1^2 & \text{in the core} \\ n_2^2 & \text{in the cladding} \end{cases} \tag{9.22}$$

Now, all field components of the TM modes have been obtained. They are summarized as

$$\begin{aligned} E_x &= \frac{\beta}{\omega\epsilon_r\epsilon_0} H_y \\ E_y &= 0 \\ E_z &= \frac{j}{\omega\epsilon_r\epsilon_0}\frac{\partial H_y}{\partial x} \\ H_x &= 0 \\ H_y &= A\cos Kx + B\sin Kx \\ H_z &= 0 \end{aligned} \tag{9.23}$$

where $B = 0$ for the even TM modes, and $A = 0$ for the odd TM modes. The equation $E_y = 0$ was derived from Maxwell's equations as follows. The y component of Eq. (9.18) and the x component of Eq. (9.19) are combined to give

$$E_y(\omega^2\mu\epsilon_r\epsilon_0 - \beta^2) = 0 \tag{9.24}$$

Since the value inside the parentheses is K^2 from Eq. (9.6) and is nonzero, $E_y = 0$. $H_x = 0$ was derived by inserting $E_y = 0$ into the x component of Eq. (9.19).

9.3 CHARACTERISTIC EQUATIONS OF THE TM MODES

In the previous section, not much was said about the actual values of K^2 and $-\gamma^2$ except that the former is a positive number and the latter, a negative number. The values of K and γ are crucial to determining the modes of propagation. Some more boundary conditions are used to find these values.

9.3.1 Solutions for K and γ

First, the even TM modes are considered. Continuity of the tangential $\mathbf{E}$ field, E_z in Eq. (9.23), at $x = d$ requires that

$$n^2 AK\sin Kd = \gamma C e^{-\gamma d} \tag{9.25}$$

where Eqs. (9.11), (9.12), and (9.21) and $n = \sqrt{\epsilon_{r2}/\epsilon_{r1}} = n_2/n_1$ were used. Dividing Eq. (9.25) by Eq. (9.14) gives

$$n^2 Kd\tan Kd = \gamma d \tag{9.26}$$

Equation (9.26) is called the *characteristic equation* for the even TM modes. The characteristic equation is used to find the solutions for K and γ.

We need one more equation to find the values of K and γ. From Eqs. (9.6) and (9.7), β is eliminated to obtain

$$(Kd)^2 + (\gamma d)^2 = V^2 \tag{9.27}$$

where

$$V = kd\sqrt{n_1^2 - n_2^2} \tag{9.28}$$

Since V consists of only physical constants such as the height of the guide, the indices of refraction, and the light wavelength, V is referred to as the *normalized thickness of the guide*. The normalized thickness V is an important parameter specifying the characteristics of the guide.

Equations (9.26) and (9.27) are transcendental equations and the solution cannot be found in a closed form. Graphical solutions are available, as shown in Fig. 9.2. Equations (9.26) and (9.27) are plotted on the Kd–γd plane as solid lines in Fig. 9.2. Note that Eq. (9.27) is a circle with radius V. The shape of the curve for Eq. (9.26) is quite similar to tan Kd. Each intersection point shown in Fig. 9.2 corresponds to a solution, or mode of propagation, of an even TM mode. These intersections are called, for short, the even TM modes.

The corresponding characteristic equation for the odd TM modes is obtained using Eq. (9.13) instead of Eq. (9.12). The continuity of H_y at $x = d$ gives

$$B \sin Kd = Ce^{-\gamma d} \tag{9.29}$$

and the continuity of E_z in Eq. (9.23) gives

$$n^2 BK \cos Kd = -C\gamma e^{-\gamma d} \tag{9.30}$$

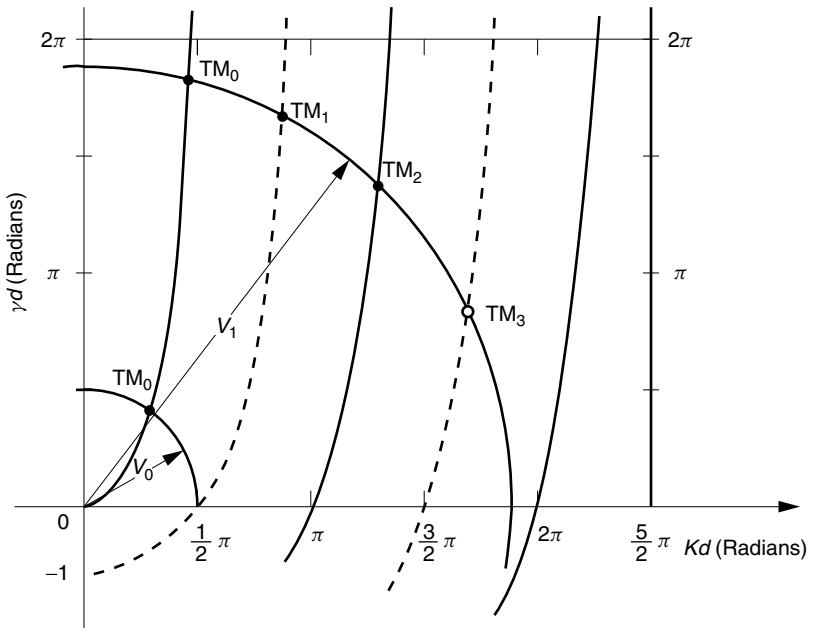


Figure 9.2 Graphical solutions for the even TM modes ($\gamma d = n^2 Kd \tan Kd$) in solid lines, and odd TM modes ($-\gamma d = n^2 Kd \cot Kd$) in dashed lines. $V = kd\sqrt{n_1^2 - n_2^2}$.

Division of Eq. (9.30) by Eq. (9.29) renders the characteristic equation for the odd TM modes as

$$-n^2 Kd \cot Kd = \gamma d \tag{9.31}$$

Equation (9.31) is plotted as the dashed lines in Fig. 9.2.

The intersections between the solid lines and the circle are the even (order) TM modes and those between the dashed lines and the circle are the odd (order) TM modes. Even numbered subscripts are used for even modes and odd numbers for odd modes. The subscript is called the *order of the mode* or the *mode number*.

With a decrease in V, the number of modes that the guide can support decreases one by one. For example, referring to Fig. 9.2, an optical guide whose normalized thickness is V_1 can support four modes, whereas that with V_0 can support only one mode. For every $\pi/2$-radian decrease in V, even and odd modes alternately disappear. The disappearance of a particular mode is called the *cutoff* of that mode. For instance, the cutoff condition for the TM_3 mode is $V = \frac{3}{2}\pi$. As the cutoff condition is approached, the value of γ approaches zero, and the effective depth of the evanescent wave (Section 2.8) in the cladding layer increases. When γ reaches zero, the evanescent wave is present throughout the cladding layer and the light energy cannot be sustained inside the core.

As long as V is greater than $\pi/2$, more than one mode can be excited simultaneously. Among the excited modes, the higher order modes are more susceptible to the conditions outside the guide because γ is smaller and the effective depth of the evanescent wave is deeper.

If V is less than $\pi/2$, there exists only one mode and no other modes can be excited. The mode that is capable of being the only excited mode is called the *dominant mode*. The dominant mode of the TM modes in the slab optical guide is the TM_0 mode. Note also that there is no cutoff for the TM_0 mode, which remains excited down to $V = 0$.

A slab optical guide that exclusively supports the dominant mode is called a single-mode guide or monomode guide. Guides that support more than one mode are called multimode guides. When light is launched into a multimode guide such that several modes are excited, then the incident light power is divided among the excited modes. Each mode, however, has a different propagation constant β_N. Thus, each mode arrives at the receiving point at a different phase and the signal is distorted. This distortion is called *mode dispersion*. A signal in a monomode guide is not distorted by mode dispersion.

Since the number of modes increases for every $\pi/2$-radian increase in V, the highest mode number N is the largest integer that still satisfies

$$\frac{2\pi}{\lambda} d \sqrt{n_1^2 - n_2^2} > N \frac{\pi}{2} \tag{9.32}$$

and the total number of TM modes including the zero-order mode is $N + 1$.

9.3.2 Examples Involving TM Modes

Example 9.1 Optical communication systems are normally operated at a light wavelength of 0.85, 1.3, or 1.55 μm. These wavelengths are outside the visible range. A

He–Ne laser emitting light at 0.63 μm is often used to test devices because 0.63 μm lies in the visible wavelength range.

A symmetric guide designed to be monomode for a wavelength of 1.3 μm was excited by a He–Ne laser. At most, how many TM modes will be excited by the He–Ne laser in this guide?

Solution According to Fig. 9.2, the largest normalized thickness V of a single-mode guide is $\pi/2$. Setting $V = \pi/2$ at 1.3 μm, the thickness d of such a guide is

$$\frac{2\pi}{1.3} d\sqrt{n_1^2 - n_2^2} = \frac{\pi}{2}$$

With this d, for 0.63 μm, the normalized thickness becomes

$$\begin{aligned} V &= \frac{2\pi}{0.63} d\sqrt{n_1^2 - n_2^2} \\ &= \frac{\pi}{2} \cdot \frac{1.3}{0.63} \\ &= 1.03\pi \text{ radians} \end{aligned}$$

From Fig. 9.2, at most three modes are excited. □

Example 9.2 The TM_2 mode in a symmetric guide was observed to be cut off when the wavelength was increased beyond 1.5 μm. The refractive index of the core is $n_1 = 1.55$ and that of the cladding is $n_2 = 1.54$.

(a) What is the thickness $2d$ of the guide?
(b) What is K_2 at the cutoff?
(c) What is β_2 at the cutoff?

The subscripts on K and β refer to the mode number.

Solution

(a) From Fig. 9.2, the TM_2 mode has its cutoff at $V = \pi$.

$$\frac{2\pi}{\lambda} d\sqrt{n_1^2 - n_2^2} = \pi$$

$$2d = \frac{\lambda}{\sqrt{n_1^2 - n_2^2}} = \frac{1.5}{\sqrt{1.55^2 - 1.54^2}} = 8.53 \text{ μm}$$

(b) At the cutoff, the normalized thickness is

$$V = K_2 d = \pi$$

and the value of K_2 is

$$K_2 = 0.74 \text{ rad/μm}$$

(c) At the cutoff, $\gamma_2 = 0$. Putting $\gamma_2 = 0$ in Eq. (9.7), one finds that β_c at the cutoff is $n_2 k$:

$$\beta_c = 6.4 \text{ rad}/\mu\text{m}$$

This relationship can also be derived from Eq. (9.6). Expressing Eq. (9.6) in terms of β_2 gives

$$\beta_2 = \sqrt{(n_1 k)^2 - K_2^2}$$

At the cutoff, $V = K_2 d$ and

$$K_2^2 = k^2(n_1^2 - n_2^2)$$

Inserting this K_2^2 value into the equation for β_2 gives β_c at the cutoff:

$$\beta_c = n_2 k$$

It is interesting to note that at the cutoff, the propagation constant is $n_2 k$ regardless of the order of the mode. □

Example 9.3 Show that if the slab guide can be excited up to the $(N_{\max})$th $\text{TM}_{N_{\max}}$ mode, the propagation constant β_N for the Nth ($N \ll N_{\max}$) TM_N mode can be approximated as

$$\beta_N = n_1 k \sqrt{1 - 2\Delta \left(\frac{N+1}{N_{\max}}\right)^2}$$

where

$$\Delta = \frac{n_1 - n_2}{n_1} \ll 1$$

$$N \ll N_{\max}$$

Solution The propagation constant β_N for the Nth TM mode is, from Eq. (9.6),

$$\beta_N = n_1 k \sqrt{1 - \left(\frac{K_N}{n_1 k}\right)^2} \tag{9.33}$$

The values of K_N in the region of

$$N \ll N_{\max}$$

are, from Fig. 9.2,

$$K_N d = \frac{\pi}{2}(N+1) \tag{9.34}$$

The missing information is d. The order of the highest mode is $N_{\max}$ and the corresponding V is

$$V = \frac{\pi}{2} N_{\max} \tag{9.35}$$

From Eq. (9.28) and the condition imposed on Δ, V can be expressed as

$$V = n_1 kd\sqrt{2\Delta} \tag{9.36}$$

Inserting the value of d obtained from Eqs. (9.35) and (9.36) into Eq. (9.34) gives the value for K_N. The final result is obtained by inserting this K_N value into Eq. (9.33), giving

$$\beta_N = n_1 k\sqrt{1 - 2\Delta\left(\frac{N+1}{N_{\max}}\right)^2} \tag{9.37}$$

□

9.4 CROSS-SECTIONAL DISTRIBUTION OF LIGHT AND ITS DECOMPOSITION INTO COMPONENT PLANE WAVES

The cross-sectional distribution of light in the guide is given by Eq. (9.16) or (9.17) and is shown in Fig. 9.3. The intensity distribution of the Nth-order TM mode has $N + 1$ loops and N nulls in the core layer. The shapes of the curves in Fig. 9.3 are determined by the value of K. If the even-mode K_N are far from the cutoff and are almost an odd multiple of $\pi/2$, then the field vanishes at $x = d$; but if the values of K_N for the even modes are smaller than an odd multiple of $\pi/2$, then the field at $x = d$ becomes a finite value. In fact, these finite values determine the amplitude of the evanescent field in the cladding layer.

It will be shown that the mode patterns are nothing but the standing-wave pattern produced by the interference of component plane waves zigzagging inside the optical guide. If the trigonometric cosine function is rewritten in exponential functions, Eq. (9.12) combined with Eq. (9.1) becomes

$$H_y = \tfrac{1}{2}A(e^{j(Kx+\beta z-\omega t)} + e^{j(-Kx+\beta z-\omega t)}) \tag{9.38}$$

Note from Chapter 1 that a plane wave propagating in the $\mathbf{k} = (k_x, k_y, k_z)$ direction is expressed by

$$e^{j\mathbf{k}\cdot\mathbf{r}} = e^{jk_x x + jk_y y + jk_z z}$$

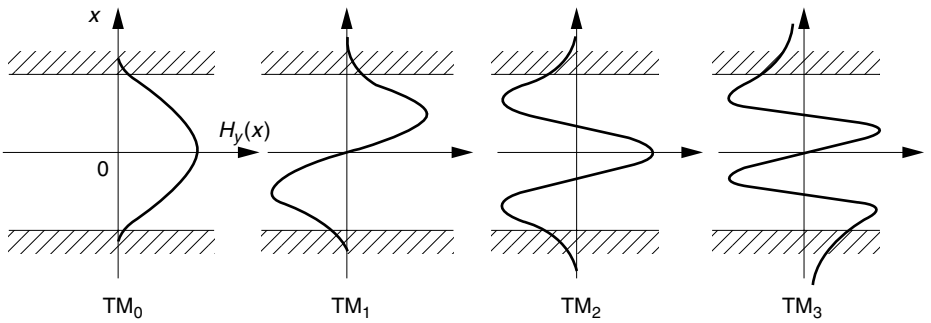


Figure 9.3 Distribution of the H_y field in the slab optical guide. The field distributions correspond to the modes in Fig. 9.2.

The first term of Eq. (9.38) is a component plane wave propagating slightly upward in the direction connecting the origin and a point (K, β) in the $x-z$ plane. The second term is a similar plane wave but propagating slightly downward in the direction connecting the origin and a point $(-K, \beta)$ in the same plane.

The interference of these two component plane waves is nothing but the field distribution of the mode. The spacing between adjacent null contour lines changes as the angle between the two plane waves is changed. This behavior can be demonstrated by drawing phase lines on two sheets of transparent paper, and then placing one over the top of the other as shown in Fig. 9.4. The phasefronts of these plane waves are designated by two kinds of lines. The 0° phase line is represented by a heavy line and the 180° phase line is represented by a fine line. The intersections of two heavy lines are the points of maximum amplitude, while the intersections of two fine lines are the points of minimum amplitude (negative extrema). The location where a fine line meets a heavy line indicates a null amplitude. The contours of the amplitude extrema and amplitude nulls are lines parallel to the z axis and alternate. The cross-sectional distribution along the x axis would look like the sinusoidal curve indicated on the right-hand side of Fig. 9.4. This sinusoidal curve is nothing but the mode pattern in the guide.

The spacing between the null amplitude contour lines starts to contract from the maximum of infinite distance to one-half wavelength by rotating from the parallel position to the perpendicular position. The angles between the component waves that can match the boundary condition are found by adjusting the rotation of the sheet. The boundary condition is met by lining up the null contours close to the upper and lower boundaries of the guide (to be exact, slightly inside the cladding layer due to the evanescent wave). For a given value $2d$ of the guide, the boundary conditions are

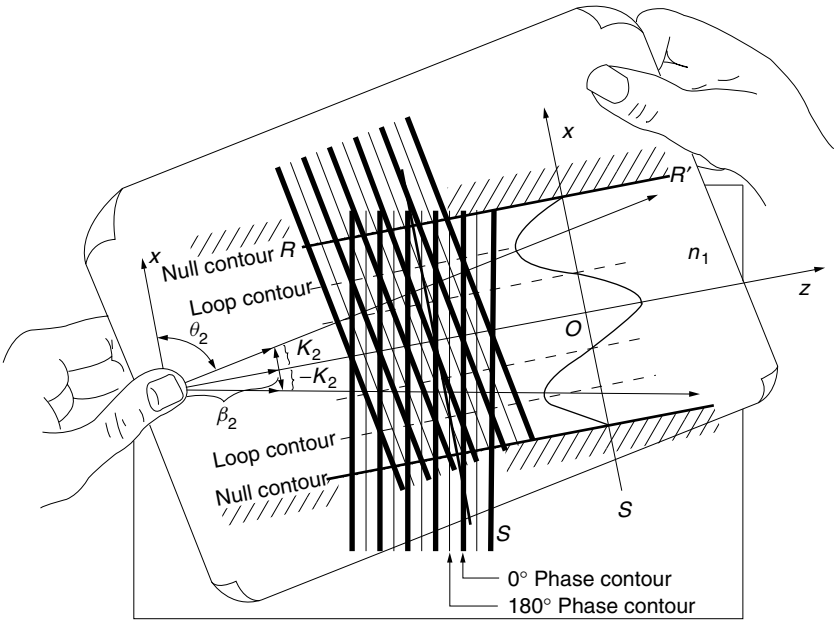


Figure 9.4 Composition of the TM_2 mode in terms of component plane waves. The effective index of refraction is $N = n_1 \sin\theta_2$.

satisfied only for a discrete number of angles θ_N. Conversely, for a given angle θ_1, the values of $2d$ that satisfy the boundary condition are discrete.

Because the field does not become exactly null on the boundary but goes into the cladding layer as an evanescent wave, there is some inaccuracy in situating the null contour line by this method. This inaccuracy can be removed by finding the value of K from Fig. 9.2 and then using Eq. (9.6) to find β and hence the angle θ_N. In this way, the directions of propagation that satisfy the boundary conditions are accurately determined.

The relationship between K and β in Eq. (9.6) is graphically represented by the K–β circle of $K^2 + \beta^2 = (n_1 k)^2$ as shown in Fig. 9.5. By using K from Fig. 9.2, β is found from Fig. 9.5. The direction of propagation is a vector connecting the point (K, β) and the origin, as shown in Fig. 9.5. The extensions of these vectors determine the directions of the component waves in the guide, shown on the right-hand side. The discrete angles θ_N of propagation are, from Fig. 9.5,

$$\theta_N = \sin^{-1}\left(\frac{\beta_N}{n_1 k}\right) \tag{9.39}$$

The range of allowed values of β is limited. The maximum value of β is $n_1 k$ from Fig. 9.5. The minimum value of β is determined by the critical angle θ_c associated with the boundary between the core and cladding layer as

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right) \tag{9.40}$$

From Eqs. (9.39) and (9.40), the minimum value of β is $n_2 k$. The allowed range of β agrees with the earlier results of Eq. (9.8). The prohibited region is to the left of the $n_2 k$ line in Fig. 9.5. In a typical glass guide, $n_2/n_1 = 1.54/1.55$ and $\theta_c = 83.5°$. The allowed range is quite small, only 6.5°.

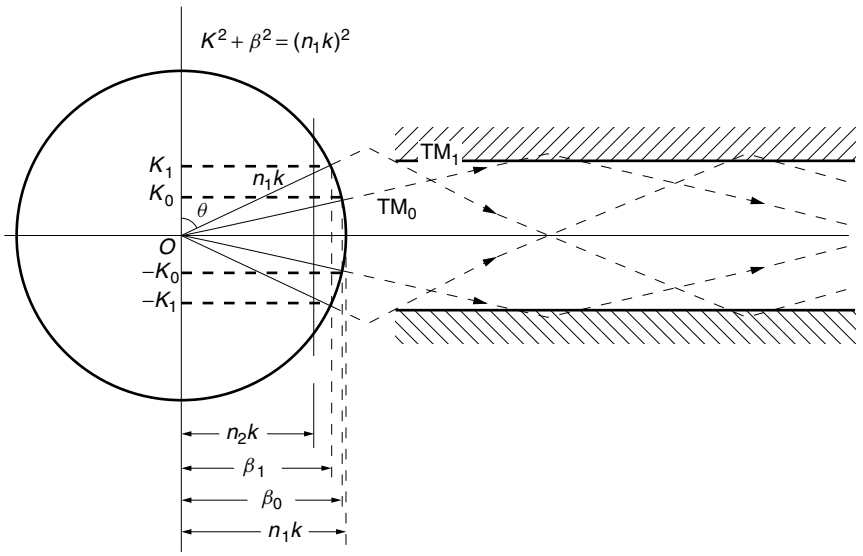


Figure 9.5 K–β circle and component waves.

Example 9.4 Using the TM_2 mode as an example, describe how the following quantities change as the thickness $2d$ of an optical guide is decreased: (a) K_2, (b) the incident angle θ_2 to the core–cladding interface, and (c) the cross-sectional field distribution of H_y.

Solution Figure 9.6 summarizes the results.

(a) The value of Kd is determined by the length p_2 shown in Fig. 9.6a:

$$K_2 = \frac{p_2}{d} \tag{9.41}$$

With a decrease in d, the value of $V = kd\sqrt{n_1^2 - n_2^2}$ decreases and hence p_2 decreases. With a decrease in d, both the numerator and denominator of Eq. (9.41) decrease. But

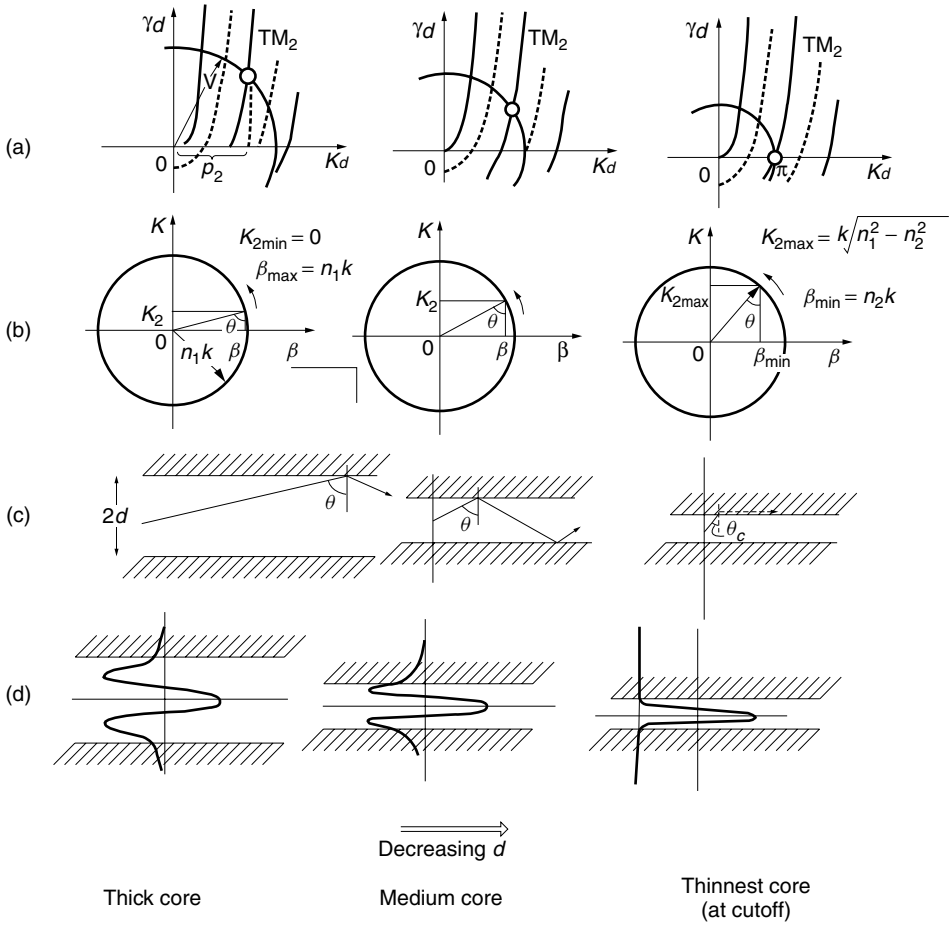


Figure 9.6 Changes to the parameters of the TM_2 mode in a slab optical guide with respect to its core thickness. The left column is the thick core and the right column is the thinnest (at cutoff) core. The free-space wavelength is kept fixed. (a) K–γ diagram. (b) K–β diagram. (c) Component rays. (d) Field pattern in the guide.

as seen from Fig. 9.6a, the decrease in p_2 is much slower than the decrease in d itself (keep in mind that V and d are linearly related) so that the value of K_2 in fact increases with a decrease in d.

(b) With an increase in K_2, the incident angle $\theta_2 = \cos^{-1}(K_2/n_1k)$ decreases. If the value of θ_2 should fall below the critical angle θ_c, then the TM_2 mode ceases to propagate since it leaves the realm of total internal reflection. This can also be explained using Fig. 9.4. With a decrease in $2d$, the null contour line $\overline{RR'}$ has to come down; hence, the angle θ_2 has to be decreased.

(c) Figure 9.6d shows the changes in the cross-sectional distribution of the field. The sinusoidal wave inside the core slightly contracts in the x direction, whereas the evanescent wave in the cladding expands significantly with a decrease in d. □

9.5 EFFECTIVE INDEX OF REFRACTION

First, the concept of the propagation constant β in $e^{j\beta z}$ is reviewed. The propagation constant β is the rate of advance in phase for unit distance (not for unit time) of advance in z. It means that the shorter the wavelength is, the larger the rate of advance in phase for a given distance, and the larger the value of β is.

In free space, the propagation constant is

$$k = \frac{2\pi}{\lambda}$$

where λ is the wavelength in vacuum. If the free space is now filled with a medium with refractive index n_0, the wavelength will be contracted to λ/n_0 and the rate of advance in phase *per distance* will become larger: namely, the propagation constant β_0 in such a medium is

$$\beta_0 = \frac{2\pi}{\lambda/n_0} = n_0 k \tag{9.42}$$

Conversely, the refractive index of the filling medium is given by

$$n_0 = \frac{\beta_0}{k} \tag{9.43}$$

The index of refraction is the ratio of the propagation constant of the medium to that of the vacuum.

Now consider the optical guide. The propagation constant depends on the direction; for instance, it is K in the x direction and it is β in the z direction. In many instances, it is propagation in the z direction that is of principal interest. In the z direction, the guide may be treated as if it were free space filled with a medium with an index of refraction N, defined as

$$N = \frac{\beta}{k} \tag{9.44}$$

where N is called the *effective index of refraction*. The effective index of refraction can be found from Fig. 9.5:

$$\beta = n_1 k \sin\theta$$

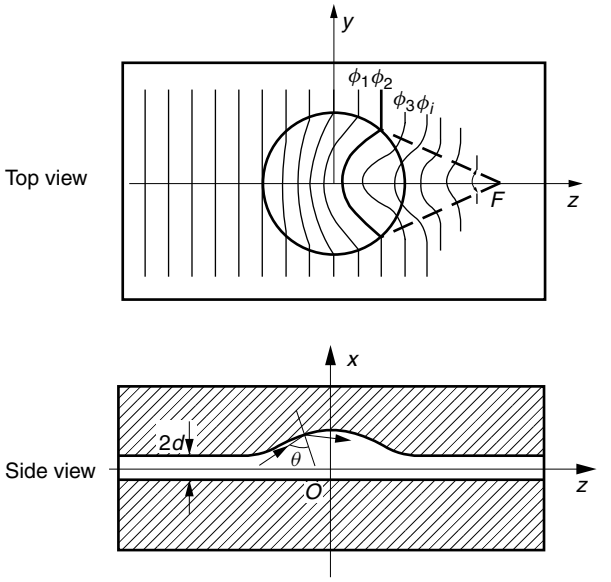


Figure 9.7 Mode-index lens used for integrated optics devices.

and

$$N = n_1 \sin\theta \tag{9.45}$$

Modulation of θ has the same effect as modulating the index of refraction. This effect is used to fabricate integrated optics devices, such as the mode-index lens on a slab guide, shown in Fig. 9.7. When a mode in the guide encounters an increase in guide thickness, the angle θ increases. (This is illustrated in Fig. 9.6b by examining the figure from the right to the left direction, which is the direction of increasing d.) As θ increases, so does the effective refractive index N. For the mode-index lens shown in Fig. 9.7, the advance in constant phase lines is slowed down the most where N is the greatest, and the spacing between the constant phase lines narrows. The result is that the phasefronts of the transmitted wave converge toward the point F.

9.6 TE MODES

The wave optics approach applied earlier to the TM modes is applied in a similar manner to the TE modes. Since the approach is so similar, only the key formulas are repeated for the benefit of summarizing the approach.

The TE modes have only transverse and no longitudinal electric field:

$$E_z = 0 \tag{9.46}$$

Again, the same assumption is made that the guide is excited such that no variation of the field occurs in the y direction, namely,

$$\frac{\partial}{\partial y} = 0 \tag{9.47}$$

With an assumed solution

$$E_y = E_y(x, y)e^{j\beta z - j\omega t} \tag{9.48}$$

the wave equation becomes

$$\frac{1}{E_y}\frac{\partial^2 E_y}{\partial x^2} + [(n_{1,2}k)^2 - \beta^2] = 0 \tag{9.49}$$

The types of solutions of this differential equation depend on the sign of the second term.

If the second term in Eq. (9.49) is a positive constant, the solutions becomes trigonometric functions; whereas if it is a negative constant, the solutions become hyperbolic or exponential functions. In order to suit the physical conditions, the former solutions are used inside the core and the latter inside the cladding:

$$\begin{aligned} (n_1k)^2 - \beta^2 &= K^2 \quad \text{in core} \\ (n_1k)^2 - \beta^2 &= -\gamma^2 \quad \text{in cladding} \end{aligned} \tag{9.50}$$

The solutions of Eq. (9.49) with Eq. (9.50) are

$$E_y = \left\{ \begin{matrix} A\cos Kx, & \text{even TE mode} \\ B\sin Kx, & \text{odd TE mode} \end{matrix} \right\} \quad \text{in core} \tag{9.51}$$

$$E_y = \left\{ \begin{matrix} Ce^{-\gamma x}, & \text{upper layer} \\ De^{\gamma x}, & \text{lower layer} \end{matrix} \right\} \quad \text{in cladding} \tag{9.52}$$

The same notations A, B, C, and D that are used for the TM modes are used for the TE modes, but there is no connection. A factor of $e^{j\beta z - j\omega t}$ is suppressed.

Maxwell's equation $\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t$ is used to find H_z:

$$H_z = \frac{1}{j\omega\mu_{1,2}}\frac{\partial E_y}{\partial x} \tag{9.53}$$

where $\mu_{1,2}$ represents the magnetic permeability of the core and cladding layers. Applying Eqs. (9.51) and (9.52) to Eq. (9.53) gives

$$H_z = \frac{K}{j\omega\mu_1}\left\{ \begin{matrix} -A\sin Kx, & \text{even TE mode} \\ B\cos Kx, & \text{odd TE mode} \end{matrix} \right\} \quad \text{in core} \tag{9.54}$$

$$H_z = \frac{\gamma}{j\omega\mu_2}\left\{ \begin{matrix} -Ce^{-\gamma x}, & \text{upper layer} \\ De^{\gamma x}, & \text{lower layer} \end{matrix} \right\} \quad \text{in cladding} \tag{9.55}$$

The other components of the TE modes [5] are

$$\begin{aligned} H_x &= -\frac{\beta}{\omega\mu}E_y \\ E_x &= H_y = E_z = 0 \end{aligned} \tag{9.56}$$

Now, the characteristic equations will be found using the boundary conditions. First, the characteristic equation of the even modes is obtained. From the continuity of E_y at $x = d$,

$$A \cos Kd = Ce^{-\gamma d} \tag{9.57}$$

and from the continuity of H_z at the same boundary,

$$-\frac{K}{j\omega\mu_1} A \sin Kd = -\frac{\gamma}{j\omega\mu_2} Ce^{-\gamma d} \tag{9.58}$$

The division of the above two equations finally gives the characteristic equation for the even TE modes.

$$\left(\frac{\mu_2}{\mu_1}\right) Kd \tan Kd = \gamma d \tag{9.59}$$

Similarly, the continuity of the odd modes gives the characteristic equation for the odd TE modes.

$$-\left(\frac{\mu_2}{\mu_1}\right) Kd \cot Kd = \gamma d \tag{9.60}$$

The difference between the characteristic equations for the TE modes and those for the TM modes is in the factor. The factor for the TE modes is μ_2/μ_1, while that for the TM modes is $(n_2/n_1)^2$. Unless magnetic material is used, $\mu_2/\mu_1 = 1$ and, for the TE modes, the characteristic equations are

$$\begin{aligned} Kd \tan Kd &= \gamma d, \quad \text{even TE mode} \\ -Kd \cot Kd &= \gamma d, \quad \text{odd TE mode} \end{aligned} \tag{9.61}$$

Since the factor $(n_2/n_1)^2$ for the TM modes is close to unity but slightly smaller than unity, the curves of the characteristic equations of the TE modes are almost identical with those of the TM modes shown in Fig. 9.2, but the curves for the TE modes are slightly higher than those of the corresponding TM modes.

9.7 OTHER METHODS FOR OBTAINING THE CHARACTERISTIC EQUATIONS

In Section 9.3, the characteristic equation was derived by the method of wave optics. In this section, a few other available methods are presented. Since the methods will be explained using the simplest geometry, there is little basis for preferring one method over the other; but for more complicated multilayer guides, the choice of the method makes a marked difference in complexity of the treatment. For the sake of comparison, the same geometry and the familiar TM wave are used for all cases.

The three methods described in this section for obtaining the characteristic equation are (1) the coefficient matrix method, (2) the transmission matrix method, and (3) the modified ray model method.

9.7.1 Coefficient Matrix Method

This method is straightforward, but the coefficient matrix is not well suited to multilayer problems because the size of the matrix becomes too cumbersome to manipulate.

The geometry shown Fig. 9.1 is used. The solution for H_y inside the core was given by Eq. (9.9), and that inside the cladding was given by Eq. (9.10). As the first step, terms such as $De^{\gamma x}$ in the upper cladding, which are obviously unfit for physical reasons, are removed. The unfit terms would eventually have been removed automatically by the method, but early removal shortens the procedure.

The constants A, B, C, and D are to be determined using the boundary conditions of continuity of the tangential components of both $\mathbf{E}$ and $\mathbf{H}$ at $x = d$ and $x = -d$. The general solution for H_y in the three layers of a symmetric guide is summarized as

$$\begin{array}{ll} & H_{y2} = Ce^{-\gamma x} \\ x = d: & \hline \\ & H_{y1} = A\cos Kx + B\sin Kx \\ x = -d: & \hline \\ & H_{y3} = De^{\gamma x} \end{array}$$

The tangential $\mathbf{H}$ field is continuous at $x = d$, giving

$$A\cos Kd + B\sin Kd = Ce^{-\gamma d} \tag{9.62}$$

and similarly at $x = -d$,

$$A\cos Kd - B\sin Kd = De^{-\gamma d} \tag{9.63}$$

Since there are four unknown constants, two more independent equations are needed to find a solution. Continuity of the tangential components of the $\mathbf{E}$ field is used. With the help of Eq. (9.21), the E_z fields in the three regions are

$$\begin{array}{ll} & E_{z2} = \dfrac{-jC\gamma}{\omega\epsilon_0\epsilon_{r2}}e^{-\gamma x} \\ x = d: & \hline \\ & E_{z1} = \dfrac{jK}{\omega\epsilon_0\epsilon_r 1}(-A\sin Kx + B\cos Kx) \\ x = -d: & \hline \\ & E_{z3} = \dfrac{jD\gamma}{\omega\epsilon_0\epsilon_{r2}}e^{\gamma x} \end{array}$$

In the top and bottom layers, the relative dielectric constants are ϵ_{r2}, whereas in the core the dielectric constant is ϵ_{r1}. The continuity of the tangential components of the $\mathbf{E}$ field at $x = d$ means that

$$\frac{K}{\epsilon_{r1}}(-A\sin Kd + B\cos Kd) = \frac{-C}{\epsilon_{r2}}\gamma e^{-\gamma d} \tag{9.64}$$

The equation analogous to Eq. (9.64) at the lower $x = -d$ boundary is

$$\frac{K}{\epsilon_{r1}}(A \sin Kd + B \cos Kd) = \frac{D}{\epsilon_{r2}} \gamma e^{-\gamma d} \tag{9.65}$$

Equations (9.62) to (9.65) can be put into matrix form as

$$\begin{matrix} H_y(x=d) \\ H_y(x=-d) \\ E_z(x=d) \\ E_z(x=-d) \end{matrix} \begin{bmatrix} \cos Kd & \sin Kd & -e^{-\gamma d} & 0 \\ \cos Kd & -\sin Kd & 0 & -e^{-\gamma d} \\ -K' \sin Kd & K' \cos Kd & \gamma' e^{-\gamma d} & 0 \\ K' \sin Kd & K' \cos Kd & 0 & -\gamma' e^{-\gamma d} \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = 0 \tag{9.66}$$

where

$$K' = K\frac{1}{\epsilon_{r1}} = \frac{K}{n_1^2}, \quad \gamma' = \gamma\frac{1}{\epsilon_{r2}} = \frac{\gamma}{n_2^2} \tag{9.67}$$

Since the right-hand side of the equation is zero, the determinant has to be zero if a nonzero solution for A, B, C, and D is to exist. The determinant is the characteristic equation. Factoring Eq. (9.66) gives

$$\Delta = e^{-2\gamma d} \cos Kd \sin Kd \begin{vmatrix} 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & -1 \\ -K' \tan Kd & K' \cot Kd & \gamma' & 0 \\ K' \tan Kd & K' \cot Kd & 0 & -\gamma' \end{vmatrix} \tag{9.68}$$

Multiplying the second row by γ', and then subtracting the fourth row from the second row leads to a matrix with the fourth column all zero except at the bottom. The determinant reduces to

$$\Delta = -\tfrac{1}{2} e^{-2\gamma d} \sin 2Kd \begin{vmatrix} 1 & 1 & -1 \\ \gamma' - K' \tan Kd & -\gamma' - K' \cot Kd & 0 \\ -K' \tan Kd & K' \cot Kd & \gamma' \end{vmatrix} \tag{9.69}$$

By multiplying the first row by γ' and then adding the third row to the first row, the determinant reduces to a 2×2 matrix and the characteristic equation finally becomes

$$\Delta = e^{-2\gamma d} \sin 2Kd(\gamma' - K' \tan Kd)(\gamma' + K' \cot Kd) = 0 \tag{9.70}$$

From Eq. (9.70), the determinant is zero when

$$2Kd = n\pi \tag{9.71}$$

or

$$\gamma d = \left(\frac{n_2}{n_1}\right)^2 Kd \tan Kd \tag{9.72}$$

or

$$\gamma d = -\left(\frac{n_2}{n_1}\right)^2 Kd \cot Kd \tag{9.73}$$

Thus, the same results as those in Section 9.3 are obtained. The values of Kd given by Eq. (9.71) are on the cutoff points of the TM modes.

9.7.2 Transmission Matrix Method (General Guides)

This method is especially suited for treating multilayer guides [6,7], such as shown in Fig. 9.8, because, regardless of the number of layers in the guide, the matrices dealt with are always 2×2 matrices. The guide is treated layer by layer, and in each layer the field is expressed by a 2×2 matrix. Using the TM case as an example, the 2×2 matrix gives the relationship between the field $[H_y(x'), E_z(x')]$ at $x = x'$ and the field $[H_y(x), E_z(x)]$ at $x = x$ in that same layer. Using the condition of the continuity of the tangential fields at the boundaries, the fields are connected between the layers.

Let us start with the general solutions of the TM wave, which have been obtained earlier. From Eqs. (9.9) and (9.21), the fields in the ith layer are

$$H_y(x) = A \cos K_i x + B \sin K_i x \tag{9.74}$$

$$E_z(x) = -AZ_i \sin K_i x + BZ_i \cos K_i x \tag{9.75}$$

where

$$Z_i = \frac{jK_i}{\omega \epsilon_0 \epsilon_{ri}} \tag{9.76}$$

or

$$\begin{aligned} Z_i &= \frac{jK_i}{\omega \epsilon_0 n_i^2} \\ K_i &= \sqrt{(n_i k)^2 - \beta^2} \end{aligned} \tag{9.77}$$

We want to choose the values of A and B such that the **H** and **E** fields become the given values of $H(x')$ and $E(x')$ at $x = x'$:

$$H_y(x') = A \cos K_i x' + B \sin K_i x' \tag{9.78}$$

$$E_z(x') = -AZ_i \sin K_i x' + BZ_i \cos K_i x' \tag{9.79}$$

Both x and x' belong to the same ith layer as indicated in Fig. 9.8.

$$x_{i-1} \leq (x', x) \leq x_i$$

where x_{i-1} is the value of x at the lower boundary of the ith layer and x_i is the value of x at the upper boundary of the ith layer. The values of A and B are obtained by

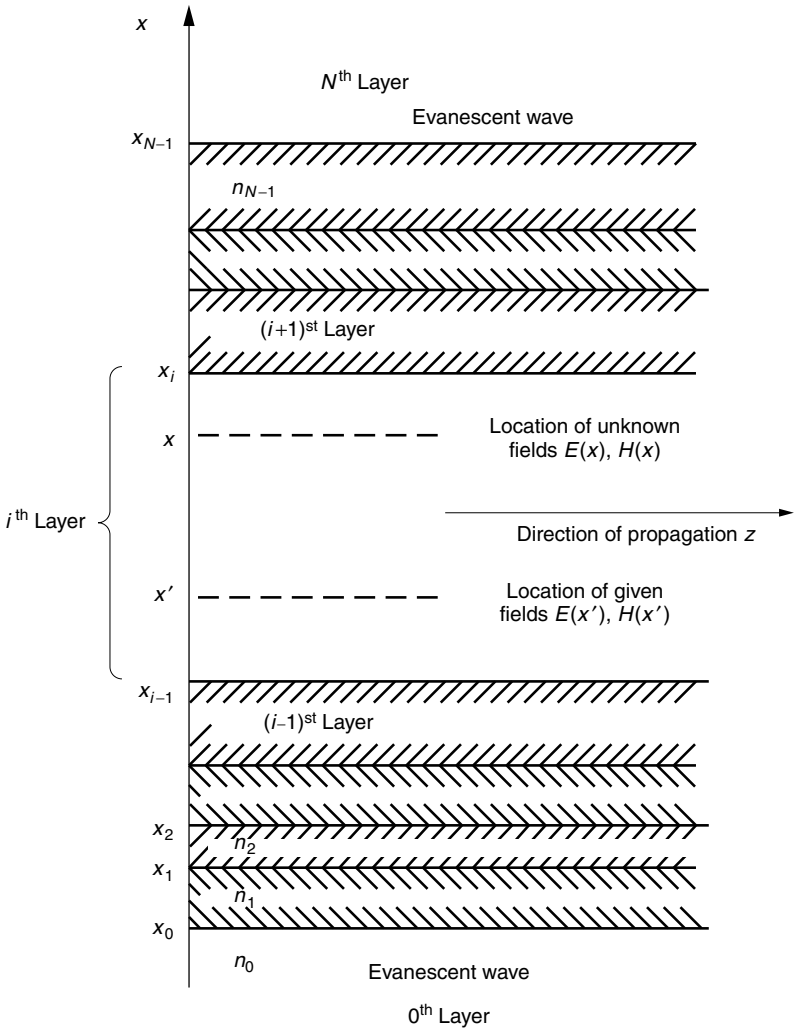


Figure 9.8 Geometry of a layered medium.

solving the simultaneous Eqs. (9.78) and (9.79). The solutions for A and B are put back into Eqs. (9.74) and (9.75) to reach

$$\begin{bmatrix} H_y(x) \\ E_z(x) \end{bmatrix} = \begin{bmatrix} \cos K_i(x-x') & \dfrac{1}{Z_i}\sin K_i(x-x') \\ -Z_i \sin K_i(x-x') & \cos K_i(x-x') \end{bmatrix} \begin{bmatrix} H_y(x') \\ E_z(x') \end{bmatrix} \tag{9.80}$$

Equation (9.80) is the transmission matrix that relates the fields at $x = x'$ with those at $x = x$. The x' and x can be set at any point as long as they belong to the same layer. Equation (9.80) is the only type of matrix that is needed for the transmission matrix method.

The transmission matrix method will first be illustrated using the example of a three-layer guide with the geometry shown in Fig. 9.9. The core-cladding lower boundary

was placed at $x = 0$ and the upper boundary at $x = 2d$. All other parameters are the same as before. Only one transmission matrix is needed for this geometry:

$$\begin{bmatrix} H_y(2d) \\ E_z(2d) \end{bmatrix} = \begin{bmatrix} \cos 2Kd & \dfrac{1}{Z_1}\sin 2Kd \\ -Z_1 \sin 2Kd & \cos 2Kd \end{bmatrix} \begin{bmatrix} H_y(0) \\ E_z(0) \end{bmatrix} \tag{9.81}$$

The geometry of the layers alone can never determine the field inside the medium. It is only after the fields on the boundary are specified that the field inside the medium is determined. The specification of the field can be the field H_y or E_z itself, as in the case of a given incident field to the boundary; or the field specification can be an unbounded traveling wave, a ratio of reflected to incident waves (standing wave), or an evanescent wave. These are called boundary fields. This section deals with guides whose boundary fields are evanescent waves.

Now let us deal with the case of a guide for which the fields in the bottom layer as well as in the top layer are unbounded waves. The unbounded field in the bottom layer is the evanescent wave $H_{y0}(x)$, which is expressed by the bottom equation of Eq. (9.11). The bottom equation of Eq. (9.11) is inserted into Eq. (9.21) to obtain

$$E_{z0}(x) = \frac{j}{\omega\epsilon_r\epsilon_0}\frac{\partial}{\partial x}De^{\gamma_0 x} = Z_0 H_{y0}(x) \tag{9.82}$$

where

$$Z_0 = \frac{j\gamma_0}{\omega\epsilon_0 n_0^2} \tag{9.83}$$

The unbounded field $H_{y2}(x)$ in the top layer is also an evanescent wave and is expressed by the top equation of Eq. (9.11). Inserting the top equation of Eq. (9.11) into Eq. (9.21) gives

$$E_{z2}(x) = \frac{j}{\omega\epsilon_0\epsilon_2}\frac{\partial}{\partial x}Ce^{-\gamma_2 x} = -Z_2 H_{y2}(x) \tag{9.84}$$

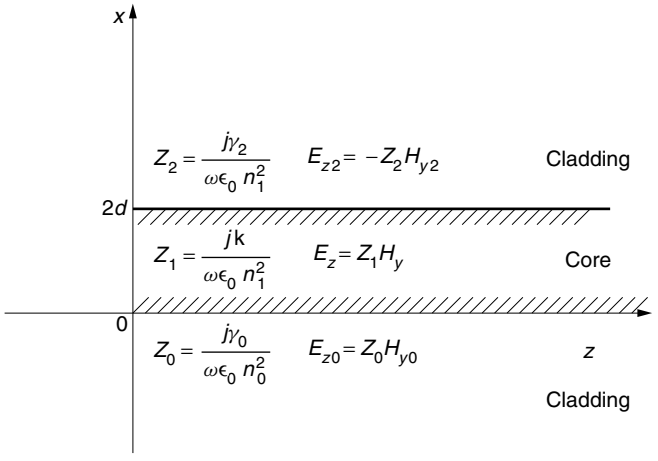


Figure 9.9 Geometry of the three-layer slab guide.

where

$$Z_2 = \frac{j\gamma_2}{\omega\epsilon_0 n_2^2} \tag{9.85}$$

The minus sign on the right-hand side of Eq. (9.84) should be noted. For simplicity, the second subscripts will be dropped as

$$\begin{aligned} H_{y0}(0) &= H_y(0) \qquad E_{z0}(0) = E_z(0) \\ H_{y2}(2d) &= H_y(2d) \quad E_{z2}(2d) = E_z(2d) \end{aligned} \tag{9.86}$$

The $E_z(2d)$ and $E_z(0)$ in Eq. (9.81) are rewritten in terms of $H_y(2d)$ and $H_y(0)$, applying the boundary conditions to Eqs. (9.82) and (9.84) and rearranged as

$$\begin{bmatrix} \cos 2Kd + \dfrac{Z_0}{Z_1}\sin 2Kd & -1 \\ Z_0 \cos 2Kd - Z_1 \sin 2Kd & Z_2 \end{bmatrix} \begin{bmatrix} H_y(0) \\ H_y(2d) \end{bmatrix} = 0 \tag{9.87}$$

In order that $H_y(0)$ and $H_y(2d)$ have nonzero solutions, the determinant has to be zero. The determinant is

$$(Z_0 + Z_2)\cos 2Kd - \left(Z_1 - \frac{Z_0 Z_2}{Z_1}\right)\sin 2Kd = 0 \tag{9.88}$$

or

$$\frac{Z_0/Z_1 + Z_2/Z_1}{1 - Z_0 Z_2/Z_1^2} = \tan 2Kd \tag{9.89}$$

If the guide is symmetric and $Z_2 = Z_0$, then

$$\frac{2(Z_0/Z_1)}{1 - (Z_0/Z_1)^2} = \tan 2Kd \tag{9.90}$$

The trigonometric identity

$$\tan 2Kd = \frac{2\tan Kd}{1 - \tan^2 Kd}$$

is inserted into Eq. (9.90). After rearrangement, Eq. (9.90) becomes

$$\left(\frac{\gamma}{n^2 K} - \tan Kd\right)\left(\frac{\gamma}{n^2 K}\tan Kd + 1\right) = 0 \tag{9.91}$$

where Eqs. (9.76) and (9.83) were used to rewrite

$$Z_0/Z_1 = \frac{\gamma}{n^2 K}$$

Thus, the first factor of Eq. (9.91) is the characteristic equation for the even TM modes and the second, for the odd TM modes.

Next, this method will be extended to an $(N-1)$-layer guide in unbounded space. The continuity condition is used to connect the field of one layer to the next. As already

stated, in the two outermost layers, the wave is necessarily an evanescent wave, and the first and last layers are set aside for now. The matrix that applies to the ith layer in the geometry shown in Fig. 9.8 is, from Eq. (9.80),

$$\begin{bmatrix} H_y(x_i) \\ E_z(x_i) \end{bmatrix} = [T_i(x_i - x_{i-1})] \begin{bmatrix} H_y(x_{i-1}) \\ E_z(x_{i-1}) \end{bmatrix} \tag{9.92}$$

where

$$T_i(x_i - x_{i-1}) = \begin{bmatrix} \cos K_i(x_i - x_{i-1}) & \dfrac{1}{Z_i} \sin K_i(x_i - x_{i-1}) \\ -Z_i \sin K_i(x_i - x_{i-1}) & \cos K_i(x_i - x_{i-1}) \end{bmatrix} \tag{9.93}$$

In the event that a particular mode that has an evanescent wave in the ith layer is desired, the transmission matrix $T_i(x_i - x_{i-1})$ of the ith layer has to be derived in a similar manner as Eq. (9.93) was derived from Eqs. (9.74) and (9.75), starting from

$$H_y(x) = C \cosh \gamma_i x + D \sinh \gamma_i x \tag{9.94}$$

$$E_z(x) = Z_i C \sinh \gamma_i x + Z_i D \cosh \gamma_i x \tag{9.95}$$

where

$$\gamma_i = \sqrt{\beta^2 - (n_i k)^2} \tag{9.96}$$

$$Z_i = \frac{j\gamma_i}{\omega \epsilon_0 \epsilon_{ri}} \tag{9.97}$$

The values of C and D are chosen such that H and E become $H(x')$ and $E(x')$ at $x = x'$. These values of C and D are put back into Eqs. (9.94) and (9.95). The obtained matrix is

$$T_i^e(x_i - x_{i-1}) = \begin{bmatrix} \cosh \gamma_i(x_i - x_{i-1}) & \dfrac{1}{Z_i} \sinh \gamma_i(x_i - x_{i-1}) \\ Z_i \sinh \gamma_i(x_i - x_{i-1}) & \cosh \gamma_i(x_i - x_{i-1}) \end{bmatrix} \tag{9.98}$$

Compared to Eq. (9.93), the lower left element has a positive sign instead of a negative sign, and the sine and cosine functions are changed to hyperbolic functions.

Repeated use of Eq. (9.92) to connect the field starting from the first layer up the $(N-1)$st layer gives $[T_{N-1}][T_{N-2}] \cdots [T_1]$:

$$\begin{bmatrix} H_y(x_{N-1}) \\ E_z(x_{N-1}) \end{bmatrix} = [T_{N-1}][T_{n-2}] \cdots [T_i] \cdots [T_1] \begin{bmatrix} H_y(x_0) \\ E_z(x_0) \end{bmatrix} \tag{9.99}$$

The product of the T matrices is rewritten as

$$\begin{bmatrix} H_y(x_{N-1}) \\ E_z(x_{N-1}) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} H_y(x_0) \\ E_z(x_0) \end{bmatrix} \tag{9.100}$$

Now, the field at the top boundary of the zeroth layer is

$$E_z(x_0) = Z_0 H_y(x_0) \tag{9.101}$$

where

$$Z_0 = \frac{j\gamma_0}{\omega \epsilon_0 n_0^2} \tag{9.102}$$

and

$$\gamma_0 = \sqrt{\beta^2 - (n_0 k)^2} \tag{9.103}$$

The field at the lower boundary of the top free-space layer, which is the top boundary of the $(N-1)$st layer has a similar relationship as Eq. (9.84) and

$$E_z(x_{N-1}) = -Z_N H_y(x_{N-1}) \tag{9.104}$$

where

$$Z_N = \frac{j\gamma_N}{\omega\epsilon_0 n_N^2} \tag{9.105}$$

and

$$\gamma_N = \sqrt{\beta^2 - (n_N k)^2} \tag{9.106}$$

Note the difference in signs in Eqs. (9.101) and (9.104). Also note that the Nth layer is free space and Z_N given by Eq. (9.104) has to be used.

It may be added that, as explained in Section 2.7.1, the z-direction propagation constant β in the zeroth layer given by Eq. (9.103) is the same as that in the Nth layer given by Eq. (9.106) because the fields in both layers satisfy Eq. (9.1).

As a matter of fact, the fields in every layer satisfy Eq. (9.1) and have the same value of β. The phase match condition in the z direction is automatically satisfied across every layer boundary. The expressions for $E_z(x_{N-1})$ and $E_z(x_0)$ are rewritten in terms of $H_y(x_{N-1})$ and $H_y(x_0)$ using Eqs. (9.101) and (9.104) and are incorporated into Eq. (9.100):

$$\begin{bmatrix} A + Z_0 B & -1 \\ C + Z_0 D & Z_N \end{bmatrix} \begin{bmatrix} H_y(x_0) \\ H_y(x_{N-1}) \end{bmatrix} = 0 \tag{9.107}$$

In order that the solutions for $H_y(x_0)$ and $H(x_{N-1})$ are nonzero, the determinant of Eq. (9.107) has to be zero. Thus, the characteristic equation is finally

$$Z_N A + Z_0 Z_N B + C + Z_0 D = 0 \tag{9.108}$$

where the values of Z_0 and Z_N are available from Eqs. (9.102) and (9.105) and the matrix elements A, B, C, and D are obtained after calculating the product of the T's in Eq. (9.99). See the example of the asymmetric three-layer case in Example 9.6.

9.7.3 Transmission Matrix Method (Symmetric Guide)

If symmetry exists in the distribution of the indices of refraction, the computation becomes significantly simpler. Let us say, as shown in Fig. 9.10, symmetry exists with respect to $x = 0$ and the indices of refraction from the bottom to the top layer are arranged as $n_0, n_1, n_2, \ldots, n_{n-1}, n_{n-1}, \ldots, n_2, n_1, n_0$. The transmission matrix then becomes

$$[T_1][T_2]\cdots[T_{n-1}]\cdot[T_{n-1}]\cdots[T_2][T_1] \tag{9.109}$$

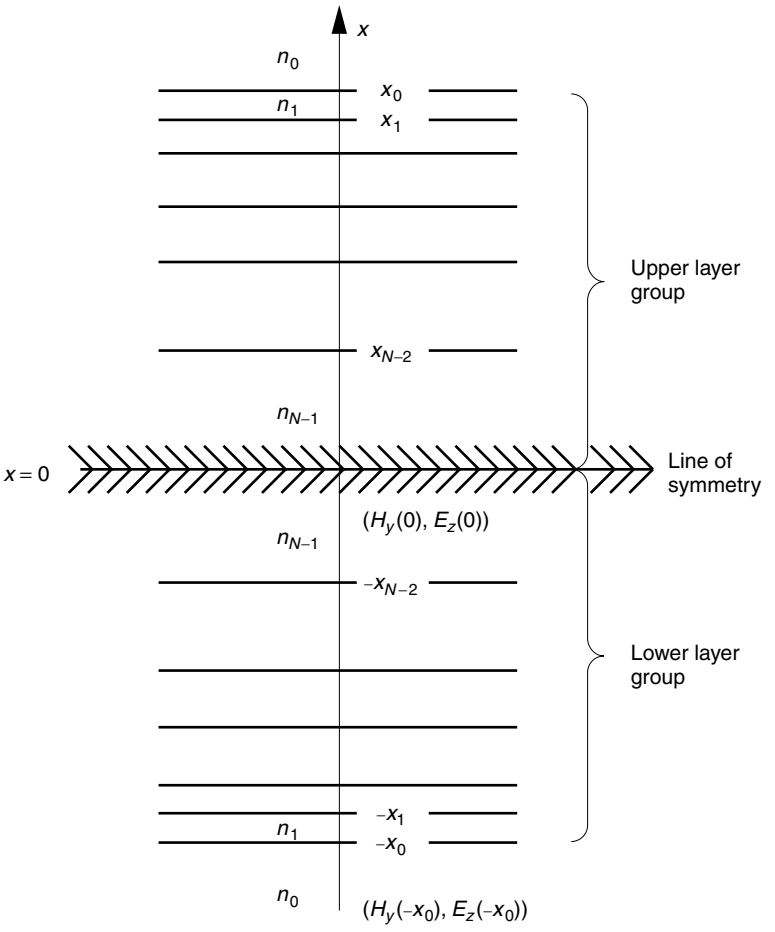


Figure 9.10 Geometry of layers with symmetry with respect to $x = 0$.

where the zeroth layer is considered unbounded and will be treated separately, as was done in Eq. (9.101). Equation (9.109) is now divided into two layer groups:

$$T_+ = [T_1][T_2]\cdots[T_{n-1}] \tag{9.110}$$

$$T_- = [T_{n-1}]\cdots[T_2][T_1] \tag{9.111}$$

Noting that with either Eq. (9.93) or (9.98) if

$$[T_1][T_2] = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

then

$$[T_2][T_1] = \begin{bmatrix} d & b \\ c & a \end{bmatrix}$$

because in both formulas the two diagonal elements in $[T_i]$ are identical.

Let

$$T_- = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{9.112}$$

then T_+, which consists of the same matrices but in reverse order in multiplication, becomes

$$T_+ = \begin{bmatrix} D & B \\ C & A \end{bmatrix} \tag{9.113}$$

Thus, the characteristic matrix

$$[T] = [T_+][T_-] \tag{9.114}$$

is obtained by the product of Eqs. (9.112) and (9.113) as

$$[T] = \begin{bmatrix} AD + BC & 2BD \\ 2AC & AD + BC \end{bmatrix} \tag{9.115}$$

The $[T]$ can be obtained by first calculating A, B, C, and D of T_- and then using them in Eq. (9.115). The characteristic matrix thus obtained involves only about one-half of the number of matrix multiplications.

The calculation of the modes in the symmetric structure can be simplified further by calculating the even and odd modes separately. Even and odd refers to H_y for the TM modes.

A special feature of the even modes is that not only is the field distribution in the upper layer group symmetric with that in the lower layer group with respect to $x = 0$, but also the derivative dH_y/dx has to vanish at $x = 0$ for a smooth connection.

$$\left.\frac{dH_y}{dx}\right|_{x=0} = 0 \quad \text{means} \quad E_z(0) = \left.\frac{j}{\omega\epsilon_0\epsilon_r}\frac{dH_y}{dx}\right|_{x=0} = 0 \tag{9.116}$$

As mentioned earlier, at the lowest boundary

$$E(-x_0) = Z_0 H_y(-x_0)$$

The field in the lower layer group is, from Eqs. (9.100) and (9.116),

$$\begin{bmatrix} H_y(0) \\ 0 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} H_y(-x_0) \\ Z_0 H_y(-x_0) \end{bmatrix} \tag{9.117}$$

Equation (9.117) can be rewritten as

$$\begin{bmatrix} A + Z_0 B & -1 \\ C + Z_0 D & 0 \end{bmatrix} \begin{bmatrix} H_y(-x_0) \\ H_y(0) \end{bmatrix} = 0 \tag{9.118}$$

For $H_y(-x_0)$ and $H_y(0)$ to exist, the determinant has to vanish and the characteristic equation for the even modes is given by

$$C + Z_0 D = 0 \tag{9.119}$$

Next, the characteristic equation for the odd TM modes is derived. A special feature of the odd modes is that not only is the field in the upper layer group the negative of the corresponding field in the lower layer group, but also the field itself has to vanish at $x = 0$ and $H_y(0) = 0$ so as to connect at $x = 0$.

$$\begin{bmatrix} 0 \\ E_z(0) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} H_y(-x_0) \\ Z_0 H_y(-x_0) \end{bmatrix} \tag{9.120}$$

Equation (9.120) can be rewritten as

$$\begin{bmatrix} A + Z_0 B & 0 \\ C + Z_0 D & -1 \end{bmatrix} \begin{bmatrix} H(-x_0) \\ E_z(0) \end{bmatrix} = 0 \tag{9.121}$$

For $H(-x_0)$ and $E_z(0)$ to exist, the determinant has to vanish and the characteristic equation for the odd modes is given by

$$A + Z_0 B = 0 \tag{9.122}$$

Example 9.5 Derive the characteristic equation of a planar type W guide (named after the shape of the index of refraction). The geometry and the indices of refraction are shown in Fig. 9.11.

Solution Calculation will take an advantage of the symmetry in the guide structure. Since $n_1 > n_2$, the guided waves can be in the center region. No total internal reflection exists at $x = \pm(d_1 + d_2)$ and this boundary alone cannot be used for confinement. The guided wave exists in $d_1 > x > -d_1$ only when the waves in $d_1 < |x| < d_1 + d_2$ are evanescent waves.

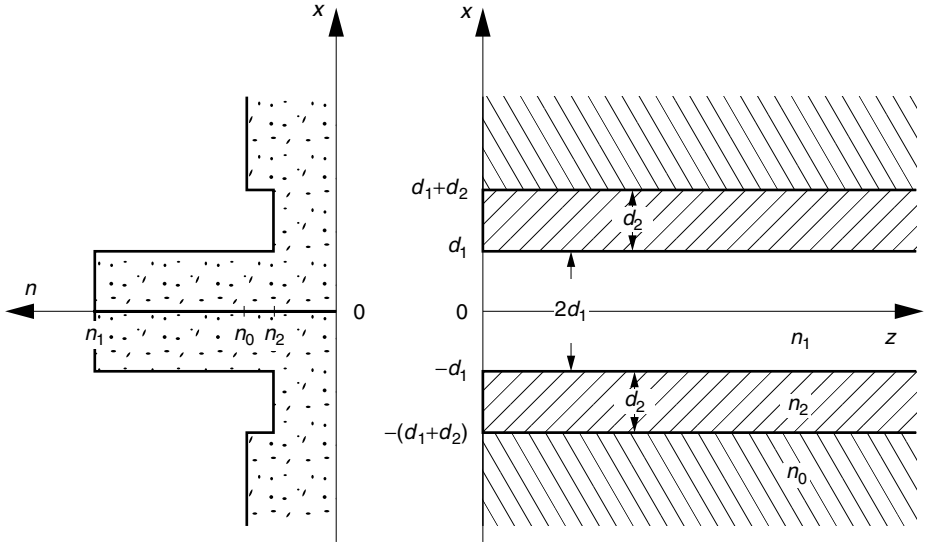


Figure 9.11 Geometry of the planar W guide.

In the region $-(d_1 + d_2) < x < -d_1$, from Eqs. (9.97) and (9.98), T_1 is

$$T_1 = \begin{bmatrix} \cosh \gamma_2 d_2 & \dfrac{1}{Z_2} \sinh \gamma_2 d_2 \\ Z_2 \sinh \gamma_2 d_2 & \cosh \gamma_2 d_2 \end{bmatrix} \tag{9.123}$$

with

$$Z_2 = \frac{j\gamma_2}{\omega \epsilon_0 n_2^2} \tag{9.124}$$

In the region $-d_1 < x < 0$, from Eqs. (9.76) and (9.93), T_2 is

$$T_2 = \begin{bmatrix} \cos K_1 d_1 & \dfrac{1}{Z_1} \sin K_1 d_1 \\ -Z_1 \sin K_1 d_1 & \cos K_1 d_1 \end{bmatrix} \tag{9.125}$$

with

$$Z_1 = \frac{jK_1}{\omega \epsilon_0 n_1^2} \tag{9.126}$$

Next, from T_2 and T_1, the matrix $T_- = [T_2][T_1]$ is calculated and then the matrix elements in T_- are used to find the characteristic equations by Eqs. (9.119) and (9.122).

For the even modes

$$\begin{aligned} &- Z_1 \tan K_1 d_1 + Z_2 \tanh \gamma_2 d_2 \\ &+ Z_0 \left(1 - \frac{Z_1}{Z_2} \tan K_1 d_1 \tanh \gamma_2 d_2 \right) = 0 \end{aligned} \tag{9.127}$$

which can be rewritten as

$$Z_0 = Z_1 \left(\frac{1 - A/B}{1 - AB} \right) \tan K_1 d_1 \tag{9.128}$$

where

$$A = \tanh \gamma_2 d_2 \tag{9.129}$$

$$B = \frac{Z_1}{Z_2} \tan K_1 d_1 \tag{9.130}$$

Equation (9.128) can be rewritten using Eqs. (9.102), (9.124), and (9.126). Thus, the characteristic equation for the even TM modes is

$$\gamma_0 d_1 = \left(\frac{n_0}{n_1} \right)^2 K_1 d_1 \left(\frac{1 - A/B}{1 - AB} \right) \tan K_1 d_1 \tag{9.131}$$

$$B = \left(\frac{n_2}{n_1} \right)^2 \frac{K_1}{\gamma_2} \tan K_1 d_1 \tag{9.132}$$

The result can be verified by setting either $d_2 \to 0$ or $n_2 \to n_0$. If $d_2 \to 0$, then $A = 0$ and Eq. (9.131) immediately reduces to Eq. (9.72). If $n_2 \to n_0$ and hence $\gamma_2 \to \gamma_0$, then B in Eq. (9.132) approaches unity because Eq. (9.72) can be approximately used in Eq. (9.132), and Eq. (9.131) approaches Eq. (9.72).

The procedure for calculating β for given physical parameters will be described. From Eqs. (9.77) and (9.96), the attentuation and propagation constants are

$$K_1 = \sqrt{(n_1 k)^2 - \beta^2} \tag{9.133}$$

$$\gamma_0 = \sqrt{\beta^2 - (n_0 k)^2} \tag{9.134}$$

$$\gamma_2 = \sqrt{\beta^2 - (n_2 k)^2} \tag{9.135}$$

From Eqs. (9.133) and (9.134), V_0^2 is expressed as

$$V_0^2 = (K_1 d_1)^2 + (\gamma_0 d_1)^2 \tag{9.136}$$

where

$$V_0 = k d_1 \sqrt{n_1^2 - n_0^2} \tag{9.137}$$

Similarly, from Eqs. (9.133) and (9.135), V_2^2 is expressed as

$$V_2^2 = (K_1 d_1)^2 + (\gamma_2 d_1)^2 \tag{9.138}$$

where

$$V_2 = k d_1 \sqrt{n_1^2 - n_2^2} \tag{9.139}$$

The procedures to find β are as follows:

1. Calculate both V_0 and V_2 from Eqs. (9.137) and (9.139).
2. Use $\gamma_2 d_1 = \sqrt{V_2^2 - (K_1 d_1)^2}$ to calculate $\gamma_2 d_1$ for a given $K_1 d_1$.
3. Insert γ_2 and the given value of $K_1 d_1$ into Eqs. (9.129) and (9.132) to find A and B and hence the corresponding value of $\gamma_0 d_1$ for the given $K_1 d_1$ from Eq. (9.131). Repeat the same with the other values of $K_1 d_1$ to complete such curves as shown in Fig. 9.12.
4. The intersections between the above curves and the circle of Eq. (9.136) are the desired solutions.
5. From Eq. (9.133), β is calculated for the obtained value of K_1.

As shown in Fig. 9.12, the value of $K_1 d_1$ for the W guide is shifted from that of the three-layer symmetric guide shown in Fig. 9.2. One of the determining factors of the distortion of a light pulse during the transmission in the guide is $d\beta^2/d\omega^2$ as will be detailed in Chapter 11. The values of n_2 and d_2 are manipulated to make a distortion-free line out of the W guide.

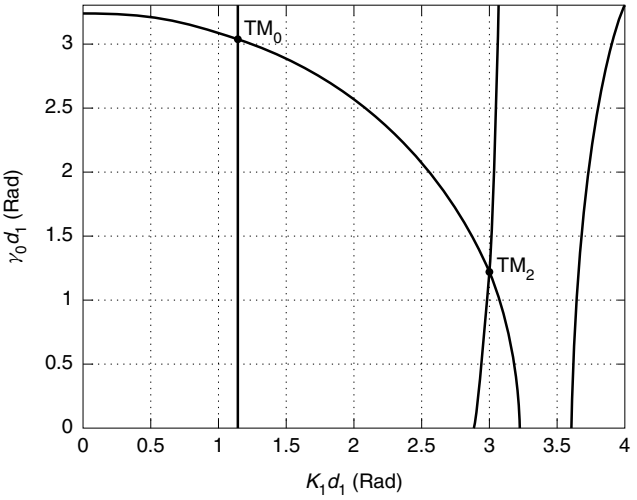


Figure 9.12 Graphical solution of the planar W guide for the even TM modes. The physical parameters are $\lambda = 1.3$ μm, $d_1 = 0.9$ μm, $d_2 = 1.0$ μm, $n_0 = 3.44$, $n_1 = 3.5$, and $n_2 = 3.42$. (Calculation courtesy of H. S. Loka.)

The characteristic equation for the odd TM modes can be obtained in a similar manner and the result is

$$\gamma_0 d_1 = -\left(\frac{n_0}{n_1}\right)^2 K_1 d_1 \cot K_1 d \cdot \left(\frac{1 + AB'}{1 + A/B'}\right) \tag{9.140}$$

where A is given by Eq. (9.129) and

$$B' = \frac{z_2}{z_1} \tan K_1 d_1$$

The result can be verified by setting either $d_2 \to 0$ or $n_2 \to n_0$, as in the case of the even TM modes. □

9.7.4 Modified Ray Model Method

The mode is essentially a quantization of the propagation constant, or a quantization of the angle of propagation of the component plane waves. At angles other than the quantized angles, no light can propagate in the guide. In Section 9.4, it was shown that a mode in the guide could be thought of as the standing-wave pattern of two component waves, one propagating in a slightly downward direction and the other in a slightly upward direction. Figure 9.13 shows the path of the component rays inside the optical guide. The hatched bell-shaped curve signifies the transverse distribution of the light. The field at $x = x_1$ for the downward component wave is considered first. As illustrated in Fig. 9.13, the field at any given point along $x = x_1$ is the sum of numerous contributing fields due to the transverse spread of the beam. If the beam were a plane wave, the transverse spread would be wider, and the interference would be even bigger. If the phases of the contributing fields are random, the resultant travelling wave field becomes null. If, however, the phase shift associated with one round trip is an integral multiple of 2π radians, all the contributing fields interfere constructively, and the field

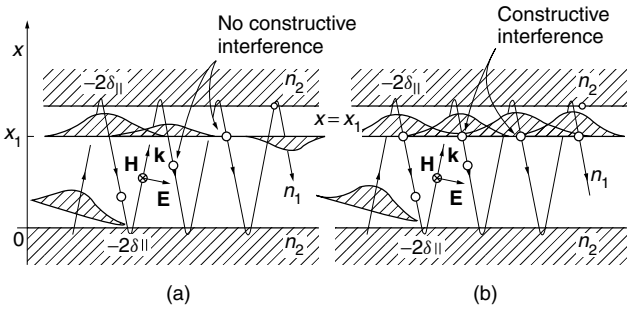


Figure 9.13 Explanation of the quantization in the x direction. (a) Random phase at $x = x_1$. (b) In phase at $x = x_1$. $4Kd - 4\delta_{\|} = 2\,\mu\pi$.

builds up. The phase shift associated with one round trip includes the phase shifts due to the total internal reflections at the top and bottom boundaries as well as the phase shift due to the round-trip distance inside the guide.

The same is true with the upward component wave. The field at any given point is the sum of many contributing fields, which must interfere constructively for the field to build up, and the same condition applies that the round-trip phase shift be an integral multiple of 2π.

The upward and downward component waves propagating in opposite directions again interfere and generate a standing-wave pattern, which is the mode pattern of the guide.

The amount of the phase shift associated with the reflection from the boundary depends on the direction of the polarization of the component plane wave. As shown in Fig. 9.13, if the TM wave is used, the direction of the **H** field is perpendicular to the plane of propagation and, hence, the direction of the polarization of the **E** field is parallel to the plane of propagation (in the plane of the page). The total round-trip phase shift for a $2d$-thick guide for constructive interaction is

$$4Kd - 4\delta_{\|} = 2\mu\pi \tag{9.141}$$

where μ is an integer and the phase shift due to the total internal reflections at each boundary is $-2\delta_{\|}$ as given by Eq. (2.86). Equation (9.141) is sometimes called the *dispersion equation*.

With Eq. (2.91), Eq. (9.141) becomes

$$\tan\left(Kd - \frac{\pi}{2}\mu\right) = \frac{\gamma}{n^2 K} \tag{9.142}$$

where

$$n = \frac{n_2}{n_1}$$

The integer μ is the mode number and separating it into even and odd modes gives

$$\tan Kd = \frac{\gamma}{n^2 K} \quad \text{for even } \mu \tag{9.143}$$

$$-\cot Kd = \frac{\gamma}{n^2 K} \quad \text{for odd } \mu \tag{9.144}$$

Thus, these results are identical to the results with the other three methods.

In the case of multilayer guides, the fields that are reflected from each interface must be taken into account in order to calculate the resultant phase shift, and calculation by this method becomes difficult.

9.8 ASYMMETRIC OPTICAL GUIDE

The three-layer optical guide with a different index of refraction for each layer is investigated in this section. Even though the analysis is slightly more complex, the asymmetric guide is of more practical importance than the symmetric guide. The geometry of the guide under consideration is shown in Fig. 9.14. The thickness of the film (core layer) is $2d$ and its index of refraction is n_1. The index of refraction of the substrate (cladding layer) is n_2. Very often, the top layer is air but sometimes it is a covering medium. The index of refraction of the top layer is n_0.

The characteristic equation for an asymmetric guide is derived by modifying the dispersion equation for the symmetric guide. The phase shift due to total internal reflection in Eq. (9.141) is separated into two parts, and with this modification, Eq. (9.141) becomes

$$4Kd - 2\delta_{\parallel}^{c} - 2\delta_{\parallel}^{s} = 2\mu\pi \tag{9.145}$$

where $-2\delta_{\parallel}^{c}$ is the phase shift due to the reflection at the air–film or cover–film interface and $-2\delta_{\parallel}^{s}$ is that of the film-substrate interface. From Eq. (2.91), the values of $\delta_{\parallel}^{c}$ and $\delta_{\parallel}^{s}$ are known and Eq. (9.145) becomes

$$(2Kd - \mu\pi) = \tan^{-1}\left(\frac{\gamma_0}{n_0'^2 K}\right) + \tan^{-1}\left(\frac{\gamma_2}{n_2'^2 K}\right) \tag{9.146}$$

where

$$n_0' = \frac{n_0}{n_1}, \quad n_2' = \frac{n_2}{n_1}$$

Equation (9.146) can be rewritten by recalling the identities

$$\tan(\theta_1 + \theta_2) = \frac{\tan\theta_1 + \tan\theta_2}{1 - \tan\theta_1 \tan\theta_2} \tag{9.147}$$

$$\tan(\theta - \mu\pi) = \tan\theta \tag{9.148}$$

With the help of these identities, the characteristic equation for the asymmetric guide is

$$\frac{K\left(\dfrac{\gamma_0}{n_0'^2} + \dfrac{\gamma_2}{n_2'^2}\right)}{K^2 - \dfrac{\gamma_0\gamma_2}{n_0'^2 n_2'^2}} = \tan 2Kd \tag{9.149}$$

Equation (9.149) becomes identical with Eq. (9.89) obtained by the transmission matrix method if the relationships Eqs. (9.102), (9.124), and (9.126) are used.

There are three unknowns — K, γ_0, and γ_2 — in Eq. (9.149), and in order to find a solution, more formulas are needed. The propagation constant relationships across a

boundary were discussed in Section 2.7.1. Because of phase matching, the propagation constant β in the z direction is the same for every layer and

$$K = \sqrt{(n_1 k)^2 - \beta^2} \tag{9.150}$$

$$\gamma_0 = \sqrt{\beta^2 - (n_0 k)^2} \tag{9.151}$$

$$\gamma_2 = \sqrt{\beta^2 - (n_2 k)^2} \tag{9.152}$$

Now, there are four unknowns — K, γ_0, γ_2, and β; and there are four equations — Eqs. (9.149) to (9.152). By substituting Eqs. (9.150) through (9.152) into Eq. (9.149), an equation in terms of β is obtained. Unfortunately, this equation is a transcendental equation and no closed-form solutions are possible, but the solutions can be obtained by numerical methods. The solutions for a specific optical guide are shown in the next example.

Example 9.6 An asymmetric optical guide such as the one shown in Fig. 9.14 is fabricated by depositing a glass film with a higher index of refraction over a glass substrate. The film thickness is $2d$. The indices of refraction are $n_0 = 1$, $n_1 = 1.55$, and $n_2 = 1.54$ and the TM modes are excited. The wavelength is λ. The positive x direction is taken downward.

Without any elaborate numerical calculations, make intelligent guesses at the following:

(a) The total number of possible modes.
(b) The conditions for a single mode guide.
(c) The field distributions for the first few mode orders.
(d) The correlation of the region of β with the field pattern and directions of propagation of the component plane waves.

Solution The change in the index of refraction at the air–film interface is much greater than that at the film–substrate interface. Therefore, if the condition of total

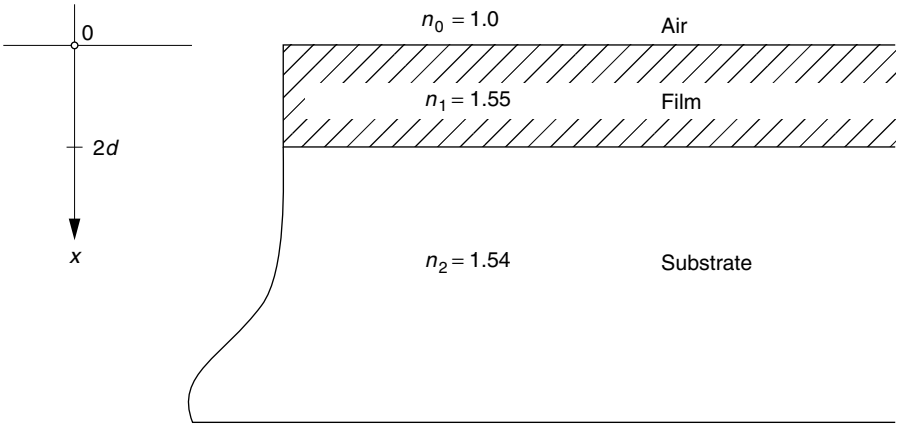


Figure 9.14 Geometry of an asymmetric slab optical guide.

internal reflection for the film–substrate is satisfied, that for the air–film is automatically satisfied. Hence, the results are primarily determined by the film–substrate boundary.

First, the region of β is investigated. Using Eqs. (9.150) to (9.152) and the given physical parameters, the curves for K, γ_0, and γ_1 are plotted as a function of β in Fig. 9.15a. The upper limit of β is determined by the index of refraction of the film $n_1 k$ from Eq. (9.150).

$$n_1 k > \beta \tag{9.153}$$

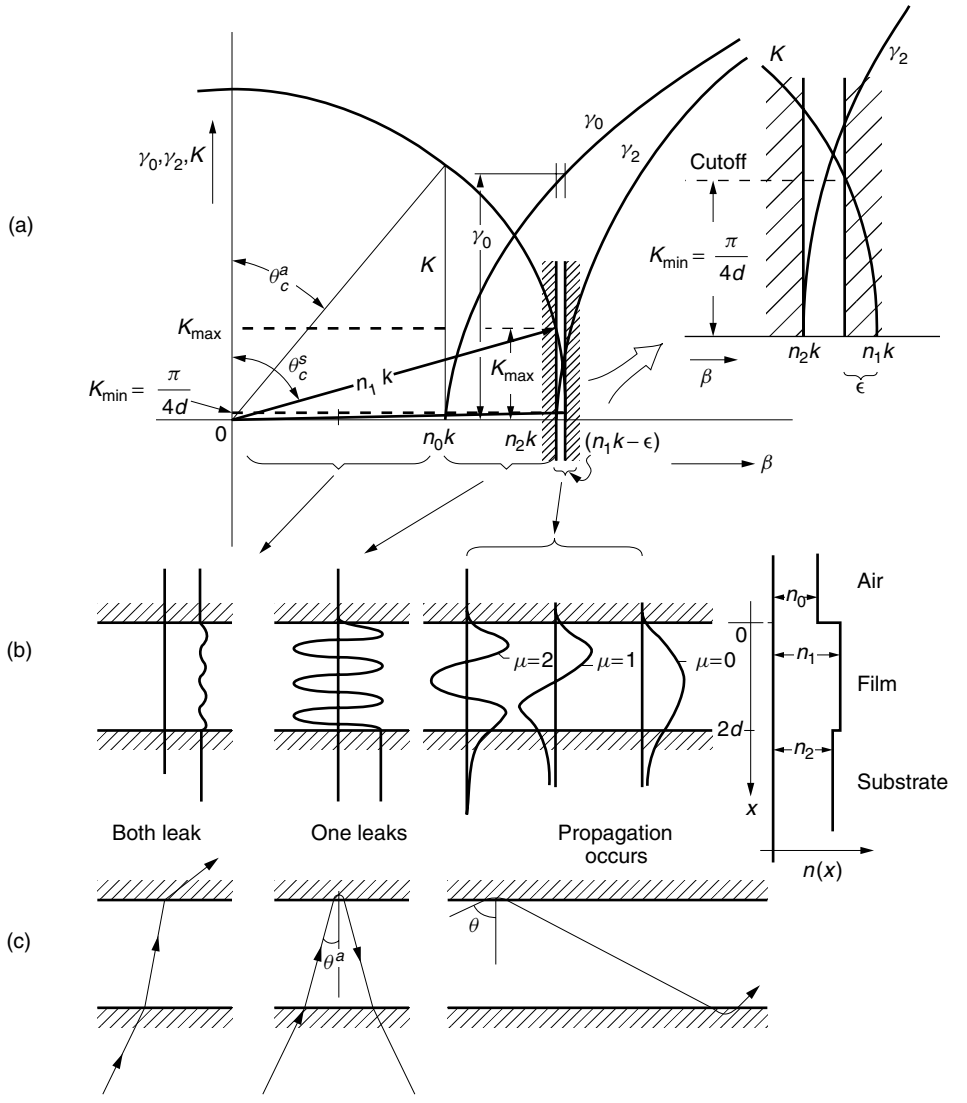


Figure 9.15 Combined chart of an asymmetric optical guide. (a) $K-\beta$ diagram. (b) Amplitude distribution. (c) Ray path.

The lower limit of β is determined by the condition of the cutoffs (total internal reflection) at the air–film and the film–substrate boundaries. From Eq. (9.151), the cutoff at the air–film boundary is

$$\beta > n_0 k \tag{9.154}$$

Whereas from Eq. (9.152), the cutoff at the film–substrate boundary is

$$\beta > n_2 k \tag{9.155}$$

Since n_2 is larger than n_0, if Eq. (9.155) is satisfied, Eq. (9.154) is automatically satisfied. Thus, for transmission, β must satisfy the condition

$$n_1 k > \beta > n_2 k \tag{9.156}$$

This is a narrow region fenced in by the hatched sections in Fig. 9.15a.

Next, an approximate expression for the characteristic equation is found. The difference in the refractive indices $n_1 - n_0$ is much larger than that of $n_2 - n_1$ and the following approximation is possible. Since the range of β where mode propagation takes place is so narrow, the value of γ_0 can be assumed constant over this range (see the curve of γ_0 in Fig. 9.15a). Moreover, γ_0 in this region is much larger than $K_{\max}$. Hence, the value of $\delta_{\parallel}^{c} = \tan^{-1}(\gamma_0 / n_0'^2 K)$ ranges from 70° to 90° depending on the value of K. For simplicity, $\delta_{\parallel}^{c} \doteqdot \pi/2$ is assumed for the time being. With this assumption, Eq. (9.146) is simplified:

$$-\cot 2Kd = \frac{\gamma_2}{n_2'^2 K} \tag{9.157}$$

It is interesting to note that Eq. (9.157) is similar to the characteristic equation [Eq. (9.31)] for the odd TM modes inside the symmetric guide, but with a factor of 2.

Another relationship has to be found to solve for K and γ_2. From Eqs. (9.150) and (9.152), we have

$$(\gamma_2 d)^2 + (Kd)^2 = V_2^2 \tag{9.158}$$

where

$$V_2 = kd\sqrt{n_1^2 - n_2^2} \tag{9.159}$$

Equations (9.157) and (9.158) are plotted in Fig. 9.16 to find the solutions. Now, we are ready to answer the questions.

(a) From Fig. 9.16, starting with $V_2 = \pi/4$, new modes are generated at an interval of $\pi/2$ radians. Modes up to the μth mode are excited if

$$\frac{\pi}{4} < V_2 < \frac{\pi}{4} + \frac{\pi}{2}(\mu + 1) \tag{9.160}$$

If μ is the largest integer that satisfies Eq. (9.160), then the number of possible modes is $\mu + 1$.

(b) The single-mode condition is again, from Fig. 9.16,

$$\frac{\pi}{4} < V_2 < \frac{3\pi}{4}$$

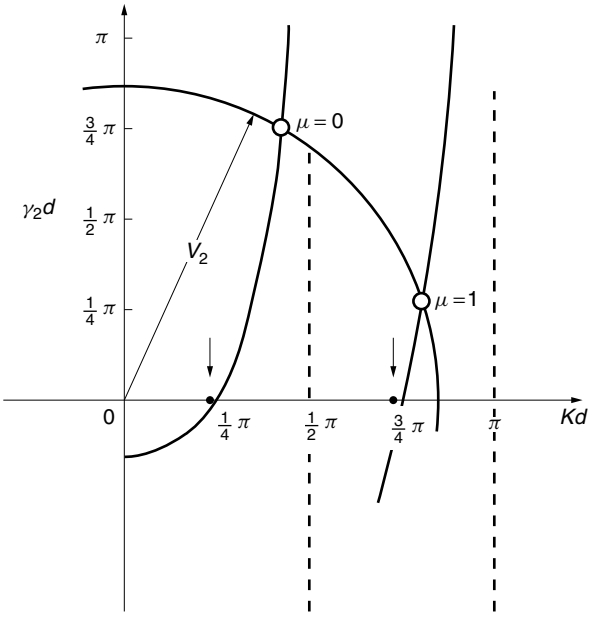


Figure 9.16 Graphical solutions for the asymmetrical guide. The exact locations of the intersections determined by the numerical methods are slightly to the left, as indicated by the arrows.

or the thickness $2d$ of the film has to satisfy

$$\frac{0.25\lambda}{\sqrt{n_1^2 - n_2^2}} < 2d < \frac{0.75\lambda}{\sqrt{n_1^2 - n_2^2}} \tag{9.161}$$

An important feature of the asymmetric guide is that, unlike the symmetric guide, the lowest order mode does have a cutoff and this cutoff occurs near $V_2 = \pi/4$.

(c) The distribution of the H_y field in the core will be found using Eq. (9.74). In the air layer, only the evanescent wave exists. Inserting Eqs. (9.74) and (9.75) into Eqs. (9.82) and (9.86), with $x = 0$, gives

$$\frac{B}{A} = \frac{Z_0}{Z_1}$$

The values of Z_0 and Z_1 are given by Eqs. (9.102) and (9.126) and

$$\frac{B}{A} = \frac{\gamma_0}{n_0'^2 K} \tag{9.162}$$

As we already found from the curves in Fig. 9.15, the right-hand side of Eq. (9.162) is very large. Hence, B is much larger than A and the distribution function of Eq. (9.74) can be approximated as

$$H_y = B \sin Kx \tag{9.163}$$

The value of K for the zero-order mode is, from Fig. 9.16,

$$\frac{\pi}{4d} < K < \frac{\pi}{2d}$$

The distribution of H_y with $\mu = 0$ is shown in Fig. 9.15b together with higher order modes.

(d) In the region

$$0 < \beta < n_0 k \tag{9.164}$$

total internal reflection does not take place at either interface and the light ray penetrates these boundaries according to the usual laws of reflection and refraction at boundaries.

We next examine the region

$$n_0 k < \beta < n_2 k \tag{9.165}$$

When β satisfies this condition, total internal reflection does not take place at the film–substrate interface, but the light is totally internally reflected at the air–film interface.

In the region

$$n_2 k < \beta < n_1 k - \epsilon \tag{9.166}$$

guided modes exist. The small number ϵ (shown in the expanded graph on the right side of Fig. 9.15a) is introduced to account for the fact that K in $\beta = \sqrt{(n_1 k)^2 - K^2}$ does not become $n_1 k$ because the lowest mode is cut off at $Kd = \pi/4$ and not at $Kd = 0$.

Ray paths of these regions are summarized in Fig. 9.15c. □

9.9 COUPLED GUIDES

Coupling between optical guides can be treated as a five-layer medium problem. Such a five-layer medium consists of two guiding layers with higher index of refraction, spaced by a center layer with lower index of refraction.

9.9.1 Characteristic Equations of the Coupled Slab Guide

Now, let us start with the calculation of the propagation constants of the coupled slab guide. The refractive index distribution is shown in Fig. 9.17. For simplicity, the distribution was chosen to be symmetric with respect to $x = 0$. The geometry is the same as the W guide explained in Example 9.5. The only difference is that this time there is a refractive index ditch in the center, and the d_2 layers in Fig. 9.11 become the guiding layers. The analysis procedures are quite similar and only a brief outline will be repeated here.

Because of the symmetry in the geometry, the product $[T_-]$ of the transmission matrices for the bottom half of the layers suffice. The layers in the bottom half are

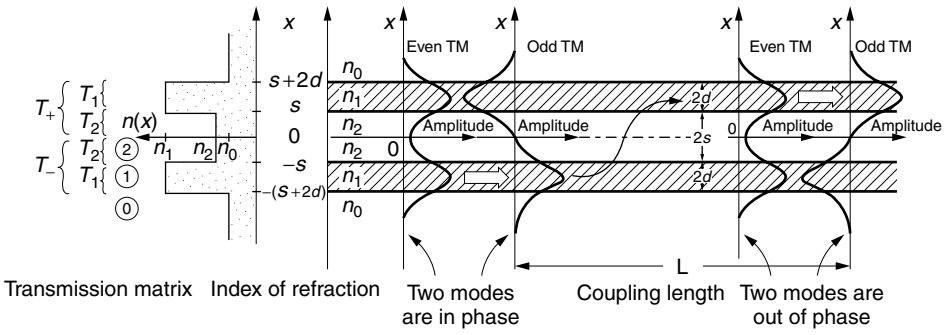


Figure 9.17 Slab optical coupler with symmetry in the refractive index distribution.

designated as

$$\begin{aligned} &\text{Region 2} \quad -s < x < 0 \\ &\text{Region 1} \quad -(s+2d) < x < -s \\ &\text{Region 0} \quad x < -(s+2d) \end{aligned} \tag{9.167}$$

The $[T_-]$ transfers

$$\begin{bmatrix} H_y(0) \\ E_z(0) \end{bmatrix} = [T_-] \begin{bmatrix} H_y(-(s+2d)) \\ E_z(-(s+2d)) \end{bmatrix} \tag{9.168}$$

where

$$[T_-] = [T_2][T_1] \tag{9.169}$$

and where $[T_1]$ and $[T_2]$ are transmission matrices for regions 1 and 2. For the mode that has the evanescent wave in region 2 and guided wave in region 1, $[T_-]$ becomes

$$T_- = \begin{bmatrix} \cosh \gamma_2 s & \dfrac{1}{Z_2} \sinh \gamma_2 s \\ Z_2 \sinh \gamma_2 s & \cosh \gamma_2 s \end{bmatrix} \begin{bmatrix} \cos 2K_1 d & \dfrac{1}{Z_1} \sin 2K_1 d \\ -Z_1 \sin 2K_1 d & \cos 2K_1 d \end{bmatrix} \tag{9.170}$$

If we put $[T_-]$ as

$$[T_-] = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{9.171}$$

then, as explained in Section 9.7.3, the characteristic equations for the even and odd modes are

$$C + Z_0 D = 0 \quad \text{even mode} \tag{9.172}$$

$$A + Z_0 B = 0 \quad \text{odd mode} \tag{9.173}$$

where

$$Z_0 = \frac{j\gamma_0}{\omega\epsilon_0 n_0^2}$$
$$Z_1 = \frac{jK_1}{\omega\epsilon_0 n_1^2} \tag{9.174}$$
$$Z_2 = \frac{j\gamma_2}{\omega\epsilon_0 n_2^2}$$

From Eqs. (9.170) to (9.172), the characteristic equation for the even modes becomes

$$\begin{aligned} &-\frac{\gamma_2}{n_2^2}\tanh\,\gamma_2 s + \frac{K_1}{n_1^2}\tan 2K_1 d \\ &-\frac{\gamma_0}{n_0^2}\left[1 + \frac{\gamma_2}{K_1}\left(\frac{n_1}{n_2}\right)^2 \tanh\,\gamma_2 s \tan 2K_1 d\right] = 0 \end{aligned} \tag{9.175}$$

Similarly, the characteristic equation for the odd modes becomes

$$\begin{aligned} &1 - \frac{K_1}{\gamma_2}\left(\frac{n_2}{n_1}\right)^2 \tanh\,\gamma_2 s \tan 2K_1 d + \frac{\gamma_0}{K_1}\left(\frac{n_1}{n_0}\right)^2 \tan 2K_1 d \\ &+ \frac{\gamma_0}{\gamma_2}\left(\frac{n_2}{n_0}\right)^2 \tanh\,\gamma_2 s = 0 \end{aligned} \tag{9.176}$$

The equations for the attenuation and propagation constants are

$$\gamma_0^2 = \beta^2 - (n_0 k)^2 \tag{9.177}$$
$$K_1^2 = (n_1 k)^2 - \beta^2 \tag{9.178}$$
$$\gamma_2^2 = \beta^2 - (n_2 k)^2 \tag{9.179}$$

The equations for the normalized thicknesses are

$$V_0^2 = (K_1 d)^2 + (\gamma_0 d)^2 \tag{9.180}$$
$$V_2^2 = (K_1 d)^2 + (\gamma_2 d)^2 \tag{9.181}$$
$$V_0 = kd\sqrt{n_1^2 - n_0^2} \tag{9.182}$$
$$V_2 = kd\sqrt{n_1^2 - n_2^2} \tag{9.183}$$

Equations (9.175) and (9.176) can be rewritten in a form similar to Eq. (9.131). The curves of $\gamma_0 d$ versus $K_1 d$ of either Eq. (9.175) or (9.176) are made incorporating Eq. (9.181). The intersections of the characteristic equation with the circle of Eq. (9.180) are the solutions for the K_{1x} even and the K_{1x} odd modes.

From the solutions for the K_{1x} even and K_{1x} odd modes, β_e and β_0 are found as

$$\beta_e = \sqrt{(n_1 k)^2 - K_{1x\,\text{even}}^2} \tag{9.184}$$

$$\beta_0 = \sqrt{(n_1 k)^2 - K_{1x\,\text{odd}}^2} \tag{9.185}$$

An actual calculation will be given in Example 9.7.

9.9.2 Amplitude Distribution in the Coupled Slab Guide

Next, the amplitude distribution across the five layers will be calculated. Only the case of the lowest order mode is treated and the subscript indicating the mode order will be suppressed. The field in the center region is calculated first, and then the outer regions in the order of 2, 1, and lastly 0.

Let us begin with the field of the even mode. The boundary field of the even mode in the $x = 0$ plane is $E_z(0) = 0$ as explained earlier in Section 9.7.3. Note that the transmission matrix is always used from the bottom to the top in each region. Only the field in the bottom half group of layers will be considered.

Field in Region 2 The field inside region 2, which extends from $x = -s$ to 0, is, from Eq. (9.98),

$$\begin{bmatrix} H_{y2}(x) \\ E_{z2}(x) \end{bmatrix} = \begin{bmatrix} \cosh\,\gamma_2(x+s) & \dfrac{1}{Z_2}\sinh\,\gamma_2(x+s) \\ Z_2\,\sinh\,\gamma_2(x+s) & \cosh\,\gamma_2(x+s) \end{bmatrix} \begin{bmatrix} H_{y2}(-s) \\ E_{z2}(-s) \end{bmatrix} \tag{9.186}$$

where $0 > x > -s$. Since we are dealing with the even mode, the condition of $E_{z2}(0) = 0$ as in Eq. (9.116) is used to simplify Eq. (9.186). From the bottom row of Eq. (9.186) with $x = 0$,

$$E_{z2}(-s) = -Z_2 H_{y2}(-s)\,\tanh\,\gamma_2 s \tag{9.187}$$

Insertion of Eq. (9.187) into the top row of Eq. (9.186) gives

$$H_{y2}(x) = \frac{\cosh\,\gamma_2 x}{\cosh\,\gamma_2 s} H_{y2}(-s)$$

and

$$H_{y2}(0) = H(0) = \frac{1}{\cosh\,\gamma_2 s} H_{y2}(-s) \tag{9.188}$$

From Eqs. (9.187) through (9.189), we have

$$H_{y2}(x) = H(0)\,\cosh\,\gamma_2 x \tag{9.189}$$

Insertion of Eq. (9.187) into the bottom row of Eq. (9.186) and the use of Eq. (9.188) give

$$E_{z2}(x) = H(0) Z_2\,\sinh\,\gamma_2 x \tag{9.190}$$

Field in Region 1 The field in region 1 is given by Eq. (9.80):

$$\begin{bmatrix} H_{y1}(x) \\ E_{z1}(x) \end{bmatrix} = \begin{bmatrix} \cos K_1(x+s+2d) & \dfrac{1}{Z_1}\sin K_1(x+s+2d) \\ -Z_1 \sin K_1(x+s+2d) & \cos K_1(x+s+2d) \end{bmatrix} \begin{bmatrix} H_{y1}(-(s+2d)) \\ E_{z1}(-(s+2d)) \end{bmatrix} \tag{9.191}$$

where

$$-s < x < -(s+2d)$$

The fields at the top of region 1 have to match those of the bottom of Region 2:

$$\begin{aligned} H_{y1}(-s) &= H_{y2}(-s) = H(0)\cosh \gamma_2 s \\ E_{z1}(-s) &= E_{z2}(-s) = -H(0)Z_2 \sinh \gamma_2 s \end{aligned} \tag{9.192}$$

where use was made of Eqs. (9.189) and (9.190).

Removing $H(0)$ from Eq. (9.192), we have

$$E_{z1}(-s) = -Z_2 H_{y1}(-s) \tanh \gamma_2 s \tag{9.193}$$

Inserting the boundary condition into Eq. (9.191) with $x = -s$ gives

$$\begin{bmatrix} H_{y1}(-s) \\ -Z_2 H_{y1}(-s) \tanh \gamma_2 s \end{bmatrix} = \begin{bmatrix} \cos 2K_1 d & \dfrac{1}{Z_1}\sin 2K_1 d \\ -Z_1 \sin 2K_1 d & \cos 2K_1 d \end{bmatrix} \begin{bmatrix} H_{y1}(-(s+2d)) \\ E_{z1}(-(s+2d)) \end{bmatrix} \tag{9.194}$$

Equation (9.194) will be inverted so that

$$\begin{bmatrix} H_{y1}(-(s+2d)) \\ E_{z1}(-(s+2d)) \end{bmatrix}$$

is used in Eq. (9.191). Recall that the matrix inversion is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\Delta}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{9.195}$$

After inserting Eq. (9.192), Eq. (9.194) is inverted using Eq. (9.195) to obtain

$$\begin{bmatrix} H_{y1}(-(s+2d)) \\ E_{z1}(-(s+2d)) \end{bmatrix} = \begin{bmatrix} H(0)\cos 2K_1 d \cosh \gamma_2 s + H(0)\dfrac{Z_2}{Z_1}\sin 2K_1 d \sinh \gamma_2 s \\ H(0)Z_1 \sin 2K_1 d \cosh \gamma_2 s - H(0)Z_2 \cos 2K_1 d \sinh \gamma_2 s \end{bmatrix} \tag{9.196}$$

Inserting Eq. (9.196) into (9.191) gives

$$H_{y1}(x) = H(0)\left[\cos K_1(x+s) - \frac{Z_2}{Z_1}\sin K_1(x+s)\tanh \gamma_2 s\right]\cosh \gamma_2 s \tag{9.197}$$

$$E_{z1}(x) = -H(0)Z_1\left[\sin K_1(x+s) + \frac{Z_2}{Z_1}\cos K_1(x+s)\tanh \gamma_2 s\right]\cosh \gamma_2 s \tag{9.198}$$

Field in Region 0 In region 0, both $H_{y0}(x)$ and $E_{z0}(x)$ decay exponentially and the relationship between them is $E_{z0}(x) = Z_0 H_{y0}(x)$ as seen from Eq. (9.82). The fields at the top of region 0 have to be *smoothly* connected with $H_{y1}(-(s+2d))$ and $E_{z1}(-(s+2d))$ in region 1. The fields are

$$H_{y0}(x) = H(0)\left[\cos 2K_1 d + \frac{Z_2}{Z_1}\sin 2K_1 d \tanh \gamma_2 s\right] \cosh \gamma_2 s(e^{\gamma_0(x+s+2d)}) \tag{9.199}$$

$$E_{z0}(x) = H(0)\left[Z_1 \sin 2K_1 d - Z_2 \cos 2K_1 d \tanh \gamma_2 s\right] \cosh \gamma_2 s(e^{\gamma_0(x+s+2d)}) \tag{9.200}$$

Table 9.1 Even TM modes in the coupled slab guide

Even TM Mode Fields			
Region	x	$H_y(x)$	$E_z(x)$
0	$s+2d$	$Ce^{-\gamma_0(x-s-2d)}$	$-Z_0 Ce^{-\gamma_0(x-s-2d)}$
1	s	$A\cos K_1(x-s) + B\sin K_1(x-s)$	$Z_1[-A\sin K_1(x-s) + B\cos K_1(x-s)]$
2	0	$H(0)\cosh \gamma_2 x$	$Z_2 H(0)\sinh \gamma_2 x$
1	$-s$	$A\cos K_1(x+s) - B\sin K_1(x+s)$	$-Z_1[A\sin K_1(x+s) + B\cos K_1(x+s)]$
0	$-s-2d$	$Ce^{\gamma_0(x+s+2d)}$	$Z_0 Ce^{\gamma_0(x+s+2d)}$

$A = H(0)\cosh \gamma_2 s$

$B = H(0)\dfrac{Z_2}{Z_1}\sinh \gamma_2 s$

$C = H(0)[\cos 2K_1 d \cosh \gamma_2 s + \dfrac{Z_2}{Z_1}\sin 2K_1 d \sinh \gamma_2 s]$

Table 9.2 Odd TM modes in the coupled slab guide

Odd TM Mode Fields			
Region	x	$H_y(x)$	$E_z(x)$
0	$s+2d$	$-C'e^{-\gamma_0(x-s-2d)}$	$Z_0 C'e^{-\gamma_0(x-s-2d)}$
1	s	$A'\cos K_1(x-s) - B'\sin K_1(x-s)$	$Z_1[A'\sin K_1(x-s) - B'\cos K_1(x-s)]$
2	0	$-H(0)\sinh \gamma_2 x$	$-Z_2 H(0)\cosh \gamma_2 x$
1	$-s$	$-A'\cos K_1(x+s) - B'\sin K_1(x+s)$	$Z_1[-A'\sin K_1(x+s) - B'\cos K_1(x+s)]$
0	$-s-2d$	$C'e^{\gamma_0(x+s+2d)}$	$Z_0 C'e^{\gamma_0(x+s+2d)}$

$A' = H(0)\sinh \gamma_2 s$

$B' = H(0)\dfrac{Z_2}{Z_1}\cosh \gamma_2 s$

$C' = H(0)[\cosh 2K_1 d \sinh \gamma_2 s + \dfrac{Z_2}{Z_1}\sin 2K_1 d \cosh \gamma_2 s]$

The smoothness of the connection is enforced by the characteristic equation, Eq. (9.172), which was originally derived from the smooth connection of both H_y and its derivative. With this condition, the ratio between the terms in square brackets in Eqs. (9.199) and (9.200) is found to be Z_0 (see Problem 9.8).

The amplitude distributions of the even TM modes in the coupled slab guide are tabulated in Table 9.1. The same for the odd TM modes are tabulated in Table 9.2.

Example 9.7 Using a lithium niobate substrate, coupled slab guides were fabricated. The dimensions and refractive indices are indicated in Fig. 9.18.

(a) Find the propagation constants of the first even and odd modes.
(b) Find the expression for the fields H_y and E_z of the even and odd modes, of the lowest order with respect to x. Draw the curve for $H_y(x)$.

Solution

(a) A slightly different approach will be taken to calculate the numerical solutions. So far, the mode values have been obtained from the intersections between the characteristic equation and the circle of the normalized thickness. The intersections provided the values of both K_d and γ_d at the same time.

Let us put the left-hand side of Eq. (9.175) as $f(K_1 d)$. Every term in Eq. (9.175) can be expressed in terms of $K_1 d$ using Eqs. (9.180) to (9.183). The function $f(K_1 d)$ is plotted with respect to $K_1 d$. Figure 9.19 shows such a graph. The intersections of $f(K_1 d)$ with the $K_1 d$ axis are the solutions for $K_1 d$. This method provides the solutions only for $K_1 d$, but is straightforward.

$$\begin{aligned} K_{1x\text{even}} d &= 1.1939 \\ K_{1x\text{odd}} d &= 1.1955 \end{aligned} \tag{9.201}$$

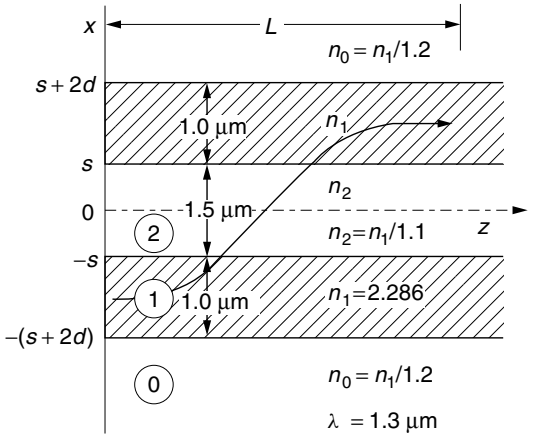


Figure 9.18 Geometry and dimensions of coupled slab guides.

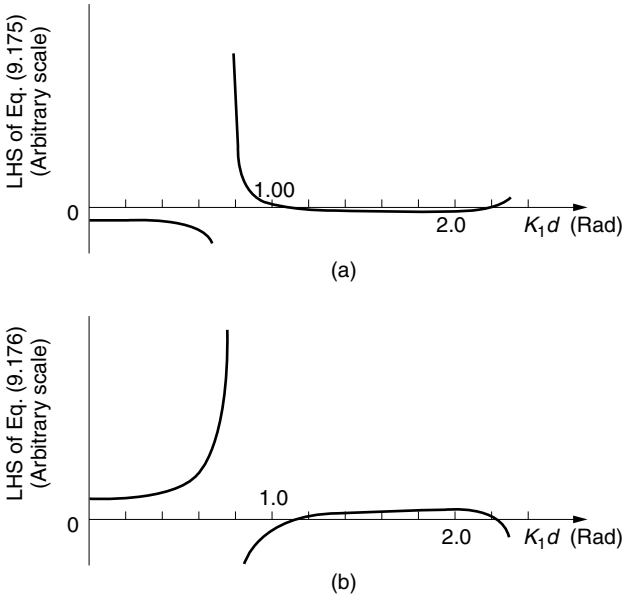


Figure 9.19 Solutions of the characteristic equations. (a) Solution for even-order modes. (b) Solution for odd-order modes. (Calculation courtesy of R. James.)

From Eq. (9.178), β_e and β_0 are calculated as

$$\begin{aligned} \beta_e &= 10.788\ \mu\text{m}^{-1} \\ \beta_0 &= 10.787\ \mu\text{m}^{-1} \end{aligned} \tag{9.202}$$

(b) In order to use the formulas in Table 9.1, γ_0, γ_2, and K_1 are calculated using Eqs. (9.180), (9.181), and (9.201).

	Even Mode	Odd Mode
γ_0 (μm^{-1})	5.6213	5.6199
K_1 (μm^{-1})	2.3877	2.3910
γ_2 (μm^{-1})	3.9351	3.9331

$$\begin{aligned} s &= 0.75\ \mu\text{m} \\ d &= 0.50\ \mu\text{m} \\ \lambda &= 1.3\ \mu\text{m} \\ n_1 &= 2.286 \\ n_0 &= n_1/1.2 \\ n_2 &= n_1/1.1 \end{aligned}$$

Figure 9.20 shows the field distribution of the even and odd TM modes calculated using the parameters and geometries shown in Fig. 9.18. □

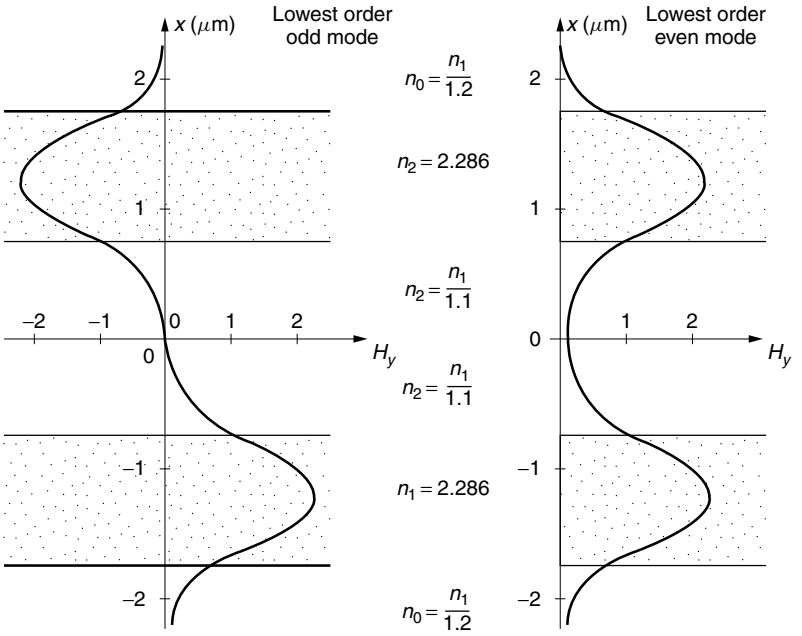


Figure 9.20 Calculation of the field $H_y(x)$ along the x axis inside the five-layer medium. (Calculation courtesy of R. James.)

9.9.3 Coupling Mechanism of the Slab Guide Coupler

We are now able to discuss the coupling mechanism of the slab guide coupler [8]. Figure 9.17 shows the cross section of a five-layer medium that forms a coupled slab guide. The coupling mechanism is interpreted from the change in the relative phase of the fundamental odd and even modes. As shown in Fig. 9.20, the field distribution of the even mode has two symmetric humps along the x axis, while that of the odd mode has two antisymmetric humps along the x direction. With the presence of the even and odd modes of equal amplitude, the two fields enhance each other and the resultant amplitude is nearly doubled in the lower guides in Fig. 9.17, whereas inside the upper guide, the two fields nearly cancel each other and the amplitude is almost null. With this distribution of the resultant fields, it appears as if only the lower guide is excited.

Because of the difference in the propagation constants of the even and odd modes, after some distance of propagation, the relative phase between the even and odd modes becomes 180°, so that the resultant field in the upper guide becomes large whereas the resultant field in the lower guide becomes null. The resultant field intensities in the upper and lower guides reverse. This is interpreted as the transfer of the light energy from the lower to the upper guide. The length of the transfer of the energy is called the *transfer length*. After another transfer length, the light energy returns to the lower guide. Light goes back and forth between the guides.

The transfer length L is given from Eqs. (9.184) and (9.185) by

$$L = \frac{\pi}{\beta_e - \beta_o} \tag{9.203}$$

The value of L with Example 9.7 is 4.41 mm.

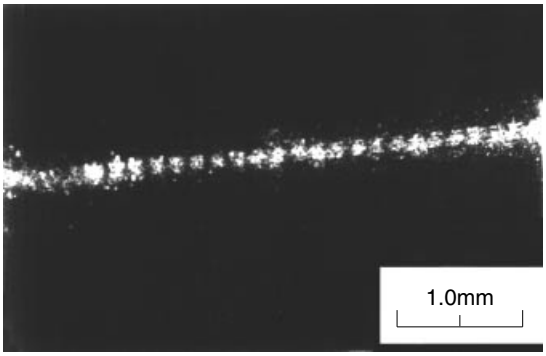


Figure 9.21 Fields seen from the top surface of the coupled slab guides. The film records light scattered from the top guiding layer. (Courtesy of N. Goto, Y. Miyazaki, and Y. Akao [9].)

The relative field strengths between the guides at a distance other than the transfer length are obtained by adding the even- and odd-mode fields with the relative phase taken into consideration.

Figure 9.21 is a photograph taken looking downward on the top surface of a five-layer medium [9]. The light is transferred back and forth between the two guiding layers. The film records only the light scattered from the upper guide layer, and the photograph shows a sequence of dots of light.

PROBLEMS

9.1 Why is only one quadrant drawn in Fig. 9.2?

9.2 Obtain the angle of propagation of the component plane waves of the highest order TM mode when the normalized thickness V is a multiple of $\pi/2$ radians.

9.3 What is the range of thickness $2d$ of a slab optical guide that has a *total* of five *possible* TM modes, where $\lambda = 0.85$ μm, $n_1 = 1.55$, and $n_2 = 1.54$.

9.4 Choose either *increase* or *decrease* in the following sentences. If the thickness of the slab guide is *decreased* with all other physical constants fixed, the normalized thickness V (*increases, decreases*) and $K_2 d$ of the TM_2 mode (*increases, decreases*) and the value of K_2 (*increases, decreases*), and this means that the value of β_2 (*increases, decreases*). Thus, in order to obtain a region of larger effective index of refraction $N = \beta_2/k$, the thickness $2d$ has to be (*increased, decreased*).

9.5 With a symmetric slab guide such as shown in Fig. P9.5, the cross-sectional distribution of the light intensity was examined. It was found that among the K values, the largest $K_{\max}$ was

$$K_{\max} = 1.0 \text{ rad}/\mu\text{m}$$

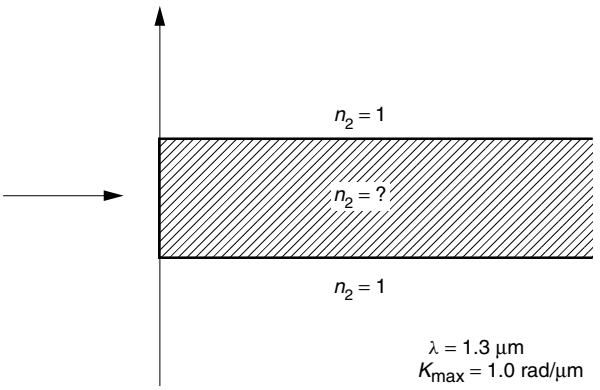


Figure P9.5 Symmetric optical guide with unknown refractive index of the core layer.

Find the index of refraction n_1 of the core layer. The index of refraction n_2 of the cladding is unity and the wavelength of the light is 1.3 μm.

9.6 An slab optical guide is bent as shown in Fig. P9.6. Find the radius of curvature r of bending that will start leaking light into the cladding for these parameters:

$$\Delta = (n_1 - n_2)/n_1 = 0.055$$

$$d = 50 \text{ μm}$$

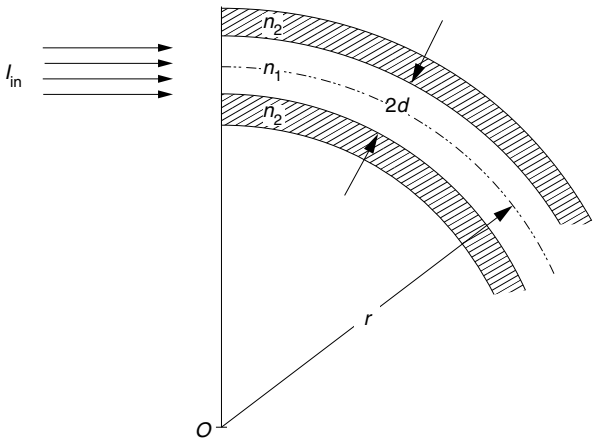


Figure P9.6 Leak from a bent optical guide.

9.7 Find the characteristic equation of an asymmetric guide with the geometry shown in Fig. P9.7, using (a) the coefficient matrix method and (b) the transmission matrix method. Compare the results. The TM wave is excited.

9.8 Verify that the ratio $E_{z0}(x)/H_{y0}(x)$ of Eqs. (9.199) and (9.200) is equal to Z_0.

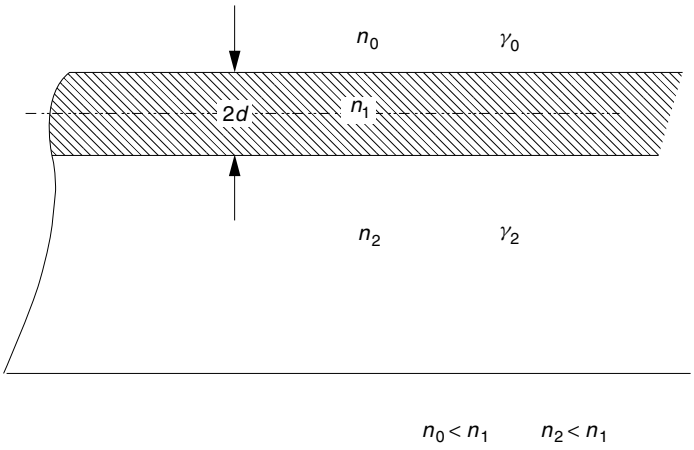


Figure P9.7 Geometry of an asymmetric guide.

REFERENCES

1. R. G. Hunsperger, *Integrated Optics: Theory and Technology*, 4th ed., Springer-Verlag, Berlin, 1995.
2. B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, Wiley, New York, 1991.
3. T. Tamir (ed.), *Guided-Wave Optoelectronics*, Springer-Verlag, Berlin, 1990.
4. A. R. Mickelson, *Guided Wave Optics*, Van Nostrand Reinhold, New York, 1993.
5. K. Iizuka, *Engineering Optics*, 2nd ed., Springer-Verlag, Berlin, 1987.
6. Y. Suematsu and K. Furuya, "Propagation mode and scattering loss of a two-dimensional dielectric waveguide with gradual distribution of refractive index," *IEEE Trans. Microwave Theory and Tech.* **MTT-20**(8), 524–531 (1972).
7. Y. Mushiake and M. Kudo, "Multilayered optical slab guide," in *Optical Guided-Wave Electronics*, by the Editorial Committee of the Special Project Research on Optical Guided-Wave Electronics of Japan, Tokyo, 1981, pp. 17–37.
8. A. B. Buckman, *Guided-Wave Photonics*, Saunders College Publishing, Austin, TX, 1992.
9. N. Goto, Y. Miyazaki, and Y. Akao, "Tunable filtering characteristics of multilayered optical couplers using surface acoustic waves," *Japanese Journal of Applied Physics*, **21**(11), 1611–1615 (1982).

10

OPTICAL WAVEGUIDES AND DEVICES FOR INTEGRATED OPTICS

In integrated optics, an optical element can be fabricated on a substrate the size of a microscope deck glass. Such elements as lenses, light deflectors, or the optical spectrum analyzer need to confine the light beam only in one dimension and the slab optical guide is widely used. However, other elements, such as the directional coupler, optical modulators, a laser diode cavity, and mode filters, need an optical guide that can confine the light in two dimensions. In this chapter, a rectangular optical guide that confines the light both in the x and y directions and transmits in the z direction is described.

The formulas developed in the previous chapter will be used freely and we highly recommend that you thoroughly review Chapter 9 before proceeding to this chapter. The latter part of this chapter also incorporates concepts from Chapter 5 on electrooptic media.

10.1 RECTANGULAR OPTICAL WAVEGUIDE

A rectangular guide [1–4] with the geometry shown in Fig. 10.1 is considered. The thickness is $2d$, and the width is $2w$. The indices of refraction n_2, n_3, n_4, n_5 of the cladding medium are not necessarily the same but are all smaller than n_1 of the core region.

10.1.1 Assumptions

The only means to confine the component waves inside the optical waveguide is total internal reflection at the boundaries. Unlike microwave waveguides, optical waveguides do not have metal walls to confine the lightwave. There are many similarities in the behavior of the modes between microwave and optical guides, but there are also differences.

In order to make the analysis manageable, certain assumptions have to be made. The first assumption is that the difference in the indices of refraction of the cladding

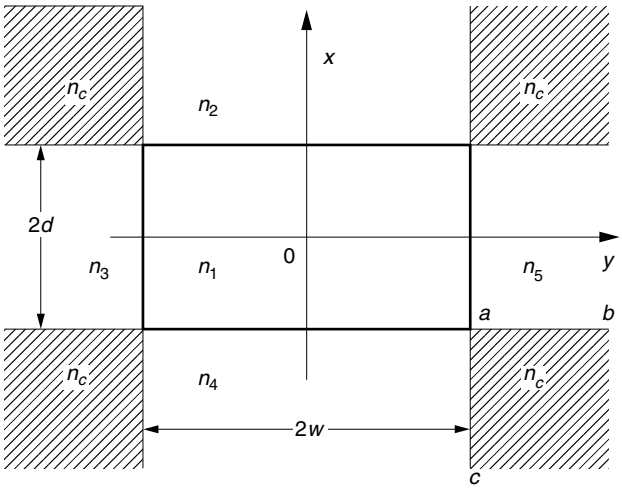


Figure 10.1 Geometry of the rectangular optical guide.

and the core is small so that all the guided component waves make a small angle with the z axis.

This assumption needs some further explanation. In Section 9.2, the assumption that $H_z = 0$ was necessary to simplify the wave equation. All the solutions so far are based on this assumption. However, the assumption that $H_z = 0$ cannot hold true in the rectangular optical waveguide. In the planar cases mentioned in the previous chapter, the component waves took a zigzag path only in the x direction, as shown in Fig. 10.2a, and $H_z = 0$ was possible. When the component waves take a zigzag path in both the x and y directions to satisfy the boundary conditions on both walls as indicated in Fig. 10.2b, the **H** component cannot be parallel to the y direction and at the same time remain perpendicular to the direction of propagation. The **H** component has to be like H_y' shown by the dotted line in Fig. 10.2b. But as long as the angle of the zigzag is small, the approximation

$$H_y \doteqdot H_y' \tag{10.1}$$

can be made.

We will still use the results in Section 9.2 as approximate expressions, and we will call the modes associated with H_y' the "TM-like modes." The common convention for designating the modes in a rectangular guide is different from what we have been using for the slab guide. In rectangular guides, the designation is with reference to the transverse **E** field. Recall that the TM mode of the slab guide had only three nonzero components: H_y, E_x, and E_z as in Eq. (9.23). The "TM-like modes" of the rectangular guide are called the E^x modes and their mode numbers are designated by the number of the extrema rather than an integer μ associated with the dispersion equation as in Eq. (9.146). The mode with p extrema in the x direction and q extrema in the y direction is designated by E_{pq}^x.

The TE mode of the slab guide again has only three nonzero components: E_y, H_x, and H_z as in Eq. (9.56). The "TE-like modes" of the rectangular guide are designated by E_{pq}^y.

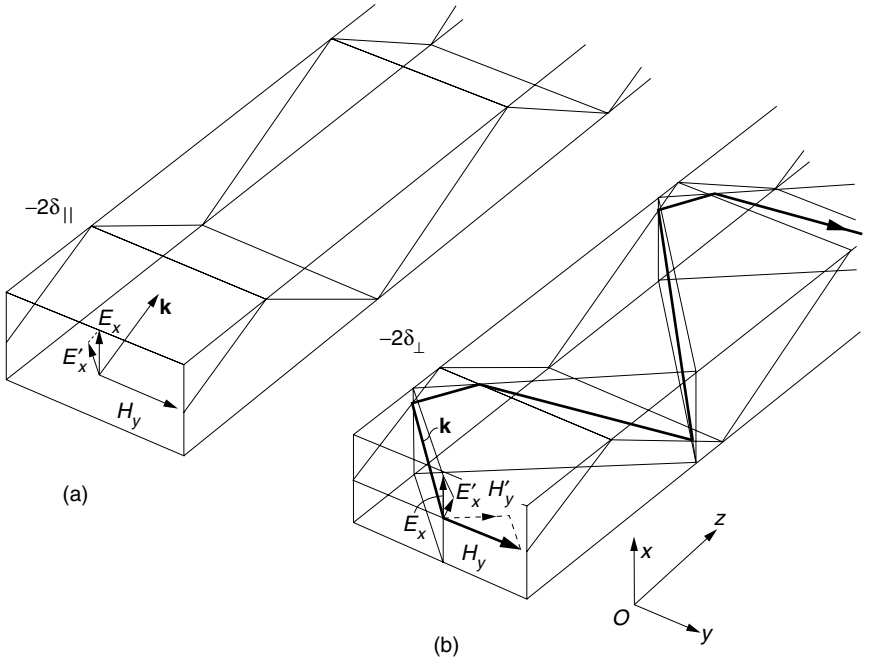


Figure 10.2 Ray path of ''TM-like mode'' or ''E^x_{pq} mode.'' (a) Zigzag only in the x direction. (b) Zigzag in both x and y directions.

Throughout the calculations, the boundary conditions associated with the hatched areas in Fig. 10.1 have been ignored. If one attempts to find a solution including, for example, the lower right hatched region, first a general solution has to be found that matches the boundaries along both a–c and a–b, and then the continuity boundary conditions have to be applied so that the solution in the hatched region connects with the decay function γ_4 along a–c and with the decay function γ_5 along a–b. Now, γ_4 is connected with K_x by a characteristic equation, and γ_5 is connected with K_y by another characteristic equation. Inclusion of the hatched region means that K_x and K_y are no longer independent, as has been assumed up to now. Furthermore, K_x and K_y are coupled in a very complicated manner (coupled equations). By ignoring the hatched regions, we are making the assumption that K_x and K_y can be chosen independently. Since the hatched areas are bordered by regions containing evanescent waves, which are much smaller in amplitude than the propagated wave, the hatched regions may be ignored without serious error.

An additional assumption made on the hatched region in Fig. 10.1 is explained in the boxed note.

10.1.2 Characteristic Equation for the Rectangular Guide

Now, the characteristic equation for the "TM-like mode" or E^x_{pq} mode will be derived. The method of the modified ray model will be used. Unlike a slab guide, the dimensions of the rectangular core are finite in both the x and y directions and standing waves (modes) exist in both the x and y directions. Let the propagation constants in the x

and y directions be K_x and K_y, respectively. The analysis is quite similar to the slab guide except that the dispersion equation has to be satisfied in both x and y directions.

The dispersion equation for the zigzag path shown in Fig. 10.2a is considered first. The field component E_x is in the plane of propagation and Eq. (2.91) is used for the phase shift due to reflection, similar to the derivation of Eq. (9.146) for the planar guide.

$$4K_x d - 2\tan^{-1}\left(\frac{\gamma_2}{n_2'^2 K_x}\right) - 2\tan^{-1}\left(\frac{\gamma_4}{n_4'^2 K_x}\right) = 2\mu\pi \tag{10.2}$$

where

$$n_2' = \frac{n_2}{n_1}, \qquad n_4' = \frac{n_4}{n_1} \tag{10.3a}$$

$$\mu = p - 1 \tag{10.3b}$$

$$K_x = \sqrt{(n_1 k)^2 - \beta^2 - K_y^2} \tag{10.3c}$$

K_y accounts for the inclusion of the $\partial^2 H_y/\partial y^2$ term in Eq. (9.5). Note the relationship between μ and p. Equation (10.2) can be reduced to the dominant mode, the TM_0 mode, with $n_2' = n_4'$ and $\mu = 0$. The TM_0 mode has one intensity maximum in the x direction. Thus, if p represents the number of intensity maxima, then μ must be replaced by $p - 1$.

The attenuation constants of the evanescent waves in the upper and lower cladding layers are

$$\gamma_2 = \sqrt{\beta^2 + K_y^2 - (n_2 k)^2} \tag{10.4}$$

$$\gamma_4 = \sqrt{\beta^2 + K_y^2 - (n_4 k)^2} \tag{10.5}$$

Now consider the y direction and the left–right zigzag path shown in Fig. 10.2b. In this case, E_x is perpendicular to the plane of the light path and Eq. (2.92) is used instead of Eq. (2.91). The dispersion equation for the left–right zigzag path is

$$4K_y w - 2\tan^{-1}\left(\frac{\gamma_3}{K_y}\right) - 2\tan^{-1}\left(\frac{\gamma_5}{K_y}\right) = 2\nu\pi \tag{10.6}$$

where $\nu = q - 1$

$$K_y = \sqrt{(n_1 k)^2 - \beta^2 - K_x^2} \tag{10.7}$$

and

$$\gamma_3 = \sqrt{\beta^2 + K_x^2 - (n_3 k)^2} \tag{10.8}$$

$$\gamma_5 = \sqrt{\beta^2 + K_x^2 - (n_5 k)^2} \tag{10.9}$$

So far, seven unknowns — K_x, K_y, γ_2, γ_3, γ_4, γ_5, and β — have been generated.

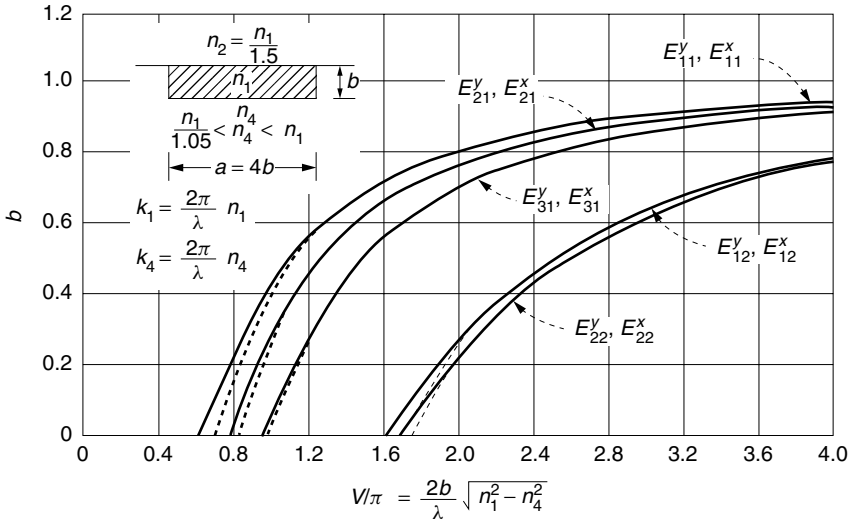


Figure 10.3 Propagation constants for different modes with respect to the dimension of the optical guide. Dispersion curves are shown for an embedded guide with height b and width $a = 4b$. The refractive index of the embedded guide is n_1, the refractive index of the medium above the guide is $n_2 = n_1/1.5$, and the refractive index of the embedding medium is n_4. The vertical axis is the normalized guide index b defined as $b = (N^2 - n_4^2)/(n_1^2 - n_4^2)$. N is the effective index of refraction. The horizontal axis is the normalized thickness V divided by π: $V/\pi = (2b/\lambda)\sqrt{n_1^2 - n_4^2}$. (After E. A. J. Marcatili [5] with x and y conventions interchanged to match the notation in this textbook.)

There are eight equations, Eqs. (10.2) through (10.9), but note that Eqs. (10.3c) and (10.7) are identical so that there are exactly the seven needed independent equations. They are transcendental equations and the solutions are obtainable only by numerical methods. The results obtained by Marcatili [5] are shown in Fig. 10.3.

For a given V, the value of b is found from the graph. From the normalized guide index b, the effective index of refraction N is calculated using

$$b = \frac{N^2 - n_4^2}{n_1^2 - n_4^2} \tag{10.10}$$

The propagation constant β of such a guide is found from N as

$$\beta = kN \tag{10.11}$$

Now, in order to clarify all the above descriptions a practical example is in order.

10.1.3 A Practical Example

Many practical devices can be generated from the geometry shown in Fig. 10.4. The next example explores some characteristics of a guide of this geometry for typical physical constants.

Example 10.1 An optical guide such as the one shown in Fig. 10.4 is called an embedded or buried guide. For the "TM-like modes," or E_{pq}^x modes, find (a) the

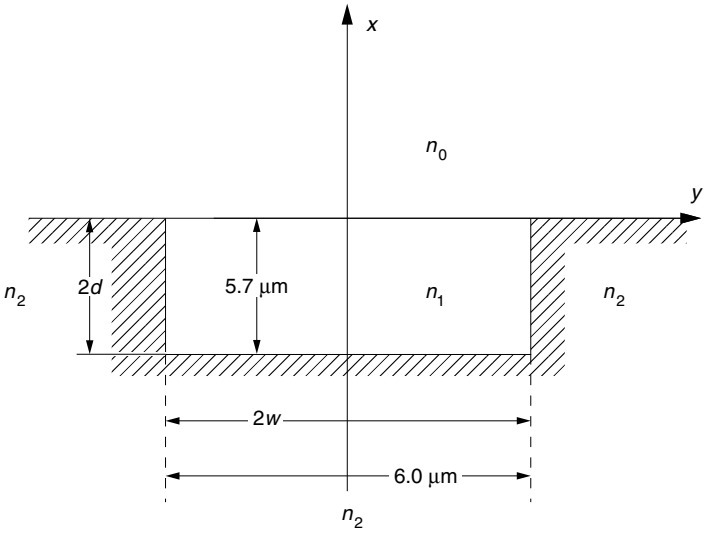


Figure 10.4 Geometry of a buried optical guide.

characteristic equation, (b) the propagation constant and the cutoff of the lowest order mode, and (c) the field distribution of each mode. The physical constants are

$$n_1 = 1.55, \quad n_2 = 1.54, \quad 2d = 5.7\ \mu\text{m}, \quad 2w = 6.0\ \mu\text{m}, \quad \lambda = 1.3\ \mu\text{m}$$

Solution Figure 10.4 is an asymmetric guide in the x direction, whereas it is a symmetric guide in the y direction. The solution can be obtained by combining the earlier results for each geometry.

The dispersion equations in the y direction are, from Eq. (10.6),

$$\tan K_y w = \frac{\gamma_2}{K_y} \quad \text{for even } \nu \tag{10.12}$$

$$-\cot K_y w = \frac{\gamma_2}{K_y} \quad \text{for odd } \nu \tag{10.13}$$

where

$$K_y = \sqrt{(n_1 k)^2 - \beta^2 - K_x^2} \tag{10.14}$$

$$\gamma_2 = \sqrt{\beta^2 + K_x^2 - (n_2 k)^2} \tag{10.15}$$

From Eqs. (10.14) and (10.15), we have

$$(K_y w)^2 + (\gamma_2 w)^2 = V_w^2 \tag{10.16}$$

$$V_w = kw\sqrt{n_1^2 - n_2^2} \tag{10.17}$$

From Fig. 9.2 with $V_w = 0.811\pi$, the TM_0 and TM_1 modes can be excited in the y direction. The corresponding $K_y w$ values are

$$\begin{aligned} K_{y0}w &= 0.37\pi \\ K_{y1}w &= 0.70\pi \end{aligned} \tag{10.18}$$

An approximate dispersion equation in the x direction is, from Eq. (9.157),

$$-\cot 2K_x d = \frac{\gamma_2}{n_2'^2 K_x} \tag{10.19}$$

and

$$(\gamma_2 d)^2 + (K_x d)^2 = V_d^2 \tag{10.20}$$

$$V_d = kd\sqrt{n_1^2 - n_2^2} \tag{10.21}$$

Figure 9.16 is used with $V_d = 0.77\pi$. In the x direction again there are two modes:

$$\begin{aligned} K_{x0}d &= 0.42\pi \quad \text{for } TM_0 \\ K_{x1}d &= 0.76\pi \quad \text{for } TM_1 \end{aligned} \tag{10.22}$$

Now, the modes for the two directions are combined. First, the modes will be expressed using the notation of E_{pq}^x, where p and q represent the number of the extrema for the x and y directions. The TM_0 mode has one and the TM_1 has two extrema. In terms of E_{pq}^x, the available modes are E_{11}^x, E_{12}^x, E_{21}^x, and E_{22}^x. The propagation constant of the lowest order mode E_{11}^x is, from Eq. (10.14),

$$\beta_{11} = \sqrt{(n_1 k)^2 - (K_{x0})^2 - (K_{y0})^2} \tag{10.23}$$

With the results in Eqs. (10.18) and (10.22) combined with the values of w and d, the propagation constant of E_{11}^x is

$$\beta_{11} = 7.46 \text{ rad}/\mu\text{m} \tag{10.24}$$

The cutoff condition for the E_{11}^x mode is determined by the asymmetric mode alone because the cutoff of the symmetric mode is $V = 0$. The cutoff for the asymmetric mode is $V_d = \pi/4$.

$$2d_c = \frac{\lambda}{4\sqrt{n_1^2 - n_2^2}} = 1.85 \ \mu\text{m} \tag{10.25}$$

Combining the distributions in the x and y directions, the field distribution shown in Fig. 10.5 is obtained in a manner similar to that described in Section 9.4. □

10.2 EFFECTIVE INDEX METHOD FOR RECTANGULAR OPTICAL GUIDES

The effective index method [6,7] is simple and useful for finding an approximate expression for the propagation constant of a rectangular optical guide. The effective index method is conceptually more straightforward than what was explained in the previous section and provides almost identical results. The method is explained

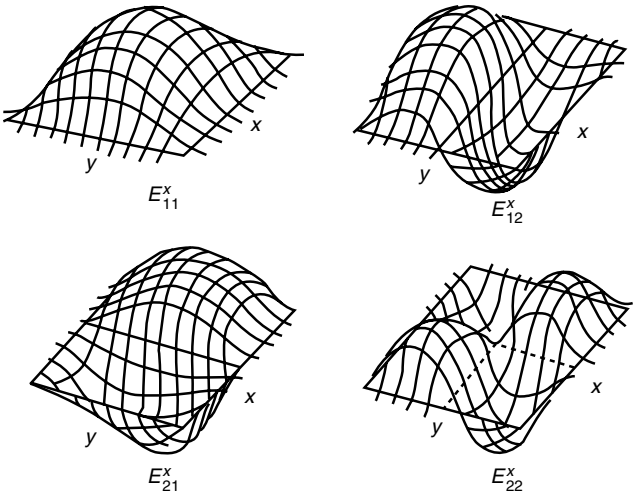


Figure 10.5 Mode patterns in a rectangular guide.

using the rectangular optical guide shown in Fig. 10.6a, which was dealt with in Example 10.1. The procedure consists of two steps; each step involves finding the propagation constant of a slab optical guide of an infinite extent. The first step is to find the propagation constant β_w of a slab optical guide with thickness $2w$ as shown in Fig. 10.6b. The second step is to calculate the propagation constant β_{wd} of the slab optical guide with thickness $2d$, but using the effective index of refraction $N = \beta_w/k$ obtained in the first step as the index of refraction of the core layer, as shown in Fig. 10.6c. The propagation constant β_{wd} of the second optical guide is the desired propagation constant of the rectangular optical guide.

Step 1. With the geometry of the first slab optical guide shown in Fig. 10.6b, the propagation constant is found. Assuming even TM modes, the characteristic equation, from Eqs. (9.26) and (9.27), is

$$n^2 Kw \tan Kw = \gamma w \tag{10.26}$$

with

$$(Kw)^2 + (\gamma w)^2 = V_w^2 \tag{10.27}$$

where

$$V_w = kw\sqrt{n_1^2 - n_2^2}$$

The intersections of the two curves provide the solutions. Let's say one of the solutions is K_y. Then, the propagation constant β_w is, from Eq. (9.6),

$$\beta_w^2 = (n_1 k)^2 - K_y^2 \tag{10.28}$$

and the effective index of refraction N is, from Eq. (9.44),

$$N_1^2 = \frac{(n_1 k)^2 - K_y^2}{k^2} \tag{10.29}$$

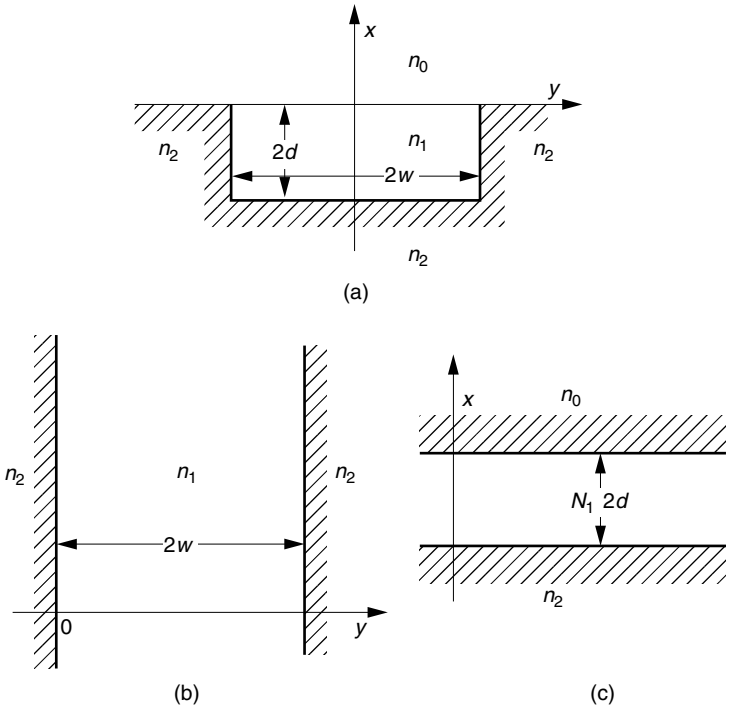


Figure 10.6 Explanation of the effective index method.

Step 2. The effective index of refraction N_1 is used as the refractive index of the core layer of the slab optical guide shown in Fig. 10.6c, which is an asymmetric guide. The characteristic equation from Eq. (9.149) is

$$\frac{K\left(\dfrac{\gamma_0}{n_0'^2}+\dfrac{\gamma_2}{n_2'^2}\right)}{K^2-\dfrac{\gamma_0\gamma_2}{n_0'^2 n_2'^2}}=\tan 2Kd \tag{10.30}$$

where

$$\begin{aligned} n_0' &= \frac{n_0}{N_1} \\ n_2' &= \frac{n_2}{N_1} \end{aligned} \tag{10.31}$$

When n_0 is air and is much smaller than any other refractive index, Eq. (10.30) can be approximated in a manner similar to that mentioned in Example 9.6 and becomes

$$-\cot 2Kd=\frac{\gamma_2}{n_2'^2 K} \tag{10.32}$$

with

$$(Kd)^2+(\gamma_2 d)^2=V^2 \tag{10.33}$$

where

$$V = kd\sqrt{N_1^2 - n_2^2}$$

Let's say one of the solutions of Eqs. (10.32) and (10.33) is K_x; then the propagation constant β_{wd} is finally, from Eq. (9.150),

$$\beta_{wd}^2 = (N_1 k)^2 - K_x^2 \tag{10.34}$$

β_{wd} is the required propagation constant of the rectangular optical guide.

The results obtained by the effective index method are compared with those of the previous section. Inserting Eq. (10.29) into (10.34) gives

$$\beta_{wd} = \sqrt{(n_1 k)^2 - K_x^2 - K_y^2} \tag{10.35}$$

This expression is identical to Eq. (10.23). The difference, however, is that K_y in Eq. (10.35) was obtained with $n_2' = n_2/N_1$, whereas K_{y0} was obtained with $n_2' = n_2/n_1$. Further discussions about method approximations can be found in Refs. 6 and 7.

10.3 COUPLING BETWEEN RECTANGULAR GUIDES

The results with the slab guides can immediately be used for calculating the coupling between rectangular guides. The geometry shown in Fig. 10.7 is considered. This geometry was generated by slicing the five-layer slab guide, such as mentioned in Section 9.9, by a plane parallel to the x–z plane with width $2w$. In order to make use of the results for the TM modes in Section 10.1.3, the E_{pq}^x mode will be considered. If we follow the assumption made in Section 10.1.1 that the x and y components are mathematically separable, the propagation constant β of the coupled rectangular guide can be obtained by combining $K_{1x\ \text{even}}$ and $K_{1x\ \text{odd}}$ of the coupled slab guide in the x direction with the K_{y0} of a symmetric guide in the y direction.

Even though the values of $K_{1x\ \text{even}}$ and $K_{1x\ \text{odd}}$ obtained in Section 9.9.1 are immediately usable, the values of K_{y0} have to be calculated using the characteristic equations (9.26) and (9.27) for the TM mode. Once these values are obtained, the required propagation constants are found as

$$\beta_{1e} = \sqrt{(n_1 k)^2 - (K_{1x\ \text{even}})^2 - K_{y0}^2} \tag{10.36}$$

$$\beta_{1o} = \sqrt{(n_1 k)^2 - (K_{1x\ \text{odd}})^2 - K_{y0}^2} \tag{10.37}$$

and the coupling length is

$$L = \frac{\pi}{\beta_{1e} - \beta_{1o}} \tag{10.38}$$

An actual calculation is demonstrated in the next example.

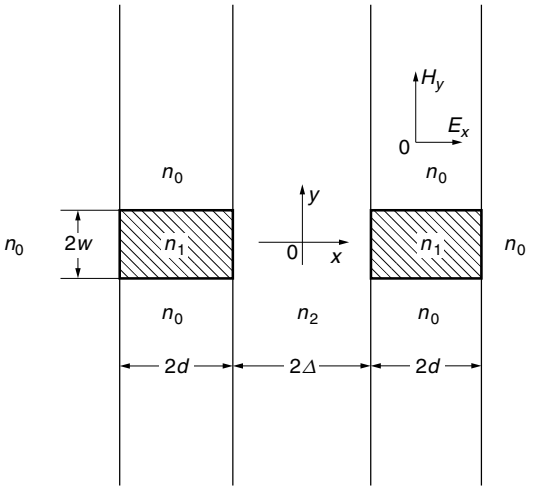


Figure 10.7 Rectangular coupled guides.

Example 10.2 Find the transfer length L of the rectangular coupled guide with the dimensions and physical parameters shown in Fig. 10.8.

Solution Simultaneous equations (9.26) and (9.27) have to be solved first. The same approach as in the previous example will be used. Set

$$f(K_y w) = n^2 K_y w \tan K_y w - \sqrt{V_0^2 - (K_y w)^2} \tag{10.39}$$

The function $f(K_y w)$ is plotted with respect to $K_y w$. The intersections of $f(K_y w)$ with the $K_y w$ axis are the solutions.

$$K_{y0} = 3.68903 \tag{10.40}$$

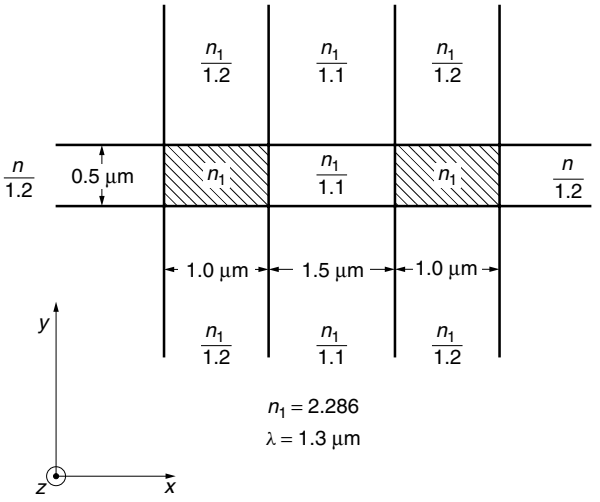


Figure 10.8 Geometry and physical parameters of the rectangular coupled guide.

The progagation constants β_{1e} and β_{1o} are found by inserting Eqs. (9.201) and (10.40) into Eqs. (10.36) and (10.37).

$$\beta_{1e} = 10.13728\ \mu\text{m}^{-1}$$

$$\beta_{1o} = 10.13652\ \mu\text{m}^{-1}$$

The coupling length L for the rectangular coupled guide is

$$L = \frac{\pi}{\beta_{1e} - \beta_{1o}} = 4.14\ \text{mm} \qquad \square$$

10.4 CONFLECTION

Reflection, refraction, and diffraction are words describing the change in the direction of light rays. The Latin verb *flectere* means "to bend" and *frangere* means "to break."

Di Francia [8] coined a new word — *conflection* — meaning the change in the direction of propagation at the joint of two cones of guiding layers.

When a light beam is launched into a uniform flat slab optical guide, it zigzags in the plane perpendicular to the boundary. However, in the plane parallel to the boundary, the light beam will no doubt propagate along a straight line because it is the fastest path to go from point A to point B. This is Fermat's principle.

Would it still be the fastest path between A and B if the guide surface were bent, and if the light ray were not allowed to leave the surface? Let us conduct a small experiment. Draw a straight line $\overline{AB}$ in the center of a piece of cardboard as shown in Fig. 10.9a. Lay a piece of string over the line and attach the string to the cardboard at points A and B. Fold the cardboard any way you wish with the string outside so that the string can be moved around but does not move off the surface, as shown in Fig. 10.9b. You will soon discover that no matter how you bend the cardboard, the string neither sags nor moves away from the pencil line $\overline{AB}$.

From this fact, one can find the path of light in a folded guide. The law of conflection is that the light path is bent such that the light path becomes a straight line when the bend is developed (unfolded) to be flat. This can be restated as saying that the incident angle θ_i with respect to the normal to the edge is the same as the transmitted angle θ_t as shown in Fig. 10.9c. This principle will be explained further by an example.

Example 10.3 Find the light path when a light beam is launched into a tetrahedral (figure with four equivalent triangular faces) guide as shown in Fig. 10.10.

Solution Let the light beam incident on the tetrahedron $abcd$ be parallel to the optical axis (dashed line), as shown in Fig. 10.10(1).

First, face abc is folded in as $ab'c$ and a straight line $\overline{AB}$ is drawn in face $ab'c$, as in Fig. 10.10(2).

Face $ab'c$ is folded back to the original place abc. Now, face bcd is unfolded so that bcd' and abc are in the same plane and the incident beam is further extended to $\overline{BC}$ as in Fig. 10.10(3).

Face bcd' is folded back to the original position and then, this time, face bcd is folded inside the tetrahedron until face $b'cd$ is pressed down onto the surface of the

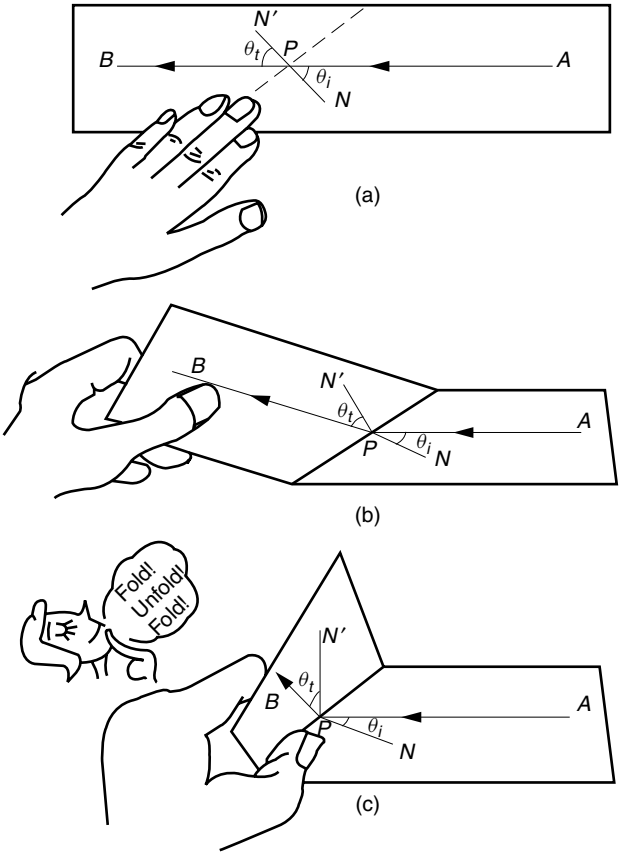


Figure 10.9 Operation of FUF (fold–unfold–fold). (a) Developed (opened up to be flat). (b) Folded halfway. (c) Folded at 90°.

substrate. Then the line $\overline{BC}$ drawn on face $b'cd$ is extended onto the substrate to finally draw the emergent beam.

Figure 10.11 explains how to draw the same light path using the equality of the angles of incidence and transmission across the bend. The incident light beam is drawn at the far right-hand side of Fig. 10.11, and the faces of the tetrahedron that the beam encounters are drawn from right to left. Angles of incidence and transmission for points A, B, and C are made equal, giving the light path shown.

A planar lens based on this principle is called a *conflection* lens. The conflection lens will be described as a practical example of the principle of conflection. □

10.4.1 Conflection Lens

Figure 10.12 shows a conflection lens [8]. A rectangular block with one of its ends shaped into a semicircular cylinder is placed on top of a larger block. It looks like the deck of a ship. The entire surface of the two-tier substrate block is covered by a thin film, which acts as the nonplanar guiding (core) layer. When light is launched in the lower deck, the light is guided in the core layer on the surface of the structure and

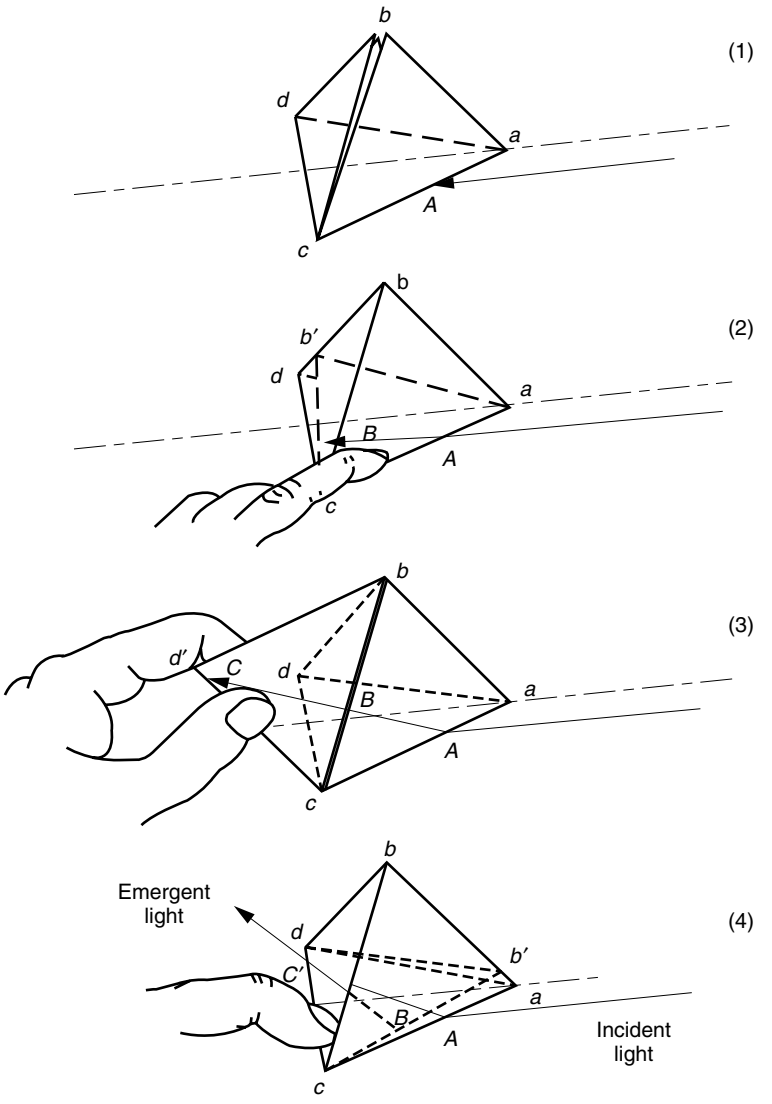


Figure 10.10 Conflection by a tetrahedron.

eventually goes up to the upper deck. If a parallel beam is launched in the lower deck, the beam will be focused to a point on the upper deck because of the conflections taking place both at the foot and top of the short semicircular cylinder. The paths of three beams are shown in Fig. 10.12. The center beam that has entered on the optical axis keeps its direction parallel to the optical axis all the way. The beams on either side, however, change their directions at each conflection point, as shown in the figure. The three beams meet at the focal point on the upper deck.

The exact path of the light will be traced. Let the light beam $\overline{AB}$ be incident parallel to the optical axis. At the foot of the semicircular cylinder, conflection takes place. The path of light is determined by unfolding the folded small section of the side wall

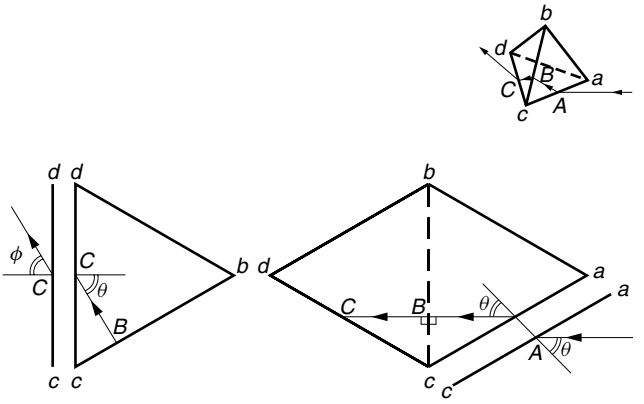


Figure 10.11 Light path in the guide formed on the surface of a tetrahedral substrate. The law of conflection is used.

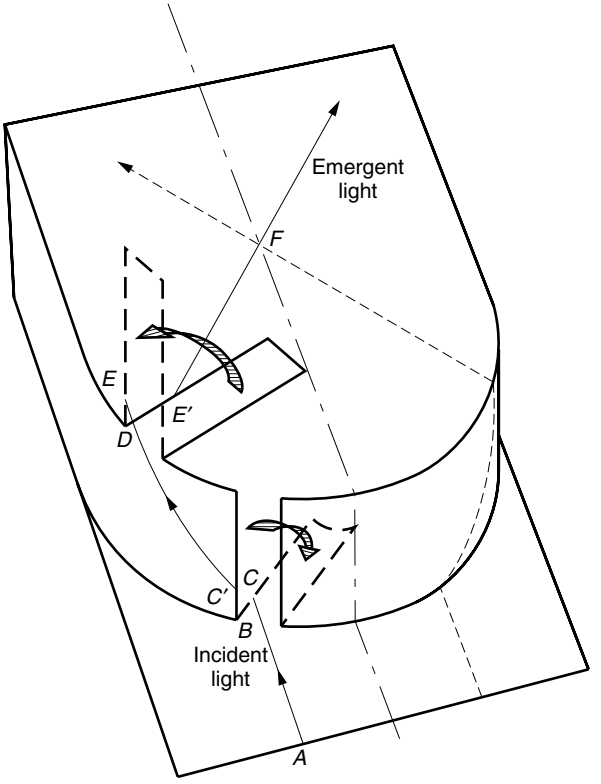


Figure 10.12 Tracing the path of light of a conflection lens by the FUF law.

into the cylinder until it coincides with the plane of incidence. The path of light is a straight line $\overline{AC}$. The small unfolded section is then folded back to its original position. The curved wall is then unfolded to be flat, and the line $\overline{AC}$ is extended as the straight line $\overline{C'D}$.

At the edge made by the top plane and the side wall another conflection takes place. Unfold the top surface until it coincides with the plane of the cylinder wall, and draw the straight line $\overline{DE}$. Curve the cylinder wall back to its usual position and then fold the section back to the top surface. The extension $\overline{E'F}$ of $\overline{ED}$ is the emergent light path.

Figure 10.12 illustrates a conflection lens in which the incident and emergent beams are not in the same plane. The incident and emergent beams can be in the same plane by replacing the semicircular cylinder by a whole circular cylinder. Even though the idea of the conflection lens is interesting, scattering at the folded sections creates loss. Careful design of the surface is important to minimize the aberration of the lens.

10.5 VARIOUS KINDS OF RECTANGULAR OPTICAL WAVEGUIDES FOR INTEGRATED OPTICS

Representative structures of rectangular optical guides are summarized in Fig. 10.13. Each one is different. The choice has to be made from considerations such as connections with adjacent devices, difficulty of fabrication, loss in the guide, kinds of substrate materials to be used, and highest temperature allowable for fabrication if other devices are to share the same substrate with the optical guide. The types of rectangular guides shown in Fig. 10.13 are the following: ridge guide, rib guide, strip-loaded guide, embedded guide, immersed guide, bulge guide, metal guide, buffered metal guide, and photochromic flexible guide.

The various types of guides are explained below.

10.5.1 Ridge Guide

Figure 10.13a shows the structure of the ridge guide [9]. The normal difference in the refractive indices between the film and substrate is only a fraction of a percent, but if air is used as the covering medium, the difference in the refractive indices at the air–film interface is significantly larger and the evanescent wave is significantly reduced. Any roughness on the air–film interface upsets the guiding condition and creates radiation loss from the guide.

The core strip is deposited directly onto the substrate of a lower index of refraction. It looks structurally simple but strenuous efforts have to be made to ensure that the three side walls of the core strip are extra smooth to minimize the radiation loss.

10.5.2 Rib Guide

The rib guide shown in Fig. 10.13b is a film layer deposited on a substrate where the thickness of a strip along the center of the film is increased. Referring back to Fig. 9.6c, with an increase in the thickness of the guide, the angle θ_2 of the direction of the component wave has to increase to match the boundary condition. The increase in θ_2 means an increase in the effective index of refraction $N = n_1 \sin\theta_1$ in the center region of the film and the light is confined to this region.

This structure again has the core–air interface on three walls and stringent requirements on the smoothness of the surfaces have to be met.

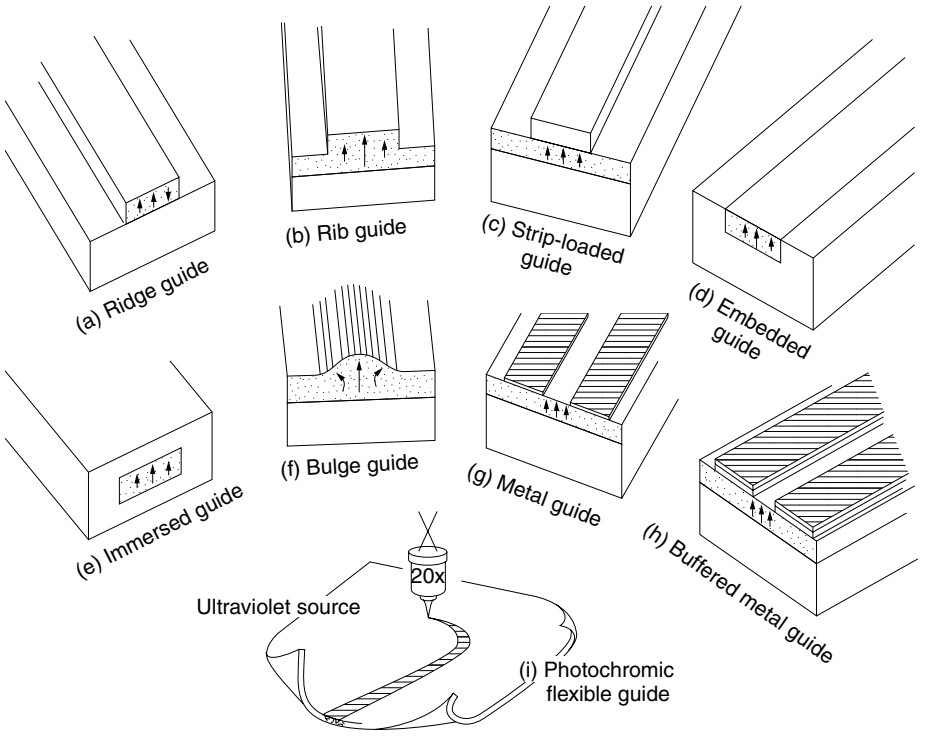


Figure 10.13 Various kinds of rectangular guides. The field lines are the **E** field of the E_{11}^{x} mode (x is in the vertical direction).

10.5.3 Strip-Loaded Guide

A film is first deposited on a substrate of lower index of refraction. Then a strip whose index of refraction is lower than the film but higher than the covering to the film (e.g., air) is deposited on the film. This is the strip-loaded guide as shown in Fig. 10.13c. The principle of this guide can be explained by referring back to Fig. 9.4. In the region where the strip is deposited, the effective depth of the evanescent wave is deeper than that in the surrounding region and the angle θ_2 of the direction of propagation of the component wave has to be increased. Therefore, the effective index of refraction N is increased compared to that in the surrounding medium, and the light is confined to the region below the strip.

An advantage of this guide is that the stringent requirement of smooth surfaces at air–film interfaces is removed.

10.5.4 Embedded Guide

In the embedded or buried guide, only one of the surfaces is exposed to air, as shown in Fig. 10.13d, and the smooth surface requirement is not as demanding as the ridge or rib guide. When an electrooptic material is used for the core, the electrodes can be deposited in direct contact with the core so that the external field can be effectively applied. If two embedded guides are placed in close proximity so that the evanescent

fields overlap, then coupling can occur between the guides via the evanescent field. This is the basis of the optical guide coupler.

10.5.5 Immersed Guide

In the immersed guide, the core is completely immersed in the cladding medium, as shown in Fig. 10.13e. The mathematical analysis is simpler because of the symmetries in both the x and y directions. There is no cutoff of the dominant mode. This geometry, however, is not suitable for devices that need electrodes.

10.5.6 Bulge Guide

The bulge guide is a variation of the rib guide, as shown in Fig. 10.13f. The shape is not too critical but again the smoothness of the surface is important to minimize radiation loss. It possesses characteristics similar to the rib guide.

10.5.7 Metal Guide

The guiding film layer of this guide is covered by a pair of metal strips, as shown in Fig. 10.13g. In the section where the film is covered by the metal, the evanescent wave is removed and the effective height of the guide is reduced. In Fig. 9.6c, as the guide thickness decreases, θ_2 decreases, and therefore so does the effective index of refraction N. The center region is bordered by regions of lower index of refraction, and a guide is formed.

Some loss of light power occurs on the metal surface because the metal is not a perfect conductor at optical frequencies. The complex refractive index of silver at optical frequencies is $\tilde{n}^2 = -16.32 - j0.5414$, and that of aluminium is $\tilde{n}^2 = -39.88 - j15.56$. The $\tilde{n}^2$ of a dielectric material like glass is a pure real number. The imaginary part of $\tilde{n}$ is associated with the conductivity of the metal and a larger number corresponds to a higher metal conductivity.

In a metal guide, the metal layers can conveniently be used as electrodes for applying an external control field for an electrooptic device. The electrooptic light modulator is an example of using such a feature.

10.5.8 Buffered Metal Guide

The light loss in the metal layer in the metal guide can be minimized by placing a thin dielectric buffer layer between the metal and film layers, as shown in Fig. 10.13h. The index of refraction of the buffer layer has to be lower than that of the film to confine the light primarily in the film layer. By adjusting the thickness of the buffer layer, the absorption for selected modes can be reduced. This property can be used as a mode filter. Since the thickness of the buffer layer is only a few tenths of a micrometer, the metal layer can still be used effectively for applying the external electric field for electrooptic devices.

10.5.9 Photochromic Flexible Guide

A photochromic material is a transparent material that changes to dark blue when it is exposed to ultraviolet light. Besides changing color, its index of refraction increases

for light wavelengths longer than the exposing ultraviolet light. This property of photochromics can be used to fabricate an optical guide [10].

A photochromic layer is deposited onto an acetate sheet or a glass substrate. Using an ultraviolet beam, an optical guide pattern is written onto the photochromic layer, as shown in Fig. 10.13i. Unlike other guides discussed in this section, the photochromic guide is flexible. The guide fades away in a few tens of minutes, which can be an advantage or disadvantage depending on the application. The film can be reactivated countless times with minimum fatigue.

10.6 POWER DIVIDERS

Figure 10.14 shows an example of a network of fiber-optic communication systems. It is evident that the simple Y junction power divider is a crucial component of such a network. Important considerations in designing the power divider are the following:

1. Accurate division of the input power into the prescribed proportions.
2. Minimum loss of light power in the divider.
3. Easy coupling to an external optical fiber circuit with minimum coupling loss.

10.6.1 The Y Junction and Arrayed-Waveguide Grating

The Y junction, although structurally very simple as shown in Fig. 10.15a, is nevertheless a very useful device. A practical difficulty in fabrication is the restriction on the angle θ between the output ports. It has to be less than 1° if the branching loss is to be held below 1 dB. This restriction means the divider will be long if the output ports are to be sufficiently separated for coupling to an external circuit.

Multiport junctions can be fabricated by staggering the Y junction, as shown in Fig. 10.15b. The number of ports increases very rapidly by staggering the junctions. However, the insertion loss of the divider increases each time a new stage is added.

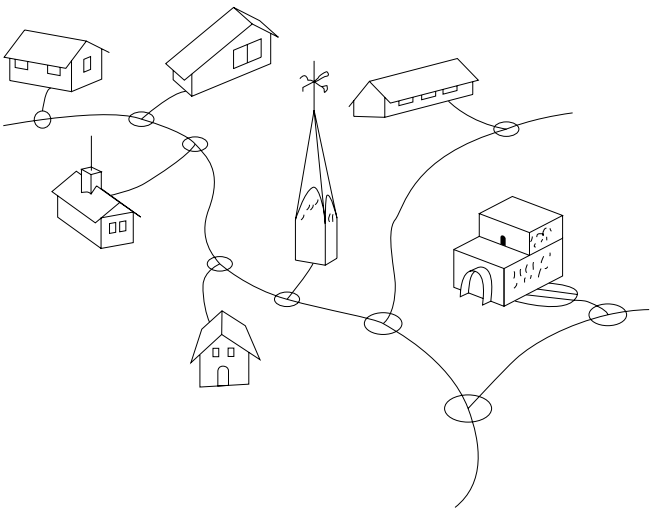


Figure 10.14 Fiber-optic network connected by three-port power dividers.

The multimode interference (MMI) splitter shown in Fig. 10.15c aleviates the difficulties associated with concatenation of the Y junctions. The MMI splitter consists of a single-mode input port and several single-mode output ports with a multimode region in the middle section. The axes of the output ports are positioned to coincide with the maxima of the standing-wave pattern established in the multimode region.

The parameters can be designed for low loss as well as for a balanced division of power in the output ports [11,12].

Figure 10.15d shows the geometry of the arrayed-waveguide grating (AWG). This device sorts the closely spaced multi-wavelength light signals in the input waveguides into the output waveguides in accordance with the wavelength of the signal. It plays an important role in wavelength division multiplexing (WDM) of fiber-optic communcation systems [13,14].

N input waveguides (typically $N = 8 \sim 128$) are connected to a slab optical guide, which forms a free space region of propagation. Light from any one of the input waveguides propagates across the free space region and illuminates the entrances of all the M arrayed waveguides (M is larger than N and typically $M \doteqdot 2N$).

The arrayed waveguides as a whole act as a grating. The lengths of the arrayed waveguides are successively incremented with a step length of ΔL. The exits of the M arrayed guides are connected to another free space region. To the end of the free space region, N output waveguides are connected.

The direction θ_{out} of the output beam is where constructive interference of the light emergent from the arrayed waveguides takes place. The condition for constructive

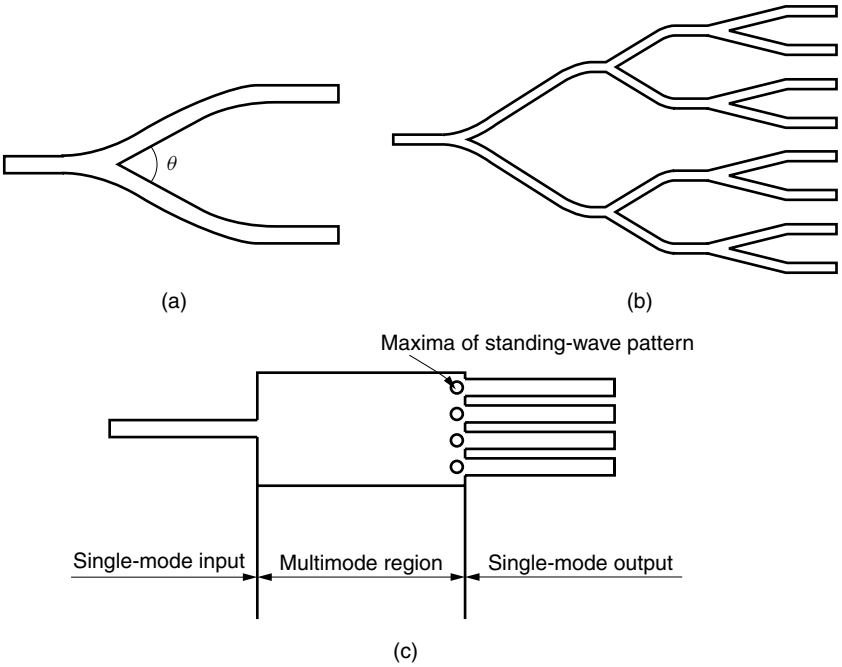
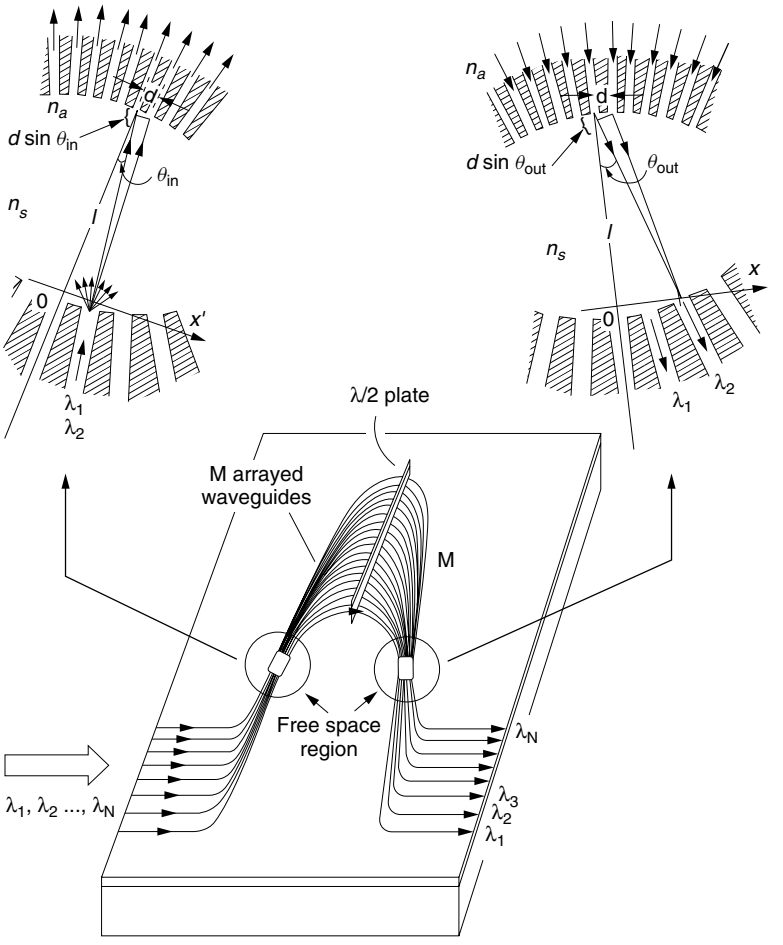
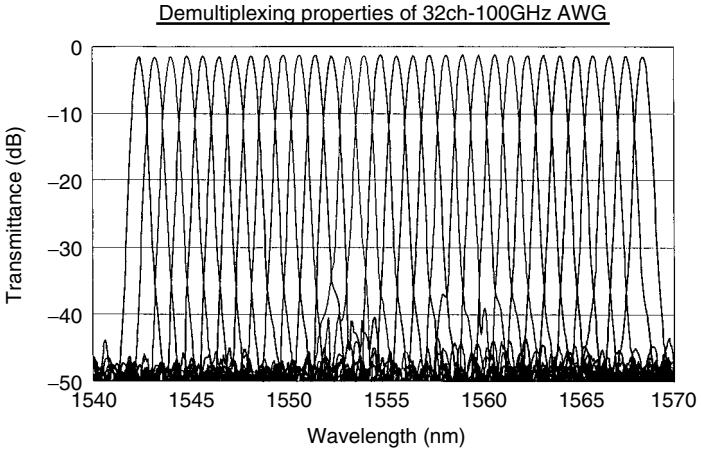


Figure 10.15 Branching networks. (a) Y junction. (b) Eight-port branching circuit made of Y junctions. (c) Multimode interference (MMI) splitter. (d) Arrayed-waveguide grating (AWG). (e) Demultiplexing properties of an arrayed-waveguide grating. (After K. Okamoto and A. Sugita [16].)



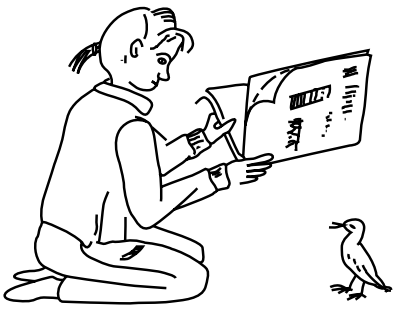
(d)



(e)

Figure 10.15 (*Continued*)

At first glance, the Y junction might seem too inefficient a device for dividing the power into many subscribers, but in actual fact the number of output ports grows very rapidly. One only needs to stagger the Y junctions 20 stages to make enough ports to connect an entire city of one million households. An interesting analogy is that a sheet of newspaper can be divided into a million pieces by folding it only 20 times: $2^{20} = 1{,}048{,}576$.



What is the least number of times you have to fold a sheet of newspaper to make a million pieces? Yes, it is only 20 times.

interference of the light with wavelength λ originating from the input waveguide angle θ_{in} is

$$n_a \Delta L + n_s d(\sin\theta_{\text{in}} + \sin\theta_{\text{out}}) = p\lambda \tag{10.41}$$

where n_a and n_s are the refractive indices of the arrayed waveguides and the free space region, respectively. The arrayed waveguide spacing is d and the diffraction order is p. Thus, from Eq. (10.41), θ_{out} for a given λ and θ_{in} is found, and by manipulating the wavelength, the input signal at θ_{in} can be steered to any desired output waveguide θ_{out}. This means that the AWG can perform an $N \times N$ matrix connection.

Next, the dispersion of the position of the output beam will be calculated. For small θ_{out}

$$\sin\theta_{\text{out}} \doteqdot x/l \tag{10.42}$$

where l is the length of the free space region. Form Eqs. (10.41) and (10.42),

$$\frac{dx}{d\lambda} = \frac{lp}{n_s d} \tag{10.43}$$

The value of the diffraction order p is obtained from Eq. (10.41) for small θ_{in} and θ_{out} as $p = n_a \Delta L/\lambda$ and Eq. (10.43) becomes

$$\frac{dx}{d\lambda} = \frac{l n_a \Delta L}{n_s \lambda d} \tag{10.44}$$

Finally, the channel spacing of the AWG will be calculated. The diffraction pattern from an array of apertures of a finite number has already been calculated in Chapter 1. The result in Fig. 1.17 indicates that the first zero of an array with array width c is situated at

$$f_x = 1/c \tag{10.45}$$

where c is the array aperture.

This result is immediately applicable to our case by changing the physical parameters as

$$\begin{aligned} \lambda &\Rightarrow \lambda/n_s \\ c &\Rightarrow Md \\ f_x &\Rightarrow \frac{x}{\lambda l} n_s \end{aligned} \tag{10.46}$$

For simplicity $\theta_{\text{in}} = 0$ is assumed.

The value x_W of the first zero of the radiation lobe in our case is

$$x_W = \frac{\lambda l}{M n_s d} \tag{10.47}$$

The quantity $2x_W$ is the "zero-crossing beam width" of the light peak at the entrance to the output waveguide.

If the channel wavelength spacing $\Delta\lambda$ is defined as the amount of change in the wavelength that shifts the peak to the first zero of the adjacent peak, namely,

$$\Delta x = x_W \tag{10.48}$$

where

$$\Delta x = \frac{dx}{d\lambda} \Delta\lambda \tag{10.49}$$

then from Eqs. (10.44), (10.47), (10.48), and (10.49) the channel spacing becomes

$$\Delta\lambda = \frac{\lambda^2}{n_a M \Delta L} \tag{10.50}$$

The channel wavelength spacing can be narrowed by increasing the number of the arrayed waveguides M and the step length ΔL.

The polyimide half-waveplate inserted in the midpoint of the arrayed waveguides acts as a TE–TM mode converter (Section 6.4.1) [15]. As seen from Eq. (10.41), the direction θ_{out} of the light peak depends on the effective index of refraction n_a, which is different for the TE and TM modes. The light in the TE mode (TM mode) that has travelled the first half of the arrayed waveguides travels in the TM mode (TE mode) in the second half of the arrayed waveguides. This averaging effect eliminates the polarization dependence of the AWG.

Figure 10.15e shows an example of the characteristics of the AWG [16]. In this example, the transmission of light is from one optical guide. The direction of transmission is from left to right and the device is used as a multiplexer, but the AWG is reciprocal and $\lambda_1, \lambda_2, \ldots, \lambda_N$ signals can be multiplexed into one optical waveguide.

10.6.2 Power Scrambler

Figure 10.16 shows an 8×8 scrambler made out of Y junctions. The power of any one input port is divided equally into the 8 output ports. Shown in Fig. 10.16a is an example of the division of inputs Ⓐ and Ⓔ into 8 output ports each of which has $A + E$. The flow of signal A is indicated by the heavy lines and that of signal E is indicated by the dashed lines. Each output has the same mixture of A and E. Such a

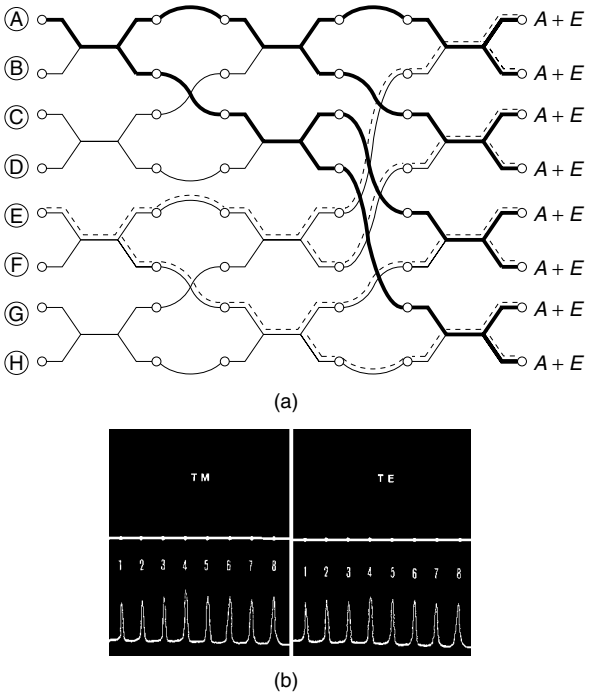


Figure 10.16 An 8×8 star coupler. (a) Geometry of 8×8 star coupler. (b) Output from the first through eighth output ports when the first input port is illuminated. (After W. Doldissen et al. [17].)

scrambler, for instance, distributes 8 different TV channels to 8 subscribers, each of whom can receive 8 TV channels.

The connecting guides for such a geometry as shown in Fig. 10.16a cannot avoid crossing each other, but the crosstalk between the intersecting lines can be suppressed below -23 dB by choosing intersecting angles larger than 10° [17].

Figure 10.16b shows the electrical output from the image made by an infrared vidicon camera.

A power scrambler such as the one shown in Fig. 10.16 is sometimes called a *star coupler* because of its configuration.

10.7 OPTICAL MAGIC T

Figure 10.17 shows the geometry of an optical magic T [18,19]. The function is quite similar to the magic T of microwave technology. Arms 1 and 2 are input arms and arms 3 and 4 are output arms. The dimensions of the input arms are identical, but those of the output arms are different. All arms are single-mode guides.

Figure 10.17a shows the field distribution when input arms 1 and 2 are equal in both amplitude and phase (symmetric feed). The distribution of the resultant field at the junction is double humped. Arm 3 is excited more than arm 4 because the shape of the resultant field at the junction conforms more with the TM_0 mode of arm 3 than with that of arm 4. The closer the shape of the illuminating light is to the amplitude distribution of the TM_0 mode, the higher the efficiency of excitation is because of the boundary condition of the continuity of the light.

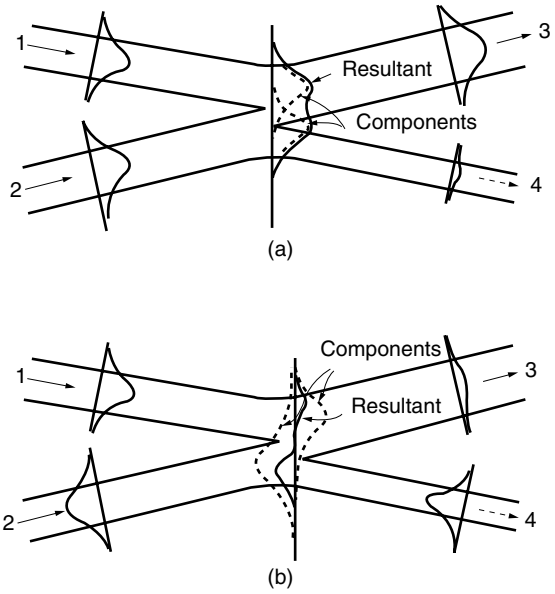


Figure 10.17 Optical magic T. (a) Arms 1 and 2 are fed in phase. (b) Arms 1 and 2 are fed out of phase.

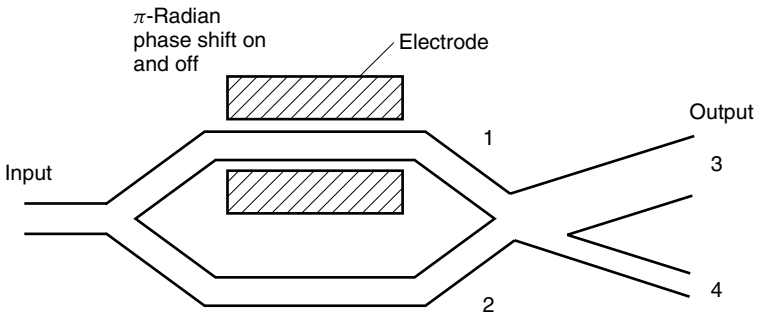


Figure 10.18 Amplitude modulator using the magic T.

Figure 10.17b shows the field distribution when the input arms 1 and 2 are equal in amplitude but in opposite phase (antisymmetric feed). The distribution of the resultant field at the junction is antisymmetric in shape. Arm 4 is excited more than arm 3 because the field in the mouth of arm 3 is far from the shape of the TM_0 mode, whereas the shape of the resultant field at the mouth of arm 4 conforms with the fundamental mode in arm 4.

Neither symmetric nor antisymmetric feed will completely turn on or off the output arms; however, this device is still useful in applications that do not require complete switching. Figure 10.18 shows an example of the application of the magic T to an optical modulator. The modulator consists of the magic T and a Mach–Zehnder interferometer. The Mach–Zehnder interferometer is made of electrooptic material and the phase of one of the arms can be controlled by an external electric field. By

modulating the phase of the control arm between 0 and π radians, each output light of arm 3 and 4 is amplitude modulated.

10.8 ELECTRODE STRUCTURES

In order to apply an external electric field to an electrooptic waveguide in the desired direction, the electrodes have to be properly designed. In this section, various electrode configurations are summarized. Configurations for bulk waves as well as for waveguides are included.

10.8.1 Laminated Electrodes

Electrodes are normally laminated by three metal layers, consisting of top, stopper, and contact layers. Figure 10.19 shows a cutout view of the lamination of an electrode (F. Saito, *private communication*).

1. The top layer is almost always gold because of the ease with which gold connecting wire is bonded to the electrode.
2. A chemical reaction would take place on the surface of some substrate materials if gold were in direct contact. GaAs is especially noted for reaction with gold. Metal for the contact layer (bottom layer) has to be chemically inactive and possess good affinity and low contact resistance to the surface of the substrate. Ti, Ni, Sn, Zn, and Cr are good candidates for this layer.

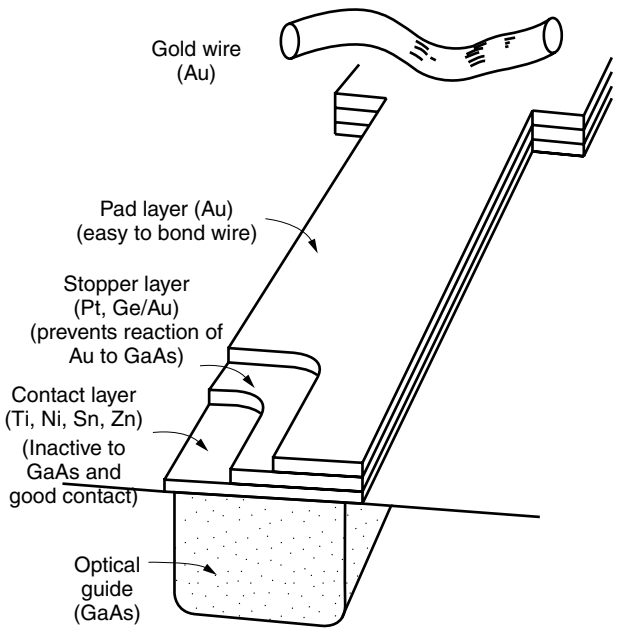


Figure 10.19 Electrode lamination.

3. The stopper layer, which is the middle layer, has to completely shield both the contact layer and the surface of the substrate from chemical reaction with the top gold layer. The criteria for selecting the metal for the stopper layer are that the metal must have a high conductivity and a good affinity to both top and bottom metal layers. Pt and an alloy of Au and Ge satisfy these criteria and are most often used as the material for the stopper layer.

The most common combinations of metals for lamination are Au–Pt–Ti and Au–Ge/Au–Ni.

A single layer of Au or Al is often used as a temporary electrode in a laboratory. The Al, however, can be used only when low electrical current density is drawn from the electrode because of its poor electrical contact.

When a metal electrode is deposited directly over the surface of the guide, not only does the optical transmission loss increase, but the effective index of refraction of the optical guide decreases. These influences are larger for the TM mode. With the TM mode, the index of refraction decreases by $(1-2) \times 10^{-4}$ while with the TE mode, the decrease is one-fifth of this value [20].

A thin buffer layer ($\sim$0.3 μm), with an index of refraction lower than the core of the guide, is inserted between the electrode and the optical guide to solve the problem of the decrease in the index of refraction. Common buffer layer materials are MgF_2, Al_2O_3, and SiO_2.

Finally, an indium oxide (In_2O_3) or indium tin oxide (ITO) coating is used when the electrodes have to be transparent.

10.8.2 Electrode Configurations

Figure 10.20 illustrates various electrode configurations for applying external electric fields to electrooptic waveguides. Figures 10.20a and 10.20b are for bulk waves, where the light may propagate in the entire cross section of the crystal. The rest of the configurations are for guided waves, where the light propagates only inside the optical guide.

10.8.2.1 Applying a Longitudinal Field to Bulk Waves

Figure 10.20a shows the electrodes that provide an electric field parallel to the direction of light propagation. Transparent electrodes have to be used. As was pointed out in Example 5.2, the amount of change $\Delta\phi$ in the phase is determined by the applied voltage and is independent of the length of the crystal. This configuration requires a high voltage. With most crystals, the required control voltage is higher than several kilovolts to obtain $\Delta\phi = \pi$ radians of phase shift.

10.8.2.2 Applying a Transverse Field to Bulk Waves

In Fig. 10.20b, the direction of the external field is perpendicular to the direction of light propagation. This time, the amount of phase shift $\Delta\phi$ can be increased by choosing a longer crystal. However, the length of the crystal cannot be increased indefinitely if the external electric field ε is modulated at a high frequency. At higher modulation frequencies, the transit time for the applied electric field ε to travel from one end to the other end of the electrodes becomes important. For instance, the direction of the ε field at one end can be opposite to that at the other end because of the long transit time,

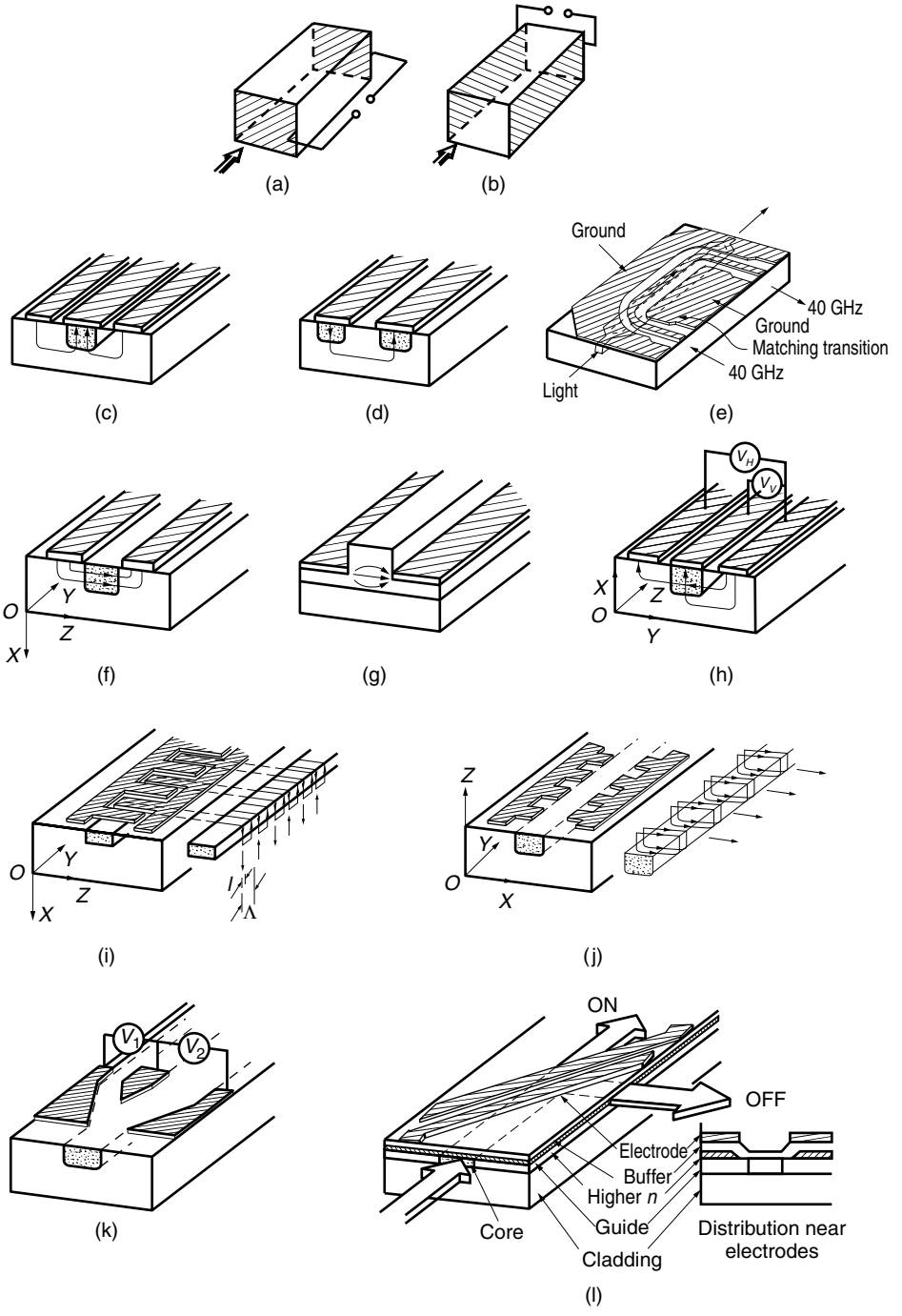


Figure 10.20 Various electrode configurations on electrooptic waveguides. (a) Bulk longitudinal field. (b) Bulk transverse field. (c) Vertical field in an embedded guide. (d) Differential field. (e) Velocity matched Mach–Zehnder interferometer. (f) Horizontal field in an embedded guide. (g) Horizontal field in a rib guide. (h) Horizontal and vertical fields. (i) Periodic vertical field. (j) Periodic horizontal field. (k) Trimming electrodes. (l) Switching electrodes.

and the efficiency of modulation decreases. This, however, can be avoided by treating the electrodes as a pair of transmission lines, and designing the phase velocity of the modulating signal ε to match with that of the light inside the crystal. In this way, the propagating light always sees the same ε throughout transmission in the crystal. The electrodes should be terminated by a proper impedance so that there is no standing wave creating periodic zero fields and decreasing the efficiency.

All the electrodes in the following configurations are for guided waves.

10.8.2.3 Vertical Field in an Embedded Guide

Figure 10.20c shows an electrode layout for applying a vertical field to an embedded guide. Since the electrical lines of force are always perpendicular to the surface of a conductor, the vertical field is obtained by depositing one electrode immediately above the optical guide, and the other electrode is split into two strips, one on each side of the center electrode.

In order to minimize the transmission loss and lowering of the effective index of refraction due to the direct contact of metal with the optical guide, a lower-refractive-index buffer layer has to be inserted. This is true with any of the electrodes for the vertical field, and as mentioned above, the necessity for the buffer layer is greater when the optical guide is excited by the TM mode.

10.8.2.4 Vertical Field in Adjacent Embedded Guides

Figure 10.20d shows a pair of electrodes that provides a vertical electric field of opposite polarity to two adjacent embedded guides. The pair of electrodes is used to cause maximum imbalance between the effective indices of refraction of the two guides. The electric lines of force leave in a downward direction from one of the electrodes and return in an upward direction to the other electrode. The directions of the electric fields in the two optical guides are opposite; one creates Δn while the other creates $-\Delta n$, and the maximum differential in the indices of refraction is achieved.

The Mach–Zehnder interferometer is an example of a device that often uses this electrode configuration. In a Mach–Zehnder interferometer, light traveling initially along a single path is split into two separate paths and then later recombined into a single path again. Any factor that affects the two separate paths differently will result in a change in the interference of the recombined light.

10.8.2.5 Velocity Matched Mach–Zehnder Interferometer

A traveling-wave type modulator is shown in Fig. 10.20e [21]. A coplanar waveguide (two-dimensional coaxial cable) is deposited over the Mach–Zehnder optical interferometer substrate. The phase velocity of the microwave signal propagating along the coplanar waveguide is matched with that of the light in the Mach–Zehnder interferometer.

As mentioned in Section 10.8.2.2, the synchronization of these two velocities achieves not only a high modulation efficiency but also a wideband frequency modulation.

10.8.2.6 Horizontal Field in an Embedded Guide

An electrode arrangement that can provide a horizontal (parallel to the top surface of the guide) electric field inside an embedded guide is shown in Fig. 10.20f. The electric lines of force are almost horizontal but there is always some deviation from

a truly horizontal field. A buffer layer is not needed because the electrodes are not in direct contact with the surface of the guide. The maximum voltage that can be applied between the electrodes is limited by either the arc breakdown voltage between the electrodes (See Problem 10.5) or crystal damage.

10.8.2.7 Horizontal Field in a Rib Guide

Figure 10.20g shows the electrodes for applying a horizontal electric field to a rib optical guide. Compared to Fig. 10.20f, the configuration in Fig. 10.20g has a higher upper limit on the applied voltage, a more closely horizontal electric field distribution, and less field divergence, but the configuration is applicable only to a rib guide.

10.8.2.8 Horizontal and Vertical Fields

When more than a few elements of the matrix of the Pockels constants are used, applied fields may be required in both the horizontal and vertical directions. Figure 10.20h shows three electrodes that provide both vertical and horizontal fields simultaneously and yet each can be controlled independently to make sophisticated use of the properties of the crystal anisotropy.

10.8.2.9 Periodic Vertical Field

Figure 10.20i shows the finger electrodes that provide periodic electric fields along the path of light propagation. Electric lines of force start perpendicularly to the surface of a tooth and terminate similarly at the adjacent tooth. The lines of force around the optical guide are shown in the side view of the figure indicating the direction of polarization of the applied electric field.

Such finger electrodes are used in a mode converter [22], a device that converts a TM mode to a TE mode or vice versa. The performance of the converter is very much wavelength dependent and the output is present only around the narrowly designed wavelength. Because of this property, the same device is also used as a wavelength filter.

The fundamental TM mode is vertically polarized, and the fundamental TE mode is horizontally polarized. Hence, the function of the TM to TE mode converter is to convert from vertical to horizontal polarization. Examples 7.13 and 10.4 provide the design details for such a mode converter.

In order to achieve a high efficiency of conversion, the tooth pattern has to be repeated. The maximum conversion is achieved when many teeth are arranged with a period such that the TE wave generated from each tooth adds up in phase at the exit of the converter.

10.8.2.10 Periodic Horizontal Field

Another type of finger electrode pattern is shown in Fig. 10.20j. These electrodes provide periodic electric fields, but in the horizontal direction. At each region where the distance between opposing teeth is the shortest, horizontal electric lines of force are set up. These finger electrodes are also used for mode converters and wavelength filters based on the same principle as described for the periodic vertical field device. As mentioned earlier, the advantage of the horizontal field electrodes is that the electrodes are not in contact with the surface of the guide, and the transmission loss is small.

10.8.2.11 Trimming Electrodes

In Fig. 10.20k, trimming electrodes have been placed at the location where one guide branches into two separate guides. In applications where splitting the light into the

two branch guides has to meet exact specifications, fabrication tolerances can be relaxed significantly if there is a provision for electronic tuning. For instance, in a Mach–Zehnder modulator, the amplitudes of the input to the two arms have to be exactly the same in order to achieve l00% modulation. Without electronic tuning, fabrication tolerance is severe. However, with the use of trimming electrodes, small fabrication faults can be corrected electronically. An imbalance in the applied voltages to the electrodes causes a small imbalance in the division of the light power into the two arms to correct for imperfect fabrication.

The earlier example of Fig. 10.20h gave an electrode configuration for applying horizontal and vertical fields. For some applications, one of the applied fields is used as the primary field, while the other is used in a trimming capacity.

10.8.2.12 Switching Electrodes

A pair of electrodes such as shown in Fig. 10.20l is used for switching the optical beam in a slab guide [23]. The switching action is based on total internal reflection. The index of refraction in the region between the electrodes is decreased by applying an external electric field so that total internal reflection takes place.

Realizable decreases in the index of refraction are small, which means that the critical angle is necessarily close to 90° so that the reflected and transmitted beams will be quite close to one another. In order to alleviate this problem, the effective index of refraction of the core slab is raised everywhere except in the region between the electrodes. This is done by first depositing a higher-index medium such as GaAs over the core, except at the location between the electrodes. In the region between the electrodes where no higher-index material was deposited, the effective index of the core dips. When an external field is applied, the dip deepens, total internal reflection takes place, and the beam is switched to the off axis.

10.9 MODE CONVERTER

Mode converters are used to convert a TM mode to a TE mode or vice versa. The Poincaré sphere interpretation of the operation of a mode converter was given in Example 7.13. In the mode converter example below, parameters for optimum conversion are explored using the indicatrix.

Example 10.4 A TE $\leftrightarrow$ TM converter is fabricated, using the configuration shown in Fig. 10.20i. A lithium niobate slab with its crystal axis in the Z direction is used.

(a) Find the cross-sectional ellipse of the indicatrix in the region where an X-directed electric field is applied by a pair of fingers.
(b) Consider a single pair of finger electrodes of length l, and let a vertically polarized TM mode be incident. Find the expression for the horizontally polarized TE mode at the exit of the finger electrode region.
(c) What is the spacing Λ between the adjacent fingers for the optimum conversion?

Solution

(a) The material is $LiNbO_3$ and Eq. (5.10) applies. Since light propagates in the Y direction, the "cross-sectional ellipse" of the indicatrix is obtained from Eq. (5.4) with

$y = 0$ and Eq. (5.10):

$$\frac{X^2}{n_o^2} + \frac{Z^2}{n_e^2} + 2r_{51}\varepsilon_x XZ = 1 \tag{10.51}$$

Equation (10.51) represents an ellipse rotated due to ε_x and the angle θ of the rotation is obtained immediately from Eq. (5.35):

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2r_{51}\varepsilon_x}{1/n_0^2 - 1/n_e^2}\right) \tag{10.52}$$

The lengths of the major and minor axes of the ellipse are represented in the rotated coordinates by

$$\frac{x'^2}{(N - \Delta N)^2} + \frac{z'^2}{(N + \Delta N)^2} = 1 \tag{10.53}$$

where

$$\begin{aligned} \Delta N &= \frac{1}{2}N^3\sqrt{B^2 + d^2} \\ \frac{1}{N^2} &= \frac{1}{2}\left(\frac{1}{n_o^2} + \frac{1}{n_e^2}\right) \\ d &= \frac{1}{2}\left(\frac{1}{n_o^2} - \frac{1}{n_e^2}\right) \\ B &= r_{51}\varepsilon_x \end{aligned} \tag{10.54}$$

(b) Since the cross-sectional ellipse is rotated by θ, the allowed directions of polarization are at θ and $\theta + \pi/2$. It is this rotation that is used for the mode conversion. First, the amplitude E_{TM} of the incident vertically polarized TM mode is decomposed into the θ and $\theta + \pi/2$ directions and then each of these components propagates for the length l of a finger. At the exit of the finger, the horizontal components of each of the allowed polarizations are found and combined to give the desired field. Referring to Fig. 10.21a the components in the θ and $\theta + \pi/2$ directions are

$$E_{X'} = E_{\mathrm{TM}}\sin\theta \quad \text{in the } X' \text{ direction}$$

$$E_{Z'} = E_{\mathrm{TM}}\cos\theta \quad \text{in the } Z' \text{ direction}$$

After each wave propagates a length l, $E_{X'}$ and $E_{Z'}$ become

$$E_{X'} = E_{\mathrm{TM}}\sin\theta\ e^{j(N-\Delta N)kl}$$

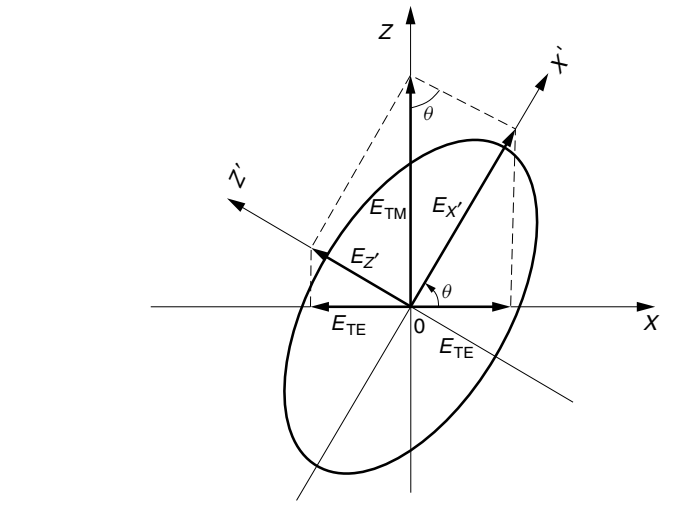
$$E_{Z'} = E_{\mathrm{TM}}\cos\theta\ e^{j(N+\Delta N)kl}$$

The resultant horizontal component E_{TE} becomes

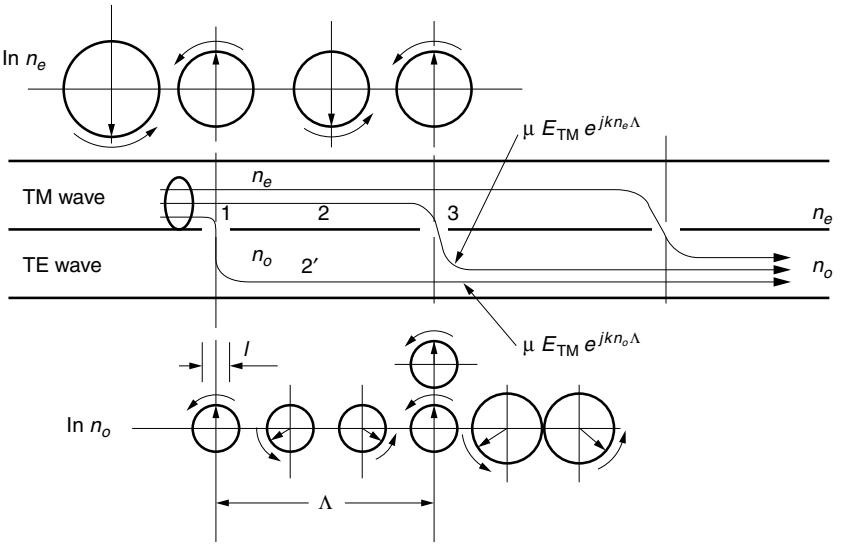
$$\begin{aligned} E_{\mathrm{TE}} &= E_{\mathrm{TM}}\left(e^{j(N-\Delta N)kl} - e^{j(N+\Delta N)kl}\right)\sin\theta\cos\theta \\ &= -jE_{\mathrm{TM}}e^{jNkl}\sin(\Delta N \cdot kl)\sin 2\theta \\ E_{\mathrm{TE}} &= \mu E_{\mathrm{TM}} \end{aligned} \tag{10.55}$$

where

$$\mu = -je^{jNkl}\sin(\Delta N \cdot kl)\sin 2\theta \tag{10.56}$$



(a)



(b)

Figure 10.21 Mode converter principle. (a) Rotation of the elliptical cross section of the indicatrix due to ε_x. (b) Coupling from the TM wave to the TE wave.

Thus, μ is the quantity associated with the amount of conversion in one finger pair. It increases with l (as long as $\Delta Nkl < \pi/2$), and it increases as θ approaches $\pi/4$ rad.

(c) The optimum spacing Λ between the fingers will now be found. For the sake of simplicity, let us assume that the length l is much shorter than the spacing Λ and let us use the fictitious model shown in Fig. 10.21b for explanation. Two guides are coupled through a series of holes; the TM wave is incident in the top guide and the TE mode is

generated in the bottom guide, even though in reality both propagate in the same guide. Each hole represents a finger electrode pair. The TM and TE modes see a different index of refraction. If the angle of the zigzag propagation in the guide is very small (or the fundamental mode is far away from the cutoff), the propagation constant of the TM mode is approximately $n_e k$ and that of the TE mode is $n_0 k$ in the region with $\varepsilon_x = 0$. The next question is how to maximize the emergent TE wave, which is the sum of all the waves leaked from the coupling holes (fingers). Since the propagation constants of the TM and TE modes are different, the phases of the contributing waves depend on which hole they have come from, and the spacings between the holes.

Compare the phase of the leaked component of the TE mode that leaked at hole 1 and took the route of $2'$ to reach hole 3, with the phase of the component of the TE mode that leaked at hole 3 and joined the TE mode. Taking into account the difference in the propagation constants in the upper and lower guides, the TE component taking the path of $1{-}2'{-}3$ is

$$\mu \mathrm{TE}_{\mathrm{TM}} e^{jkn_0\Lambda}$$

while that taking the path $1{-}2{-}3$ is

$$\mu \mathrm{TE}_{\mathrm{TM}} e^{jkn_e\Lambda}$$

The spacing Λ of the holes that makes these two waves in phase is

$$k(n_o - n_e)\Lambda = 2\pi \tag{10.57}$$

With this choice of the spacing Λ, the outputs from the N holes will add most efficiently, increasing the output N times.

For $\lambda = 1.3$ μm and $LiNbO_3$, the optimum spacing is

$$\Lambda = 15.1 \text{ μm} \tag{10.58}$$

If this condition is satisfied, the amplitude of the TE wave from N holes becomes

$$E_{\mathrm{TE}} = \mu N E_{\mathrm{TM}} \tag{10.59}$$

□

PROBLEMS

10.1 Find the number of possible modes of the ridge guide shown in Fig. P10.1. The TM-like mode is excited with physical parameters

$$n_1 = 1.55, \quad n_2 = 1.54, \quad 2d = 5.7 \text{ μm}, \quad 2w = 6.0 \text{ μm}, \quad \lambda = 1.3 \text{ μm}$$

10.2 A slab optical guide with a cone-shaped indentation such as shown in Fig. P10.2 was illuminated by a parallel beam. What is the direction of the emergent beam? For what can such a guide be used?

10.3 Two cone-shaped indentations are connected by their tips as shown in Fig. P10.3. For what can such a slab optical guide be used?

10.4 An optical guide was formed on the surface of tetrahedron, *abcd*. The tetrahedron is placed such that the bisect of the bottom face is along the optical axis. The

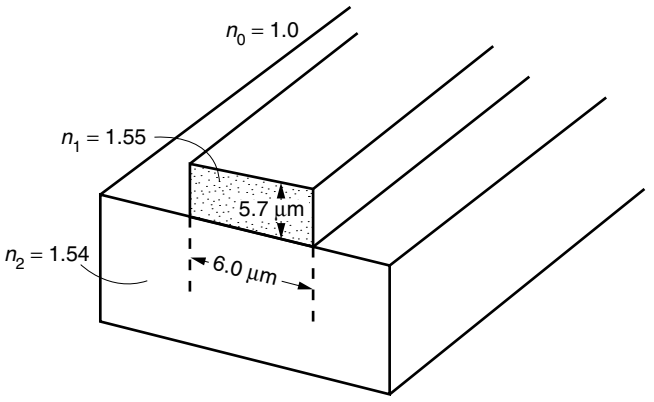


Figure P10.1 Geometry of a ridge guide.

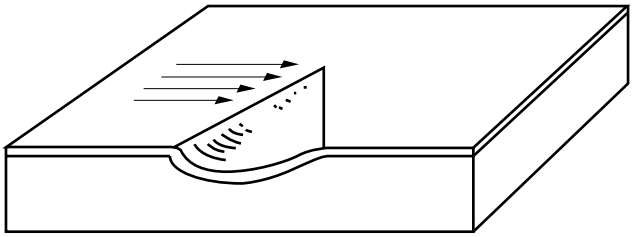


Figure P10.2 A slab optical guide with cone-shaped indentation.

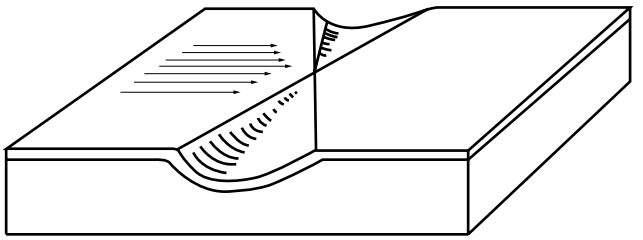


Figure P10.3 A slab optical guide with double cone-shaped indentations.

light path on the surface of the tetrahedron looks like ABC as indicated in Fig. P10.4. Find the location A (length aA) such that the point C of exit is at $dC = 2$ cm. The length of the edge of the tetrahedron is l0 cm. The incident light is parallel to the optical axis.

10.5 The arc breakdown voltage is 30 kV/cm or 3 V/μm in air. What is the breakdown-voltage-limited phase shift $\Delta\phi$ obtainable by a phase shifter such as shown in Fig. 10.20f? The phase shifter is made of X-cut lithium niobate, and the horizontally polarized TE mode is excited with free-space wavelength $\lambda = 1.3$ μm. The length in the Y direction is 1 mm.

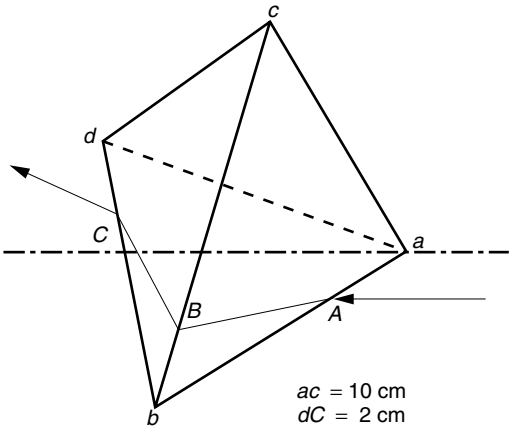


Figure P10.4 Path of conflection.

REFERENCES

1. T. Tamir (Ed.), *Guided-Wave Optoelectronics*, Springer-Verlag, Berlin, 1990.
2. R. G. Hunsperger, *Integrated Optics: Theory and Technology*, 4th ed., Springer-Verlag, Berlin, 1995.
3. A. R. Mickelson, *Guided Wave Optics*, Van Nostrand Reinhold, New York, 1992.
4. Y. Mushiake and M. Kudo, "Multilayered optical slab guide," in *Optical Guided-Wave Electronics*, by the Editorial Committee of the Special Project Research on Optical Guided-Wave Electronics of Japan, Tokyo, 1981, pp. 17–37.
5. E. A. J. Marcatili, "Dielectric rectangular waveguide and directional coupler for integrated optics," *Bell Syst. Tech. J.* **48**(7), 2071–2102 (1969).
6. K. S. Chiang, "Effective-index analysis of optical waveguides," *Proc. SPIE* **2399**, 2–12 (1995).
7. K. S. Chiang, "Analysis of rectangular dielectric waveguides: effective-index method with built-in perturbation correction," *Electron. Lett.* **28**(4), 388–390 (1992).
8. G. T. di Francia, "Conflection doublets," *J. Opt. Soc. Am.* **45**(8), 621–624 (1955).
9. H. P. Zappe, *Introduction to Semiconductor Integrated Optics*, Artech House, Boston, 1995.
10. A. G. Hallam, I. Bennion, and W. J. Stewart, "Photochromic stripe waveguide for integrated optics," in *First European Conference on Integrated Optics*, Oct. 14–15, pp. 26–28, 1981.
11. J. A. Besley, J. D. Love, and W. Langer, "A multimode planar power splitter," *J. Light wave Technol.* **16**(4), 678–684 (1998).
12. J. Leuthold and C. H. Joyner, "Multimode interference couplers with tunable power splitting ratios," *J. Lightwave Technol.* **19**(5), 700–707 (2001).
13. A. E. White, "Wavelength division multiplexing (WDM) has arrived. Critical to future system developments are integrated optical components fabricated using silicon optical bench technology," *Optics and Photonics News*, 27–30, March 2000.
14. H. Takahashi, S. Suzuki, K. Kato, and I. Nishi, "Arrayed-waveguide grating for wavelength division multi/demultiplexer with nanometre resolution," *Elec. Lett.* **26**(2), 87–88 (1990).
15. Y. Inoue, H. Takahashi, S. Ando, T. Sawada, A. Himeno, and M. Kawachi, "Elimination of polarization sensitivity in silica-based wavelength division multiplexer using a polyimide half waveplate," *J. Lightwave Tech.* **15**(10), 1947–1957 (1997).

16. K. Okamoto and A. Sugita, "Flat spectral response arrayed-waveguide grating multiplexer with parabolic waveguide horns," *Electron. Lett.* **32**(18), 1661–1662 (1996).
17. W. Döldissen, H. Heidrich, D. Hoffmann, H. Ahlers, A. Döhler, and M. Klug, "Integrated optical 8 × 8 scrambler in integrated optics," in *Proceedings of the 3rd European Conference ECIO 1985*, May 6–8, Berlin, Germany, pp. 225–228, 1985.
18. M. Izutsu, A. Enokihara, and T. Sueta, "Optical-waveguide hybrid coupler," *Opt. Lett.* **7**(11), 549–551 (Nov. 1982).
19. P. Sewell, T. M. Benson, T. Anada, and P. C. Kendall, "Bi-oblique propagation analysis of symmetric and asymmetric Y-junctions," *J. Lightwave Technol.* **15**(4), 688–696 (1997).
20. M. Masuda and J. Koyama, "Effects of a buffer layer on TM modes in a metal-clad optical waveguide using Ti-diffused $LiNbO_3$ C-plate," *Appl. Opt.* **16**(11), 2994–3000 (1977).
21. W. K. Burns, M. M. Howerton, R. P. Moeller, R. Krähenbühl, R. W. McElhanon, and A. S. Greenblatt, "Low drive voltage, broad-band $LiNbO_3$ modulators with and without etched ridges," *J. Lightwave Tech.* **17**(12), 2551–2555 (1999).
22. R. C. Alferness, "Electrooptic guided-wave device for general polarization transforms," *IEEE J. Quantum Electron.* **QE-17**(6), 965–969 (1981).
23. J. Koyama, H. Nishihara, M. Haruna, and T. Suhara, "Optical guide switches and their applications," in *Optical Guided-Wave Electronics*, by Editorial Committee of the Special Project Research on Optical Guided-Wave Electronics of Japan, Tokyo, 1981, pp. 419–431.

11

MODES AND DISPERSION IN OPTICAL FIBERS

John Tyndall demonstrated in 1870 in front of an audience of the Royal Academy of London that light can travel within a curved jet stream of water from a hole made on the side of a water pail. Propagation was based on repeated total internal reflection at the air–water boundary. However, it was not until 1966, when Kao and Hockham [1] promoted the use of a glass fiber as the transmission medium, that the doors were opened to using the principle of total internal reflection as a viable means of communication. What followed was an intensive development of the purification process of glass materials to reduce transmission loss. Concurrently, improvements were realized in the design and control of critical parameters such as the dimensions and indices of refraction of the fiber layers.

Light propagates in an optical fiber by repeated total internal reflection at the core–cladding glass boundary. Optical fibers have a phenomenally large capacity to carry information and are able to deliver this information to extremely distant locations. Fiber-optic communication systems possess such advantages as:

- Low transmission loss
- Large capacity of information transmission
- No electromagnetic interference
- Lighter weight than copper
- No sparks even when short-circuited
- Higher melting point than copper
- Practically inexhaustible raw material supply

On the other hand, the disadvantages are:

- Connections and taps are more difficult to make than for copper wire.
- Fiber is not as flexible as a copper wire.

Among the advantages, low transmission loss and large capacity of information transmission are the two most important. The transmission loss of highly purified glass has become as low as 0.15 dB/km at $\lambda = 1.55$ μm. Considering the fact that an ordinary glass window just a few centimeters thick cuts the intensity of transmitted light in half, the fiber's loss is unbelievably low. A copper coaxial cable has a loss of approximately 10 dB/km at 50 MHz, so that a significant increase in the span between repeater stations has become possible using fiber-optic cables. In terms of frequency bandwidth, the first copper submarine cable of TAT 1 (Trans Atlantic Transmission) could send 64 voice channels in 1956 but now the submarine fiber-optic link around the globe (FLAG), which links Great Britain and Japan by way of the Middle East and South East Asia, can send 120,000 voice channels [2]. The undersea cables of TyCom Global Network (TGN) will connect major world cities via the Atlantic and Pacific Oceans and the Mediterranean Sea with a transmission capacity of over ten million voice channels [3].

11.1 PRACTICAL ASPECTS OF OPTICAL FIBERS

In this section, practical aspects of optical fibers that are necessary for designing an optical communication system will be explained. The emphasis here is on understanding concepts through simple arguments, leaving the more rigorous and quantitative expressions to a later part of this chapter.

11.1.1 Numerical Aperture of a Fiber

An optical fiber consists of core glass of higher index of refraction in the center, surrounded by cladding glass of lower index of refraction. As shown in Fig. 11.1, light traverses the core glass by *repeated total internal reflection at the boundary between the core and cladding glasses*. There are two paths of propagation in a fiber. One is a zigzag path in a plane containing the fiber axis as shown in Fig. 11.1a. A ray that takes this path is called a *meridional ray*. The other is a helical path around the center of the fiber as shown in Fig. 11.1b. A ray that takes this path is called a *skew ray*.

Since transmission of light relies on total internal reflection, there exists a maximum acceptance angle. Light rays whose angles of incidence exceed this limit do not undergo

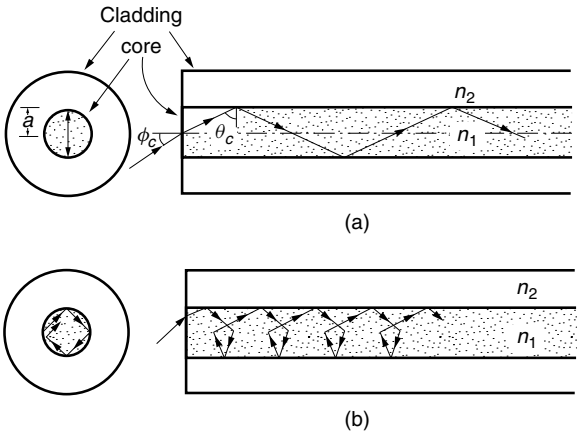


Figure 11.1 Rays in an optical fiber. (a) Meridional ray. (b) Skew ray.

total internal reflection. Referring to Fig. 11.1a, the critical angle at the core–cladding interface is

$$n_1 \sin\theta_c = n_2 \tag{11.1}$$

where n_1 and n_2 are the refractive indices of the core and cladding glasses. The corresponding cutoff incident angle ϕ_c to the end of the fiber is, from Snell's law,

$$n_0 \sin\phi_c = n_1 \sin(90^\circ - \theta_c) \tag{11.2}$$

Combining Eqs. (11.1) and (11.2) gives

$$n_0 \sin\phi_c = \sqrt{n_1^2 - n_2^2} \tag{11.3}$$

Equation (11.3) specifies the maximum incident angle to the end face of a fiber. The quantity $n_0 \sin\phi_c$ is called the *numerical aperture* (NA), and

$$\text{NA} = \sqrt{n_1^2 - n_2^2} \tag{11.4}$$

For example, the NA for a fiber with $n_1 = 1.454$ and $n_2 = 1.450$ is NA $= 0.11$, which means that the angle of incidence can be no larger than 6.3°.

When the light is coupled to the fiber using a lens, not only the angle of incidence of the converging beam but also the diameter of the incident light beam is restricted. The diameter of the converging incident light beam has to be equal to or smaller than the core diameter for efficient coupling of light to the fiber.

11.1.2 Transmission Loss of Fibers

Figure 11.2 shows the transmission loss (in dB/km) of a silica glass optical fiber with respect to wavelength λ (in μm) [4]. The general shape of the calculated loss curve is a V shape with its minimum at $\lambda = 1.55$ μm. In the vicinity of 1.55 μm, the measured loss curve closely follows the theoretically calculated attainable limit. The region of the measured loss curve from 1.0 to 1.5 μm deviates somewhat from the calculated curve, and local minima are observed at $\lambda = 1.15$ and 1.3 μm. On the left side of the V in Fig. 11.2, Rayleigh scattering is the prime contributor to the loss curve. This scattering is caused by irregularities in the fused silica glass as well as irregularities associated with the dopant used to raise the index of refraction of the core glass. No matter how slowly the glass fiber is cooled, when the glass solidifies, 1–10-nm-diameter irregularities in the index of refraction are incurred, and these act as sources of scattering. Scattering from a particle much smaller than the wavelength of the light is known as *Rayleigh scattering*. Rayleigh scattering is proportional to λ^{-4}, so that the amount of Rayleigh scattering decreases rapidly as λ increases.

The amount of Rayleigh scattering loss can be reduced by decreasing the dopant concentration. In fact, a fiber with a pure silica core and flourine-doped cladding has lower loss than a fiber with a GeO_2-doped core.

An additional loss on the left-hand side of the V curve is the tail of the electronic transition of fused silica, which occurs at $\lambda = 0.1$ μm in the ultraviolet region. The amount of this loss is an order of magnitude smaller than the Rayleigh scattering loss.

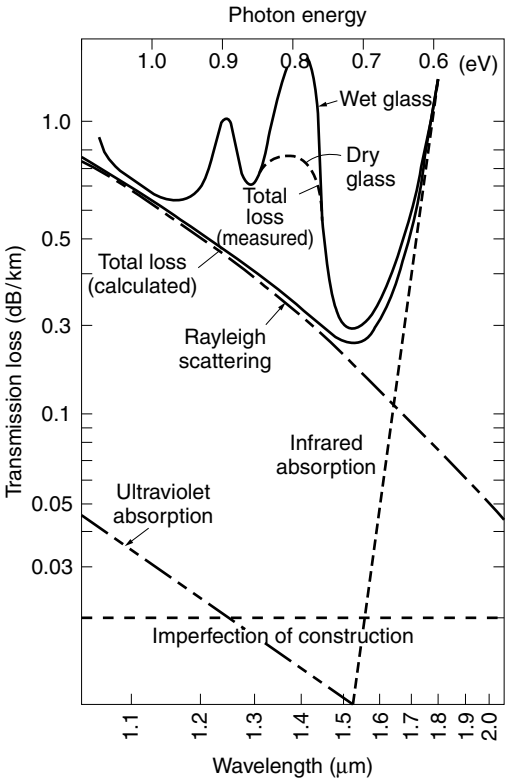


Figure 11.2 Transmission loss with wavelength. (After T. Miya et al. [4].)

Next, the loss in the right-hand side of the V curve will be explained. The right-hand side of the V curve represents the tail of the infrared absorption at $\lambda = 10$ μm. This absorption at $\lambda = 10$ μm coincides with a vibrational transition of the Si—O lattice in the glass.

Lastly, the peak appearing at $\lambda = 1.38$ μm in the measured loss curve is due to the presence of OH ion impurities. The increase in loss at $\lambda = 1.38$ μm occurs because of the second harmonic of the OH ion absorption whose fundamental is located at $\lambda = 2.72$ μm. Absorption by the OH ion is strong, and even small amounts create a significant increase in the fiber loss. In order to remove this loss, the OH ion content has to be reduced to less than 0.1 ppb (parts per billion, 10^{-9}).

11.1.3 Loss Increase Due to Hydrogen and Gamma-Ray Irradiation

It was found that the transmission loss of an optical fiber slowly increases when it is placed in a hydrogen atmosphere. The test results at an elevated temperature are shown in Fig. 11.3a and b. The loss increase is observed over the entire wavelength range, with its maximum at $\lambda = 1.38$ μm. Even though the process is slow at room temperature, such deterioration threatens the life expectancy of fiber-optic communication systems, which is normally 25 years.

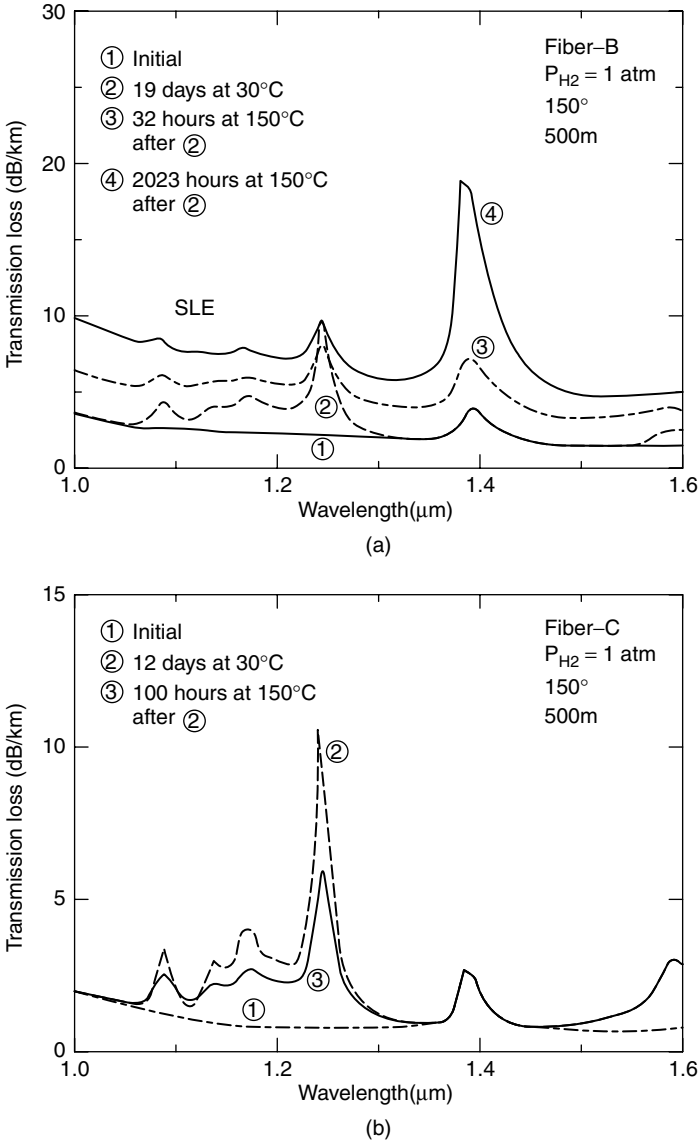
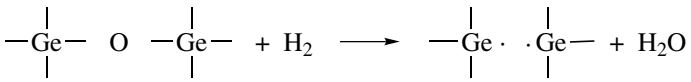


Figure 11.3 Loss increase due to hydrogen exposure for (a) a GeO_2-doped core fiber, and (b) pure SiO_2 core fiber. (After M. Kuwazuru et al. [6].)

There are two types of deterioration: one is reversible and the other is irreversible. The former does not involve a chemical reaction but the latter does. A large class of solid materials will absorb gas from the surface, albeit very small amounts of gas, by diffusion of the gas into the solid. Glass is not an exception. Vibration of the absorbed hydrogen (H_2) gas contributes to the increase in the transmission loss. Such loss, however, is removed when the fiber is removed from the hydrogen atmosphere and then its temperature is raised. The hydrogen gas boils out. This loss is still a problem as long as the fiber has to operate in the hydrogen atmosphere. Some of the

increases in loss caused by the hydrogen gas cannot be removed. These are irreversible deteriorations because they involve chemical reactions between hydrogen and the fiber core. Any chemical reaction resulting in hydroxyl (OH) formation increases the loss at $\lambda = 1.38$ μm. Another irreversible effect is the "short-wavelength-loss edge" (SLE) [5].

The monotonic increase toward short wavelengths is due to SLE. Because of the partial reduction,



an oxygen vacancy is created. Such a defect has a strong absorption line at 215 nm. The SLE is considered to be the tail of this strong absorption. Even when the same fiber ingredients are used, the amounts of OH ion loss and SLE vary, and it is thought that both loss mechanisms have a lot to do with defects of the molecular network [6].

The degree of increase in loss also depends on the kind of dopant in the core. Cores doped with P_2O_5 suffer the highest loss, followed by GeO_2. The best is the pure silica core with fluorine-doped cladding. This fiber is practically immune to hydrogen deterioration [7] because it has the least defects in the molecular network.

The major source of hydrogen is the evolution of hydrogen from organic cable materials. Silicone resin evolves a large amount of hydrogen due to the excess organohydrogen polysiloxane monomer used as a hardener [8]. The amount of hydrogen evolution of UV-light-cured resin is about one-hundredth that of H_2-rich silicone resin, and one-half that of H_2-lean silicone resin.

Another important factor that determines the hydrogen loss is the permeability of hydrogen gas through the cabling material. Even though hydrogen may be evolved from the coating material, if hydrogen escapes from the cabling material into the atmosphere, then the hydrogen pressure, which is directly related to the degree of deterioration, is low, and no problem arises. Coatings of laminated aluminium polyethylene (LAP) have to be properly permeable to hydrogen to prevent the buildup of hydrogen pressure inside the cable.

Next, transmission loss due to gamma-ray (γ-ray) irradiation will be explained. The effects of γ-ray irradiation can be significant if the optical fiber is used within the premises of a nuclear reactor, or in any environment where there is a possibility of exposure to γ rays, such as outer space or deep-sea applications.

When γ rays strike the Si—O lattice, some atoms may become dislodged from the lattice, creating dislocation spots. The dislocation displays a strong absorption at 215 nm with a long tail [9], causing SLE. Figure 11.4 shows the attenuation of light at $\lambda = 1.3$ μm when the fiber is irradiated by 10^5 rads/h of γ-ray irradiation. As soon as irradiation starts, the attenuation increases, and as soon as irradiation stops the fiber begins to recover. Figure 11.4 also shows that a fiber with a pure silica core is less susceptible to the irradiation and has superior recovery characteristics compared to a GeO_2-doped fiber.

It is known that the effect of the γ-ray irradiation can be lessened by introducing some H_2O in the core glass, at the cost of a significant increase in the transmission loss. Figure 11.5 shows the attenuation of the fiber with respect to wavelength before and after 10^5 rads/h irradiation for one hour. Even though the attenuation increases at wavelengths shorter than $\lambda = 0.9$ μm, it is practically unchanged at the wavelengths longer than $\lambda = 0.9$ μm.

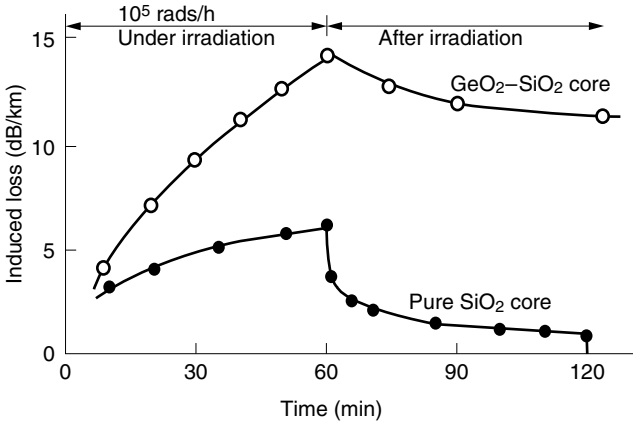


Figure 11.4 Loss increase due to γ-ray irradiation (at $\lambda = 1.3$ μm and 25°C). (After I. Yoshida et al. [9].)

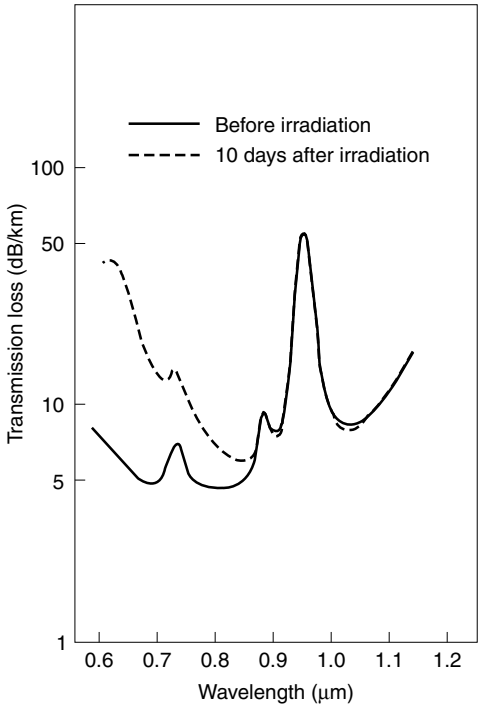


Figure 11.5 Transmission loss of a fiber containing H_2O before and after irradiation with 10^5 rads/h γ rays for one hour. (After G. Tanaka et al. [7].)

For most applications, increased fiber loss due to radiation is undesirable, but in the biomedical field [10], this property has been used to advantage. A fiber sensor has been constructed for monitoring the radiation dose in close vicinity to the site of the radiation therapy. A short length of radiation-sensitive optical fiber such as lead glass fiber or Ge-doped SiO_2 fiber is connected to a radiation-resistant fiber and is inserted

into the body of the patient. The end of the sensor has a mirror and the radiation dose is found by measuring the reflected light power.

11.1.4 Dispersion

Next to the fiber loss, the capacity of information transmission is an important consideration in designing a fiber-optic communication system. The dispersion of the fiber essentially determines the maximum bit rate or modulation frequency that can be attained. There are three types of dispersion:

1. Mode dispersion.
2. Material dispersion.
3. Waveguide dispersion.

Variation in propagation time among different modes creates mode dispersion. If the source were perfectly monochromatic, then mode dispersion would be the only dispersion with which to contend. In reality, all sources, especially when modulated, emit light over a spread of optical frequencies, and the frequency spread of the source leads to other types of dispersion. The variation in propagation time due to the wavelength dependence of the refractive index creates material dispersion. The wavelength dependence of the propagation pattern causes waveguide dispersion.

11.1.5 Mode Dispersion

In the case of the slab optical guide treated in Chapter 9, the mode pattern was considered as a standing-wave pattern made of two component waves propagating in a zigzag path. The component wave of the lowest order mode propagated almost straight, whereas that of the higher order mode propagated with a larger angle of zigzag. The same argument holds for a step-index optical fiber. When a pulse of light is injected into the fiber end, as shown in Fig. 11.6, the pulse energy is partitioned among all possible discrete angles δ_m, where $\delta_0 \doteqdot 0$ corresponds to the lowest order mode and $\delta_m = \delta_c$ corresponds to the highest order mode. The distance of travel for the lowest order mode is the shortest, while that of the highest order mode is the longest. There is a spread in the arrival time for the pulse energies transmitted in different modes. Such a spread seen in the received pulse is called *mode dispersion*. The time τ needed for the light to travel a unit distance is called the *group delay*, and dispersion

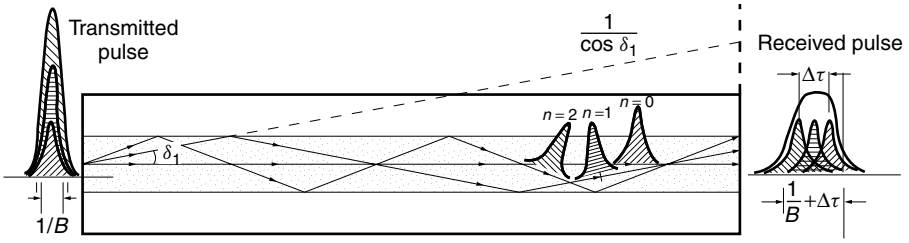


Figure 11.6 Explanation of mode dispersion. Component waves take different zigzag paths. The higher order mode components take a longer time to reach the receiver than the lower order mode components, creating a spread in arrival time.

is expressed as the spread $\Delta\tau$ of the group delay. As shown in Fig. 11.6, the total length of the zigzag propagation for the first mode is identical to the path shown by the dashed line. The spread $\Delta\tau$ due to mode dispersion, therefore, is

$$\Delta\tau = \frac{n_1}{c}\left(\frac{1}{\cos\delta_m} - 1\right) \tag{11.5}$$

where δ_m is the zigzag angle of the highest order mode. The path of the lowest order mode is assumed a straight line. Only meridional rays are considered. Using Snell's law at the entrance, the numerical aperture is expressed as

$$\text{NA} = n_1 \sin\delta_m \tag{11.6}$$

Equations (11.5) and (11.6) combined with the assumption that δ_m is very small gives

$$\Delta\tau = \frac{1}{2cn_1}(\text{NA})^2 \tag{11.7}$$

It is generally accepted that when the spread $\Delta\tau$ of the light pulse becomes wider than 70% of the bit period $1/B$ seconds, where B is the number of bits per second of the digital signal, information transmission becomes unreliable (see Section 16.6.5). From Eq. (11.7) the maximum bit rate of a multimode fiber is approximately

$$B = 1.4\frac{cn_1}{(\text{NA})^2} \tag{11.8}$$

For a multimode fiber with $n_1 = 1.55$ and $n_2 = 1.54$, we have, from Eq. (11.4), NA $= 0.176$. The information capacity for a transmission of one kilometer is $B = 21$ (Mb/s).

Equation (11.8) is valid for a step-index fiber, that is, a fiber whose refractive index profile is a step function. One obvious way to eliminate mode dispersion is to use a fiber in which, only one mode is excited and the mode dispersion is completely removed. Such a fiber is called a single-mode fiber or monomode fiber.

For some applications, single-mode fiber is the preferred choice, despite the more critical alignment tolerances. Just how small is the core of a single-mode fiber? Some calculations are presented here to answer that question. The number of possible modes in a slab guide was determined by the normalized thickness in Section 9.3. Similarly, as will be detailed later, the number of modes in an optical fiber is determined by the normalized radius V of the core

$$V = ka\sqrt{n_1^2 - n_2^2} \tag{11.9}$$

According to the calculation in Eq. (11.107), single-mode propagation is obtained when

$$V < 2.4$$

Equation (11.9) is rewritten using Eq. (11.4) as

$$V = \frac{2\pi}{\lambda}a \cdot (\text{NA}) \tag{11.10}$$

In order to reduce V, there is a choice of either reducing the core radius or the numerical aperture (NA). Choosing too small a value for the radius a makes translational alignment of the source beam too critical. On the other hand, too small a value of NA makes the angular alignment of the source beam too critical. The radius of a single-mode fiber is normally chosen around $a = 4.5–5$ μm.

11.1.6 Material and Waveguide Dispersions

Dispersion of an optical fiber is not limited to mode dispersion. Even a single-mode fiber suffers from two additional dispersions. Associated with the light spectrum is a finite wavelength (frequency) bandwidth, either because of signal modulation or characteristics of the light source, and the wavelength (frequency) dependence of the group delay causes the additional dispersion. Glass materials change their index of refraction with wavelength, as shown in Fig. 11.7 (see boxed note). The variation of the index of refraction causes not only a spread in the group velocity of the component waves but also a variation in the zigzag pattern. Dispersion due to the spread of the index of refraction is called *material dispersion*. Dispersion due to the variation in the zigzag pattern is called *waveguide dispersion*.

Material dispersion, waveguide dispersion, and the sum of the two are all plotted with respect to wavelength in Fig. 11.9. In this figure, the dispersion parameter, which has units of ps/(nm·km), is used as a practical unit for representing the dispersion. This unit means the number of picoseconds of spread in arrival time for a source wavelength spread of 1 nm and a fiber length of 1 kilometer. For example, for a pulse whose carrier frequency is $\lambda = 1.55$ μm, the dispersion parameter D from the graph in Fig. 11.9 is $D = 16.7$. If this pulse has a wavelength band of 1.5 nm and travels 20 km, the expected spread of the arrival time is (16.7) (1.5) (20) = 501 ps.

Material dispersion increases from a negative quantity to a positive quantity with wavelength. At $\lambda = 1.27$ μm, the curve crosses the horizontal axis and the material dispersion becomes zero, leaving only the waveguide dispersion as well as the mode dispersion.

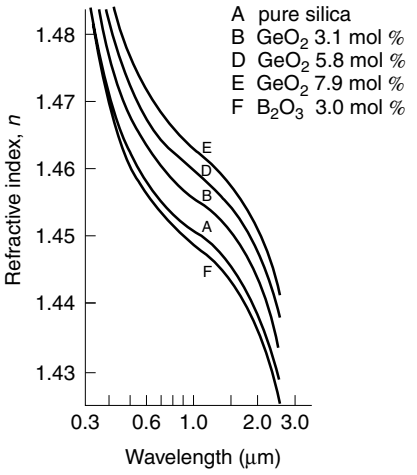


Figure 11.7 Refractive index of doped fused silica and pure fused silica. (After S. Kobayashi et al. [11].)

There is a formula that very nicely fits the wavelength dependence of the index of refraction of pure silica glass [11]:

$$n^2 = 1 + \sum_{i=1}^{3} \frac{a_i \lambda^2}{\lambda^2 - b_i} \tag{11.11}$$

with

$$\begin{aligned} a_1 &= 0.6961663 \qquad b_1 = 0.004629148 \\ a_2 &= 0.4079426 \qquad b_2 = 0.01351206 \\ a_3 &= 0.8974994 \qquad b_3 = 97.934062 \end{aligned} \tag{11.12}$$

This formula is known as *Sellmeier's formula*. As shown in Fig. 11.8, the measured index of refraction of the glass has two regions of steep slope in the range of 0.04–40 μm. The steep slope in the region of 7–15 μm is due to the dispersion caused by vibration of the Si — O lattice in the crystal. The electronic transition bands around 0.1 μm are responsible for the steep slope in the region of 0.07–0.15 μm. Each term

$$\frac{a_i \lambda^2}{\lambda^2 - b_i} \tag{11.13}$$

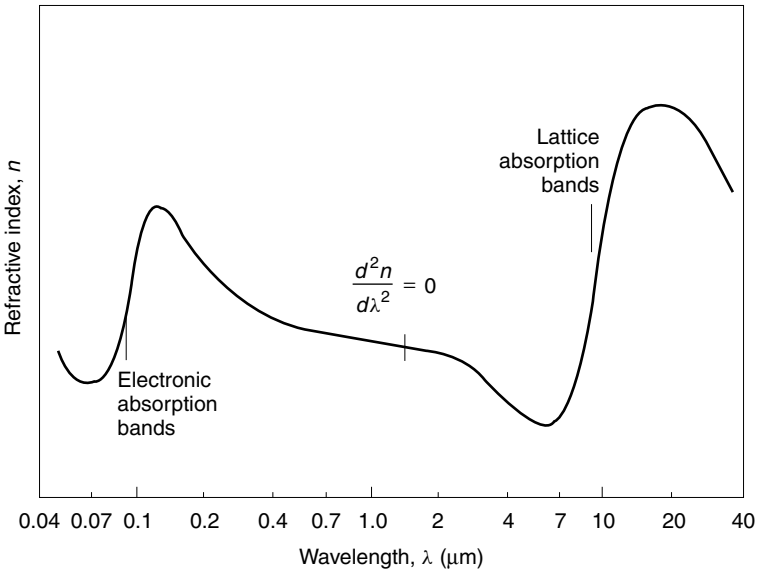


Figure 11.8 Refractive index variation of fused silica. (After S. R. Nagel.)

corresponds to a peak at $\lambda = \sqrt{b_i}$, and the sum of the terms nicely represents the index of refraction, except in very close proximity to $\lambda = \sqrt{b_i}$, where the associated term becomes infinite. Figure 11.7 was obtained by expanding the vertical scale in the region between 0.3 and 3 μm.

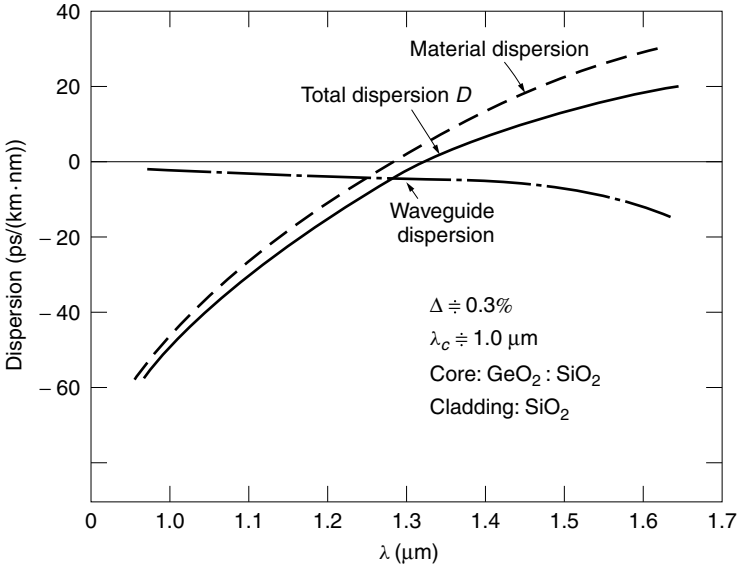


Figure 11.9 Dispersion of a single mode fiber. (After D. B. Keck [12].)

The waveguide dispersion remains a small negative quantity throughout the region shown in Fig. 11.9 [12]. At $\lambda = 1.32$ μm, the waveguide dispersion cancels the material dispersion, and the single-mode optical fiber becomes completely free from these two dispersions.

As seen from Fig. 11.9, the wavelength for dispersion-free transmission can be shifted toward a longer wavelength, more specifically to $\lambda = 1.55$ μm, if the magnitude of the waveguide dispersion is somehow increased. An optical fiber whose dispersion-free wavelength is shifted to the wavelength of lowest loss can have both the merits of lowest loss and no dispersion. Such a fiber is called a *dispersion-shifted fiber* and is discussed further in Section 11.7.

11.1.7 Various Kinds of Optical Fibers

A survey will be made of the various kinds of optical fibers in this section. A brief description of features will be given for each type. A standard cladding diameter of 125 μm is assumed unless otherwise specified.

11.1.7.1 Multimode Step-Index Fiber

The geometry of the multimode step-index fiber is shown in Fig. 11.10a. The core diameter is 50 μm, and the refractive index difference is $\Delta = 0.5–1.0\%$. The normalized radius is $V \doteqdot 30$, and the number of modes is on the order of hundreds. The advantage of this fiber is the ease of coupling to a source or connecting to another fiber because of the large core diameter and large NA value. But this fiber has a limited capacity for information transmission due to mode dispersion and is used primarily for short-distance communication. The capacity (bandwidth–length product) is about 65 (Mb/s)·km with NA = 0.1.

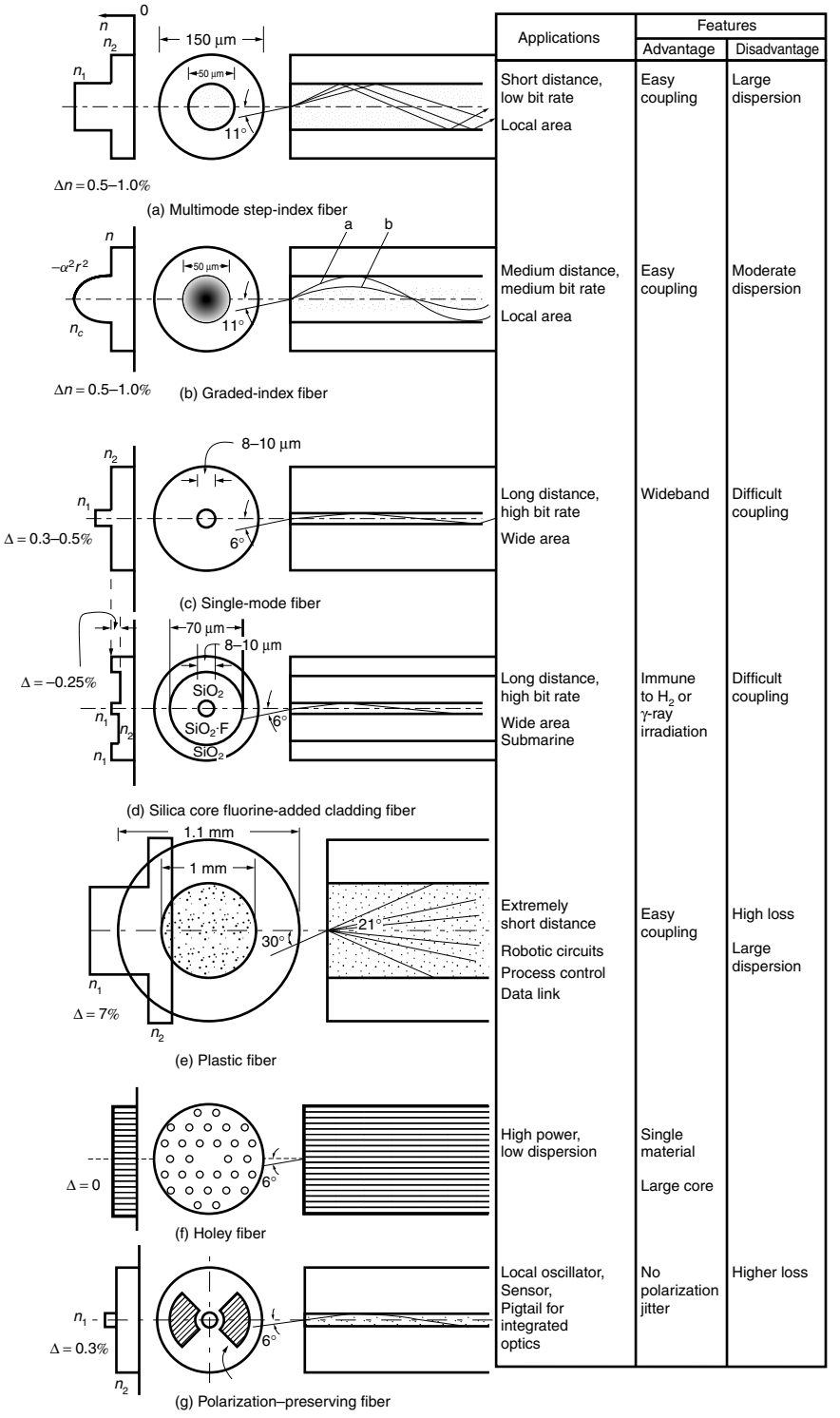


Figure 11.10 Various kinds of optical fibers.

11.1.7.2 Multimode Graded-Index Fiber

A quadratic refractive index distribution is employed as shown in Fig. 11.10b in order to reduce the dispersion. Compare the travel paths of two sinusoidal meridional rays a and b in the figure. Even though the total travel path of ray a is longer, this ray travels mostly in an outer region where the index of refraction is lower and hence the velocity is higher. On the other hand, the travel path of ray b is shorter but it travels in the center region where the index of refraction is higher and hence the velocity is lower. As a result, the difference between the travel times of ray a and b is minimized in the graded index fiber. The dispersion of the graded index fiber is much smaller than that of the multimode step index fiber.

The graded-index fiber retains the coupling ease of the multimode step-index fiber, with the added bonus of a much larger capacity for information transmission, up to 10 (Gb/s)·km. Both the step-index multimode fiber and the graded-index multimode fiber are unsuitable for coherent optical communication systems, or any other system where the phase and state of polarization are important. Because of the hundreds of modes present in these fibers, the phase and state of polarization of the light change in a highly unpredictable manner.

11.1.7.3 Single-Mode Fiber

The geometry of a single mode fiber is shown in Fig. 11.10c. The core diameter is reduced to 8–10 μm, and $\Delta = 0.3–0.5$, so that the normalized radius V of the fiber in Eq. (11.9) is smaller than the cutoff $V = 2.4$ of the mode that is the next higher mode from the dominant mode. The advantage and disadvantage of the single-mode fiber are reversed compared to the multimode fiber. Coupling is difficult in the single-mode fiber, but the information transmission capacity is significantly larger than that of the multimode fiber. Unlike the multimode fiber, phase information can be transmitted. Even though the state of polarization is subject to change during transmission, it can be compensated for at the receiver (see Section 12.9.1). This fiber is suitable for coherent communication.

11.1.7.4 Dispersion-Shifted Fiber

The dispersion-shifted fiber is not only nominally free from dispersion, but also the wavelength of operation is at $\lambda = 1.55$ μm, where transmission loss is least. The dispersion of the fiber is removed by choosing a refractive index distribution such that waveguide dispersion cancels material dispersion at $\lambda = 1.55$ μm.

Transmission loss is about 0.2 dB/km and the practical limit on the capacity of information transmission is about a (Tb/s)·km.

The refractive index distribution is either one large peak of $\Delta = 0.6–0.9$, or a concentric step modulation in the cladding layer. The former is shown in Fig. 11.34 and the latter in Fig. 11.35. The physical dimension of the fiber is more or less the same as that of the single-mode fiber shown in Fig. 11.10c. Dispersion-shifted fibers have a larger bending loss than ordinary single-mode fibers due to less confinement of the light within the core.

11.1.7.5 Silica Core Fluorine-Added Cladding Fiber

The cladding glass is normally pure SiO_2. Germanium dioxide (GeO_2) is usually used to raise the refractive index of the core glass with respect to that of the cladding glass. The inclusion of GeO_2 is an additional inhomogeneity in the core and is deemed to increase the transmission loss of the fiber. If, however, pure SiO_2 is used in the core and the refractive

index of the cladding is lowered by adding fluorine (see Fig. 11.10d), the transmission loss of the fiber is reduced. The transmission loss of such a fiber is 0.154 dB/km at $\lambda = 1.55$ μm.

An even more significant advantage of the silica core fluorine-added cladding fiber is that it is less susceptible to the loss associated with exposure to hydrogen atmosphere or γ-ray irradiation, as already mentioned in Section 11.1.3.

11.1.7.6 Plastic Fiber

This fiber is made out of low-transmission-loss plastic material. Being made of plastic, the diameter of the core can be as large as 1–2 mm. A fiber with such a large diameter would shatter if it were made of glass when it is bent. The plastic fiber not only has a large diameter but also has a large NA value of 0.5, which means $\phi_c = 30°$. Because of the large NA, the plastic fiber can easily be coupled to a light source. The disadvantages of this fiber are a high transmission loss and a large dispersion. It is primarily used for short-distance communication. Figure 11.10e shows the geometry of a plastic fiber. Plastic fiber whose core is made of polymethyl methacrylate (PMMA) and whose cladding is made of fluorinated alkyl methacrylate copolymer is useful in the region of $\lambda = 0.6$–0.8 μm and the minimum transmission loss is 20 dB/km at 0.68 μm [13].

11.1.7.7 Multi-Ingredient Fiber

The melting point of fused silica is around 1900°C. Light doping with one or two of the commonly used dopants does not affect the melting point of fused silica significantly. Glass that is used for fabricating lenses, however, contains many kinds of metal oxides, such as Na_2O, MgO, B_2O_3, Tl_2O, CaO, GeO_2, and Li_2O. The melting point of such glass is lower and is around 1200°C. The lower melting point makes it possible to fabricate an optical fiber by pulling it out of a nozzle at the bottom of a double crucible, where the inner crucible is filled with molten core glass, and the outer crucible is filled with molten cladding glass (see Fig. 11.45). The fiber is made directly from the molten material into a finished fiber, and not only is the fabrication cost lower, but the fiber can be endlessly long. Another merit of this fiber is that the difference in the refractive indices of the core and cladding can be made large and the NA value can be as high as NA = 0.5 for easy coupling to the source.

The transmission loss of such a fiber is at its lowest value of 4 dB/km near $\lambda = 0.8$ μm. Fiber of this kind is used primarily for short-distance optical communication.

11.1.7.8 Holey Optical Fiber (HF)

Holey optical fiber (HF) is a single material fiber with a periodic or aperiodic array of circular [14,15] or elliptical [16] air holes running in the axial direction of the optical fiber. The presence of air holes reduces the effective index of refraction of the material. The effective refractive index difference can be achieved from a single material.

Holey optical fiber is fabricated from a preform. The preform is assembled by stacking hexagonally shaped silica rods whose centers have been bored to produce air holes. The fused stacked silica rods are repeatedly drawn through a furnace tower at 2000°C to achieve the desired air hole dimensions of 0.2–1.2 μm in diameter.

As shown in Fig. 11.10f, there are no holes in the central region, which acts as the core. The surrounding region contains the holes and acts as the cladding.

This type of fiber is single mode over an exceptionally wide wavelength range from 458 to 1550 nm, as determined by measuring the numerical aperture (NA). The NA of the fiber was determined by measuring ϕ_c, where $NA = \sin \phi_c$ in Fig. 11.1. It was found that the value of NA almost linearly increases with wavelength. (NA = 0.13 at

The single-mode excitation was confirmed by observing the field pattern projected onto a distant screen. The field pattern was not affected either by bending the fiber or changing the launching condition into the fiber. If more than one mode were excited, the projected field pattern would change due to the change in the relative phase between the excited modes when the fiber is disturbed.

$\lambda = 458$ nm, and NA $= 0.36$ at $\lambda = 1,550$ nm.) Thus, the value of NA$/\lambda$, hence V, given by Eq. (11.10) maintains an almost constant value with wavelength, and the HF is single mode for such a wide range.

The holey fiber remains single mode even with large core dimensions. For example, a 23 μm core remains single mode at $\lambda = 458$ μm. An advantage of this type of fiber is that high-power light can be transmitted through the large-core-area fiber without concerns about the nonlinearity effects of the core material.

11.1.7.9 Polarization-Preserving Fiber

A polarization-preserving fiber is used in applications where the stability of the state of polarization is important, such as in coherent optical communication systems or optical fiber sensors. A single-mode optical fiber is actually dual mode in the sense that it is capable of supporting two orthogonal polarization modes. In a cylindrically symmetric optical fiber, perturbations such as bends and stresses on the fiber will cause coupling between the two nearly degenerate polarization modes, with the result that the output state of polarization fluctuates. Building a large asymmetry into the fiber reduces the coupling between polarization modes, so that if only one polarization mode is excited at the input, this state of polarization is preserved during transmission through the fiber.

If the effective index of refraction is raised diametrically, then the two polarization modes are orthogonally linearly polarized — one parallel to this direction, and the other perpendicular to it. Incident linearly polarized light aligned to either of these directions will be preserved.

Polarization-preserving fiber is categorized in Fig. 11.11 by the way the asymmetry of the effective index is generated. The first category relies on stress-applying members (SAM) to produce a stress-induced asymmetry of the refractive index. The second category makes use of geometric asymmetry [17], or uses an additional material with a refractive index different from the core or cladding refractive index. An example of a fiber belonging to the first category is the bow-tie fiber. The heat expansion coefficient of the bow-tie-shaped stress member is higher than that of the cladding glass. When the fiber solidifies as it is cooled, an asymmetric tensile stress is built up in the fiber, and this establishes an asymmetric distribution of the index of refraction. The Panda fiber and the elliptical SAM fiber shown on the left-hand side of Fig. 11.11 also belong to this category.

Examples of polarization-preserving fiber belonging to the second category are shown on the right-hand side of Fig. 11.11. The elliptical core fiber makes use of a geometric asymmetry, while the side pit and side tunnel fibers make use of regions with a third index of refraction.

The direction of the increased effective index of refraction is called the slow axis, and the direction of the decreased effective index of refraction is called the fast axis. The direction of the slow axis of both the bow-tie and Panda fibers is in the direction of the stress members.

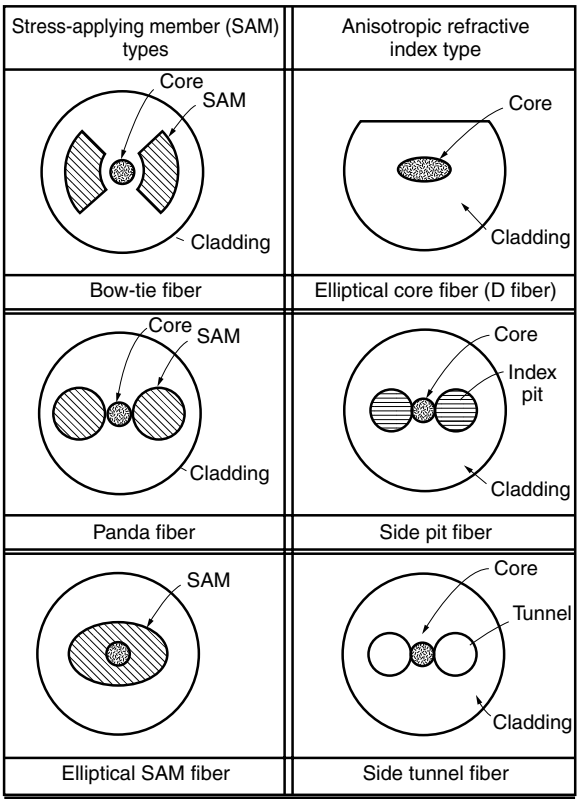


Figure 11.11 Various types of polarization-preserving fiber.

The performance of polarization-preserving fibers is measured by the amount of crosstalk, which is the power received polarized in the direction of the fast axis when linearly polarized light is incident with its polarization parallel to the direction of the slow axis. Crosstalk is typically less than −30 dB/km with a transmission loss of about 2.0 dB/km.

11.1.8 Optical Fibers Other Than Silica Based Fibers

The loss curve of the silicon dioxide (SiO_2)-based fiber was shown in Fig. 11.2. It is V shaped, with the left side of the V dominated by Rayleigh scattering loss, and the right side of the V dominated by the infrared absorption associated with the vibration of the Si — O network. Searches have been made for a material whose vibration frequency of the molecular network is lower than that of the Si — O network, so that the point of the intercept would be lower. If such a material could be found, it might be possible to manufacture a fiber with better transmission characteristics than silicon dioxide.

Figure 11.12 shows the infrared absorption curves due to vibration for various materials. In the order of decreasing vibration frequency, they are the metal fluorides like LiF, MgF_2, and fused BeF_2; the metal chlorides like NaCl, $ZnCl_2$, and KCl; and metal halides like KBr and TlBr [18].

Besides the transmission loss, the degree of difficulty in purifying the material, the flexibility, the degree of dispersion, and the availability of sources have to be taken into

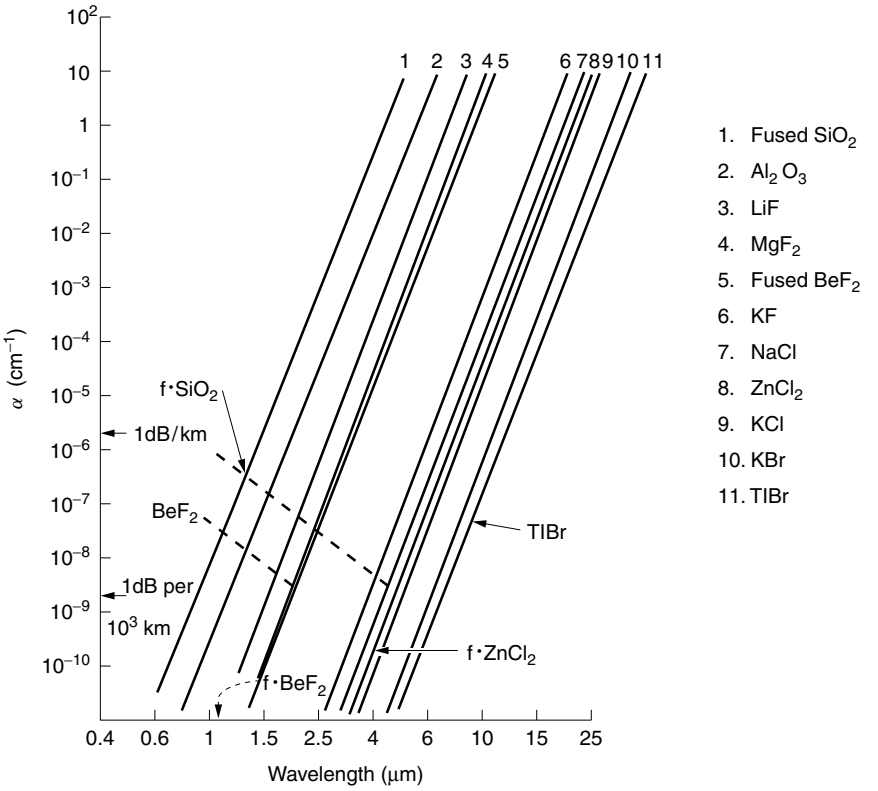


Figure 11.12 Infrared absorption due to lattice vibration for various materials. The arrows indicate the positions of λ_c for the various materials. The upper dashed line indicates scattering-loss values for fused SiO_2; the lower, the predicted values for BeF_2. (After L. G. Van Uitert and S. H. Wemple [18].)

consideration to find new materials for fiber-optic communications. Fiber materials like the metal halides whose minimum loss matches the wavelength of the CO_2 laser have a special significance outside communication as a means of transmitting the energy for a CO_2 laser surgery tool.

11.2 THEORY OF STEP-INDEX FIBERS

Earlier sections of this chapter concentrated on characterization of optical fibers from a practical viewpoint, but the time has now come to face the reality of solving Maxwell's equations to obtain more rigorous expressions. Topics to be covered in this section include hybrid modes with their cutoff conditions, the linearly polarized (LP) modes representation, distortion of a light pulse during transmission inside an optical fiber, and the principles of dispersion-shifted fibers.

11.2.1 Solutions of the Wave Equations in Cylindrical Coordinates

Figure 11.13 shows the geometry of a step-index fiber with a core radius a and cladding radius b. Practically speaking, b is large enough that the evanescent field is negligible, and the assumption of $b = \infty$ is made.

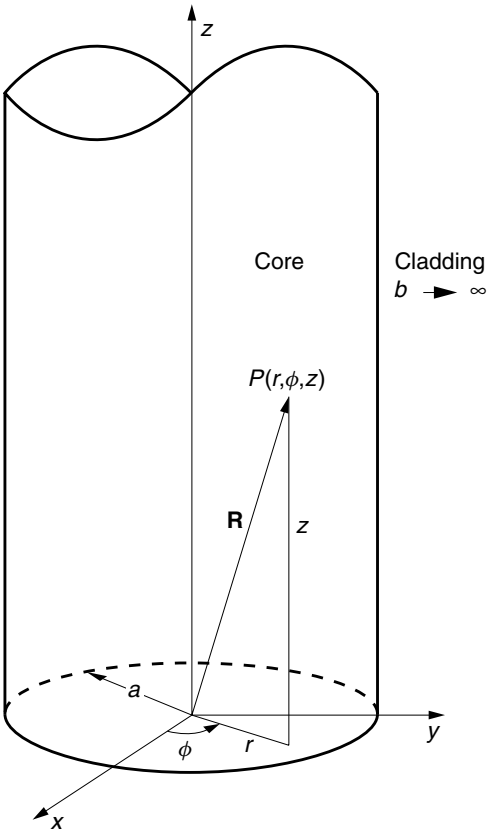


Figure 11.13 Optical fiber in a cylindrical coordinate system.

Since it is easiest to choose a coordinate system that matches the fiber geometry, cylindrical coordinates will be used. The wave equations to be solved are

$$\nabla^2\mathbf{E} + (nk)^2\mathbf{E} = 0 \tag{11.14}$$

$$\nabla^2\mathbf{H} + (nk)^2\mathbf{H} = 0 \tag{11.15}$$

where

$$n = \begin{cases} n_1 & \text{inside the core} \\ n_2 & \text{inside the cladding} \end{cases}$$

$\mathbf{E}$ and H in cylindrical coordinates are

$$\mathbf{E} = E_r\hat{\mathbf{r}} + E_\phi\hat{\boldsymbol{\phi}} + E_z\hat{\mathbf{z}} \tag{11.16}$$

$$\mathbf{H} = H_r\hat{\mathbf{r}} + H_\phi\hat{\boldsymbol{\phi}} + H_z\hat{\mathbf{z}} \tag{11.17}$$

The differential operators in cylindrical coordinates are more complicated than those in rectangular coordinates. The wave equation, Eq. (11.15), in cylindrical coordinates is

$$\nabla^2\mathbf{E} + (nk)^2\mathbf{E} = \left(\nabla^2 E_r - \frac{2}{r^2}\frac{\partial E_\phi}{\partial \phi} - \frac{E_r}{r^2} + (nk)^2 E_r\right)\hat{\mathbf{r}} + \left(\nabla^2 E_\phi + \frac{2}{r^2}\frac{\partial E_r}{\partial \phi} - \frac{E_\phi}{r^2} + (nk)^2 E_\phi\right)\hat{\boldsymbol{\phi}} + \left(\nabla^2 E_z + (nk)^2 E_z\right)\hat{\mathbf{z}} = 0 \tag{11.18}$$

where the Laplacian operator is

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2} \tag{11.19}$$

Note that in Eq. (11.18), the $\hat{\mathbf{r}}$ component contains both E_r and E_ϕ and likewise with the $\hat{\boldsymbol{\phi}}$ component, but the $\hat{\mathbf{z}}$ component contains only E_z. Because of this fact, the z component is first solved and then the other components E_r and E_ϕ will be obtained directly from Maxwell's equation.

11.2.2 Expressions for the E_z and H_z Components

Equation (11.18) being a vector equation, each component has to vanish individually in order for the vector sum of the components to be zero. Using Eq. (11.19), the z component of Eq. (11.18) is expressed as

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial E_z}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 E_z}{\partial \phi^2} + \frac{\partial^2 E_z}{\partial z^2} + (nk)^2 E_z = 0 \tag{11.20}$$

The differential equation, Eq. (11.20), is solved by the method of separation of variables. The component E_z can be expressed as a product of three functions, each individual functions of r, ϕ, and z only:

$$E_z = F(r)\Phi(\phi)Z(z) \tag{11.21}$$

Inserting Eq. (11.21) into (11.20) gives

$$F''\Phi Z + \frac{1}{r}F'\Phi Z + \frac{1}{r^2}F\Phi''Z + F\Phi Z'' + (nk)^2 F\Phi Z = 0 \tag{11.22}$$

Dividing both sides of Eq. (11.22) by $F\Phi Z$ gives

$$\begin{array}{ccccc} \left[\dfrac{F''}{F} + \dfrac{1}{r}\dfrac{F'}{F} + \dfrac{1}{r^2}\dfrac{\Phi''}{\Phi} + (nk)^2\right] & + & \dfrac{Z''}{Z} & = 0 \\ \beta^2 & - & \beta^2 & = 0 \end{array} \tag{11.23}$$

The functions inside the square brackets of Eq. (11.23) are functions of both r and ϕ, but the last term is a function of z only.

Since Eq. (11.23) has to be zero regardless of the combination of r, ϕ, and z, or in other words, since Eq. (11.23) must hold for all points in space, the value inside

the square brackets and the last term have to be individually constant, and the sum of these constants has to be zero. If one defines the value in the square brackets to be β^2,

$$\frac{F''}{F} + \frac{1}{r}\frac{F'}{F} + \frac{1}{r^2}\frac{\Phi''}{\Phi} + (nk)^2 = \beta^2 \tag{11.24}$$

then the last term has to be $-\beta^2$:

$$\frac{Z''}{Z} = -\beta^2 \tag{11.25}$$

Rewriting Eq. (11.24) gives

$$\left[r^2\frac{F''}{F} + r\frac{F'}{F} + ((nk)^2 - \beta^2)r^2\right] + \frac{\Phi''}{\Phi} = 0 \tag{11.26}$$

The function inside the square brackets depends solely on r and the last term is a function solely of ϕ. Similarly, these can be separated by putting the value of the bracketed terms as ν^2 and the last term as $-\nu^2$.

$$\frac{\Phi''}{\Phi} = -\nu^2 \tag{11.27}$$

and

$$F'' + \frac{1}{r}F' + [(nk)^2 - \beta^2]F = \frac{\nu^2}{r^2}F \tag{11.28}$$

For large r, Eq. (11.28) can be approximated as

$$F'' + [(nk)^2 - \beta^2]F = 0 \tag{11.29}$$

The solution of Eq. (11.29) depends on the sign of $[(nk)^2 - \beta^2]$. If it is positive, the solution for large r is sinusoidal, and if it is negative, the solution is either an exponential increase or an exponential decay at large r. The solution of Eq. (11.28) is similarly influenced by the sign of $[(nk)^2 - \beta^2]$. This suggests that in the cladding region $[(nk)^2 - \beta^2]$ should be negative and in the core it should be positive.

With the assumption of Eq. (11.21), the differential Eq. (11.20) finally becomes

$$Z'' + \beta^2 Z = 0 \tag{11.30}$$

$$\Phi'' + \nu^2\Phi = 0 \tag{11.31}$$

$$F'' + \frac{1}{r}F' + [(nk)^2 - \beta^2]F - \frac{\nu^2}{r^2}F = 0 \tag{11.32}$$

There is more than one type of solution for each differential equation given by Eqs. (11.30)–(11.32). Some solutions are sinusoidal, while others are exponential. The selection is made according to physical considerations. The differential equations for the core and cladding regions are solved separately, and the boundary conditions between the two regions are matched by adjusting the coefficients.

11.2.2.1 Solutions in the Core Region

First, the solutions inside the core region are considered.

$$Z(z) = ae^{j\beta z} + be^{-j\beta z} \tag{11.33}$$

$$\Phi(\phi) = ce^{j\nu\phi} + de^{-j\nu\phi} \tag{11.34}$$

or

$$\Phi(\phi) = c' \cos \nu\phi + d' \sin \nu\phi \tag{11.35}$$

and lastly the solution of Eq. (11.32) is considered. Since sinusoidal-like variation rather than exponential variation in r is desired inside the core, a positive value of $K^2 = [(nk)^2 - \beta^2]$ is selected.

$$F(r) = eJ_\nu(Kr) + fN_\nu(Kr) \tag{11.36}$$

where

$$K^2 = (n_1 k)^2 - \beta^2 \tag{11.37}$$

$$n = n_1 \tag{11.38}$$

Equation (11.33) consists of forward and backward waves. For the most part, fibers are reciprocal. There do exist some nonreciprocal effects that are employed in certain fiber sensors, but these are unusual. Reciprocity means that, coupling conditions being equal, it does not matter which end of the fiber is connected to the source. It is sufficient to choose one term from the right-hand side of Eq. (11.33). There is a choice between Eq. (11.34) and Eq. (11.35). Equation (11.34) represents two skew rays that are rotating in opposite senses. Mode patterns are standing waves generated by oppositely winding skew rays and both terms in Eq. (11.34) are needed. Equations (11.34) and (11.35) with the proper selection of c, c', d, and d' are equivalent. There is, however, a restraint in choosing the value of ν. The location of ϕ and $\phi + 2\pi$ are the same, and

$$\Phi(\phi) = \Phi(\phi + 2\pi) \tag{11.39}$$

has to be satisfied. The requirement is that ν is an integer.

$J_\nu(Kr)$ and $N_\nu(Kr)$ are respectively, the νth order Bessel function of the first kind and the Bessel function of the second kind. The curves of $N_\nu(x)$ are shown in Fig. 11.14. $J_\nu(Kr)$ will be described later in Fig. 11.16. $N_\nu(Kr)$ becomes negative infinity at $r = 0$ and physically cannot be accepted, and f in Eq. (11.36) has to be zero.

Both **E** and **H** satisfy the same kind of wave equations, Eqs. (11.14) and (11.15). In the core region $r < a$, E_z and H_z are given by

$$E_z = AJ_\nu(Kr)e^{j\beta z + j\nu\phi} \tag{11.40}$$

$$H_z = BJ_\nu(Kr)e^{j\beta z + j\nu\phi} \tag{11.41}$$

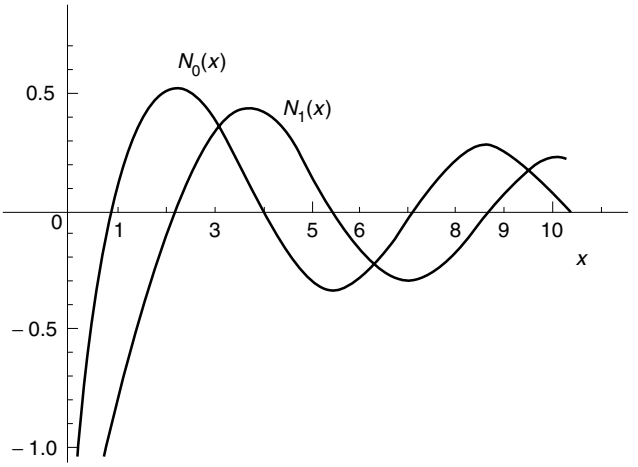


Figure 11.14 Zero- and first-order Bessel functions of the second kind.

Alternatively, the solutions could be written as

$$E_z = J_\nu(Kr)\cos(\nu\phi)e^{j\beta z}$$
$$H_z = J_\nu(Kr)\sin(\nu\phi)e^{j\beta z} \tag{11.42}$$

However, when the real part is taken at the end to obtain the final field expressions, the results are the same.

11.2.2.2 Solutions in the Cladding Region

Next, the field in the cladding region will be found. The type of solution suitable for the cladding region is one where the value of the function rapidly decreases with an increase in r. As mentioned earlier, the solution of Eq. (11.32) becomes a decaying function if the value of the square bracket is negative, that is,

$$-\gamma^2 = (n_2 k)^2 - \beta^2 \tag{11.43}$$

where

$$n = n_2 \tag{11.44}$$

The solution for F is then

$$F(r) = eI_\nu(\gamma r) + fK_\nu(\gamma r) \tag{11.45}$$

The functions $I_\nu(\gamma r)$ and $K_\nu(\gamma r)$ are, respectively, called the modified Bessel function of the first kind and the modified Bessel function of the second kind of the νth order. The value of $I_\nu(x)$ increases with an increase in x but that of $K_\nu(x)$ decreases with an increase in x, as shown in Fig. 11.15. Since $I_\nu(x)$ becomes indefinitely large with an increase in r, it is rejected on physical grounds, and e in Eq. (11.45) should be zero.

In summary, the expressions for E_z and H_z in the cladding region, $r > a$, are

$$E_z = CK_\nu(\gamma r)e^{j\beta z + j\nu\phi} \tag{11.46}$$

$$H_z = DK_\nu(\gamma r)e^{j\beta z + j\nu\phi} \tag{11.47}$$

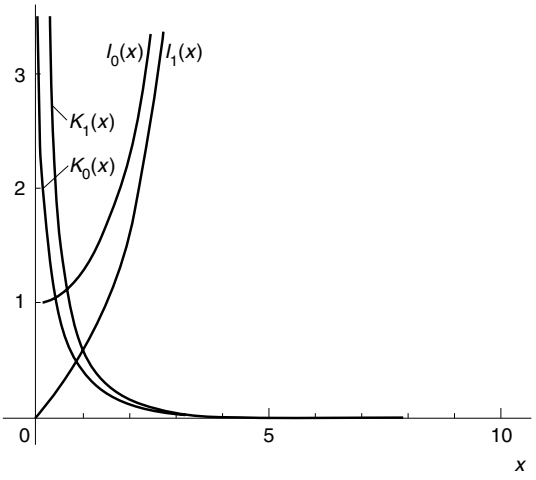


Figure 11.15 Modified Bessel functions of first $I_\nu(x)$ and the second $K_\nu(x)$ kinds.

In some literature, the Hankel function $H_\nu^{(1)}(j\gamma r)$ with a complex argument is used instead of $K_\nu(\gamma r)$. When ν is an integer, the relationship between them is

$$K_\nu(x) = \frac{\pi}{2} j^{\nu+1} H_\nu^{(1)}(jx) \tag{11.48}$$

11.2.3 Expressions for the E_r, E_ϕ, H_r, and H_ϕ Components

Having found expressions for E_z and H_z, the rest of the components are derived using Maxwell's equations. The components of

$$\nabla \times \mathbf{H} = -j\omega\epsilon\mathbf{E} \tag{11.49}$$

where

$$\nabla \times \mathbf{H} = \begin{bmatrix} \dfrac{\hat{\mathbf{r}}}{r} & \hat{\boldsymbol{\phi}} & \dfrac{\hat{\mathbf{k}}}{r} \\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial \phi} & \dfrac{\partial}{\partial z} \\ H_r & rH_\phi & H_z \end{bmatrix}$$

in cylindrical coordinates are

$$\frac{1}{r}\frac{\partial H_z}{\partial \phi} - \frac{\partial H_\phi}{\partial z} = -j\omega\epsilon E_r \tag{11.50}$$

$$\frac{\partial H_r}{\partial z} - \frac{\partial H_z}{\partial r} = -j\omega\epsilon E_\phi \tag{11.51}$$

$$\frac{1}{r}\frac{\partial (rH_\phi)}{\partial r} - \frac{1}{r}\frac{\partial H_r}{\partial \phi} = -j\omega\epsilon E_z \tag{11.52}$$

and for

$$\nabla \times \mathbf{E} = j\omega\mu\mathbf{H} \tag{11.53}$$

$$\frac{1}{r}\frac{\partial E_z}{\partial \phi} - \frac{\partial E_\phi}{\partial z} = j\omega\mu H_r \tag{11.54}$$

$$\frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} = j\omega\mu H_\phi \tag{11.55}$$

$$\frac{1}{r}\frac{\partial (rE_\phi)}{\partial r} - \frac{1}{r}\frac{\partial E_r}{\partial \phi} = j\omega\mu H_z \tag{11.56}$$

Equations (11.50) and (11.55) are solved for E_r and H_ϕ in terms of E_z and H_z, and Eqs. (11.51) and (11.54) are solved for E_ϕ and H_r in terms of E_z and H_z. The relationship $\partial/\partial z = j\beta$ is used.

$$E_r = \frac{j}{K^2}\left(\beta\frac{\partial E_z}{\partial r} + \omega\mu\frac{1}{r}\frac{\partial H_z}{\partial \phi}\right) \tag{11.57}$$

$$E_\phi = \frac{j}{K^2}\left(\frac{\beta}{r}\frac{\partial E_z}{\partial \phi} - \omega\mu\frac{\partial H_z}{\partial r}\right) \tag{11.58}$$

$$H_r = \frac{j}{K^2}\left(-\omega\epsilon\frac{1}{r}\frac{\partial E_z}{\partial \phi} + \beta\frac{\partial H_z}{\partial r}\right) \tag{11.59}$$

$$H_\phi = \frac{j}{K^2}\left(\omega\epsilon\frac{\partial E_z}{\partial r} + \frac{\beta}{r}\frac{\partial H_z}{\partial \phi}\right) \tag{11.60}$$

where

$$K^2 = (nk)^2 - \beta^2 \tag{11.61}$$

The fields inside the core region $r < a$ are obtained by inserting Eqs. (11.40) and (11.41) into Eq. (11.57) through (11.61) but with ϵ replaced by ϵ_1 and with K^2 replaced by $K^2 = (n_1k)^2 - \beta^2$:

$$E_r = \frac{j}{K^2}\left(A\beta K J'_\nu(Kr) + B\omega\mu\frac{j\nu}{r}J_\nu(Kr)\right) \tag{11.62}$$

$$E_\phi = \frac{j}{K^2}\left(A\frac{\beta}{r}j\nu J_\nu(Kr) - B\omega\mu K J'_\nu(Kr)\right) \tag{11.63}$$

$$H_r = \frac{j}{K^2}\left(-A\omega\epsilon_1\frac{j\nu}{r}J_\nu(Kr) + B\beta K J'_\nu(Kr)\right) \tag{11.64}$$

$$H_\phi = \frac{j}{K^2}\left(A\omega\epsilon_1 K J'_\nu(Kr) + B\frac{\beta}{r}j\nu J_\nu(Kr)\right) \tag{11.65}$$

where $e^{j\beta z + j\nu\phi}$ was suppressed.

The fields inside the cladding region $r > a$ are obtained by inserting Eqs. (11.46) and (11.47) into Eqs. (11.57) through (11.61), but with ϵ replaced by ϵ_2 and with K^2

replaced by $-\gamma^2 = (n_2 k)^2 - \beta^2$:

$$E_r = \frac{-j}{\gamma^2}\left(C\beta\gamma K'_\nu(\gamma r) + D\omega\mu\frac{j\nu}{r}K_\nu(\gamma r)\right) \tag{11.66}$$

$$E_\phi = \frac{-j}{\gamma^2}\left(C\frac{\beta}{r}j\nu K_\nu(\gamma r) - D\omega\mu\gamma K'_\nu(\gamma r)\right) \tag{11.67}$$

$$H_r = \frac{-j}{\gamma^2}\left(-C\omega\epsilon_2\frac{j\nu}{r}K_\nu(\gamma r) + D\beta\gamma K'_\nu(\gamma r)\right) \tag{11.68}$$

$$H_\phi = \frac{-j}{\gamma^2}\left(C\omega\epsilon_2\gamma K'_\nu(\gamma r) + D\frac{\beta}{r}j\nu K_\nu(\gamma r)\right) \tag{11.69}$$

where $e^{j\beta z + j\nu\phi}$ was suppressed. Note that if A, B, ϵ_1, K, and $J_\nu(Kr)$ in Eqs. (11.62) to (11.65) are replaced by C, D, ϵ_2, $j\gamma$, and $K_\nu(\gamma r)$, respectively, the results are Eqs. (11.66) to (11.69).

11.2.4 Characteristic Equation of an Optical Fiber

Boundary conditions are applied on the boundary between the core and cladding regions. The tangential components of **E** and **H** have to be continuous at $r = a$. The tangential components of **E** are E_z and E_ϕ and those of **H** are H_z and H_ϕ.

There are two ways to find the coefficients A, B, C, and D that satisfy these conditions. One way is by a matrix method, whereby the conditions for E_z, E_ϕ, H_z, and H_ϕ at $r = a$ are arranged in a matrix form with A, B, C, and D in Eqs. (11.62) through (11.69) as unknowns. Then the 4×4 determinant is calculated to find the condition for the nontrivial solutions of A, B, C and D [19]. The other way is to find the value of C in terms of A using the continuity condition for E_z, and to find the value of D in terms of B using the continuity condition for H_z. Then, only the 2×2 matrix of A and B has to be calculated. This method involves less manipulation and will be used here.

First, C will be found in terms of A. Continuity of E_z gives

$$AJ_\nu(Ka) = CK_\nu(\gamma a) \tag{11.70}$$

hence,

$$C = \frac{J_\nu(Ka)}{K_\nu(\gamma a)}A \tag{11.71}$$

In a similar manner, continuity of H_z is used to find D in terms of B.

$$D = \frac{J_\nu(Ka)}{K_\nu(\gamma a)}B \tag{11.72}$$

Both Eqs. (11.71) and (11.72) have the same factor.

Next, C and D in the expressions for E_ϕ and H_ϕ, Eqs. (11.67) and (11.69), are to be replaced by A and B using Eqs. (11.71) and (11.72), and then the equations necessary

for the continuity of E_ϕ and H_ϕ are reduced to

$$A\beta\nu\left(\frac{1}{(Ka)^2}+\frac{1}{(\gamma a)^2}\right)+jB\omega\mu\left(\frac{J'_\nu(Ka)}{KaJ_\nu(Ka)}+\frac{K'_\nu(\gamma a)}{\gamma aK_\nu(\gamma a)}\right)=0 \tag{11.73}$$

$$A\omega\left(\frac{\epsilon_1 J'_\nu(Ka)}{KaJ_\nu(Ka)}+\frac{\epsilon_2 K'_\nu(\gamma a)}{\gamma aK_\nu(\gamma a)}\right)+jB\beta\nu\left(\frac{1}{(Ka)^2}+\frac{1}{(\gamma a)^2}\right)=0 \tag{11.74}$$

For nontrivial solutions for A and B to exist, the determinant of the coefficients has to vanish. The result is

$$\left(\frac{1}{Ka}\frac{J'_\nu(Ka)}{J_\nu(Ka)}+\frac{1}{\gamma a}\frac{K'_\nu(\gamma a)}{K_\nu(\gamma a)}\right)\left(\frac{J'_\nu(Ka)}{KaJ_\nu(Ka)}+\left(\frac{n_2}{n_1}\right)^2\frac{K'_\nu(\gamma a)}{\gamma aK_\nu(\gamma a)}\right)$$
$$=\left[\frac{\beta\nu}{n_1 k}\left(\frac{1}{(Ka)^2}+\frac{1}{(\gamma a)^2}\right)\right]^2 \tag{11.75}$$

This is the characteristic equation for a step-index fiber.

Only values of Ka and γa that satisfy Eq. (11.75) will satisfy the boundary condition at the core–cladding boundary. There are two unknowns, Ka and γa, in Eq. (11.75). One more equation is needed to determine the unknowns. That is obtained by combining Eqs. (11.37) and (11.43):

$$V^2=(Ka)^2+(\gamma a)^2 \tag{11.76}$$

where

$$V=ka\sqrt{n_1^2-n_2^2} \tag{11.77}$$

In short, one has to solve for the γa and Ka that satisfy Eqs. (11.75) and (11.76) for given physical parameters of the normalized radius V and the value of n_2/n_1. The procedures for obtaining the solutions are exactly the same as those of the slab optical guides described in Chapter 9.

11.2.5 Modes in Optical Fibers

A graph of Eq. (11.75) in $(\gamma a, Ka)$ coordinates is a combination of Bessel function curves. The graph of Eq. (11.76) is a circle with radius V. The intersects between the two curves are solutions. Each intersect is designated as a mode that can be excited and gives such information as the cutoff conditions and the cross-sectional distribution of light in the fiber.

The solutions of the characteristic equation vary a great deal depending on whether $\nu=0$ or $\nu\neq 0$. Hence, they will be treated in two separate sections.

11.2.5.1 Meridional Modes: $\nu=0$

As seen from Eq. (11.34) or (11.35), as soon as $\nu=0$, the ϕ-dependent factor becomes a constant and a meridional ray is excited. The characteristic equation is significantly

simplified. With $\nu = 0$ the order of the Bessel function reduces, and the right-hand side of Eq. (11.75) becomes zero so that Eq. (11.75) reduces to

$$\frac{J_1(Ka)}{KaJ_0(Ka)} + \frac{K_1(\gamma a)}{\gamma a K_0(\gamma a)} = 0, \quad \text{TE mode} \tag{11.78}$$

and

$$\frac{J_1(Ka)}{KaJ_0(Ka)} + \left(\frac{n_2}{n_1}\right)^2 \frac{K_1(\gamma a)}{\gamma a K_0(\gamma a)} = 0, \quad \text{TM mode} \tag{11.79}$$

where the Bessel function relationships $J_0'(x) = -J_1(x)$ and $K_0'(x) = -K_1(x)$ were used. Equation (11.78) is the characteristic equation for the TE mode and Eq. (11.79) is for the TM mode.

There is a reason for naming these characteristic equations the TE mode or the TM mode, as will be explained before proceeding. From Eqs. (11.40) and (11.41), the ratio of the intensities of H_z to E_z is $(B/A)^2$. Using Eq. (11.73), the ratio $(B/A)^2$ is calculated, and then the term $\beta\nu(1/(Ka)^2 + 1/(\gamma a)^2)$ is rewritten using Eq. (11.74) to obtain

$$\left(\frac{B}{A}\right)^2 = \frac{1}{\eta_1^2} \frac{\dfrac{J_\nu'(Ka)}{KaJ_\nu(Ka)} + \left(\dfrac{n_2}{n_1}\right)^2 \dfrac{1}{\gamma a} \dfrac{K_\nu'(Ka)}{K_\nu(Ka)}}{\dfrac{J_\nu'(Ka)}{KaJ_\nu(Ka)} + \dfrac{K_\nu'(\gamma a)}{\gamma a K_\nu(\gamma a)}} \tag{11.80}$$

where $\eta_1 = \sqrt{\mu/\epsilon_1}$.

If Eq. (11.78) is satisfied, the denominator of the right-hand side of Eq. (11.80) becomes zero for the meridional mode; hence, $A = 0$. This means $E_z = 0$ and there is no longitudinal component of $\mathbf{E}$ and its mode is called a TE (transverse electric) mode. On the other hand, when Eq. (11.79) is satisfied, $B = 0$ and hence $H_z = 0$ and its mode is a TM (transverse magnetic) mode.

Let us leave the exact shape of the Ka versus γa curves of Eqs. (11.78) and (11.79) to the magic power of an electronic computer, and concentrate our attention on making some sense out of the curves. Let us start by examining the curve near the zero and infinity of γ.

The mode is cut off at $\gamma a = 0$ because the decay of the field with respect to r in the cladding is absent and the field is not confined inside the core. The operating point of a single-mode fiber is normally placed at as large a value of γa as possible, within the limit that the next higher mode is not excited.

First, the cutoff conditions of the TE modes will be examined by setting

$$\gamma a = 0 \tag{11.81}$$

Judging from the curves of $K_0(x)$ and $K_1(x)$ in Fig. 11.15 (or more exactly by Eq. (11.98), which will appear later), the limit as γa approaches zero gives

$$\lim_{\gamma a \to 0} \frac{K_1(\gamma a)}{K_0(\gamma a)} \gg 1 \tag{11.82}$$

Hence, the condition for Eq. (11.78) to be valid at $\gamma a = 0$ is

$$J_0(Ka) = 0_- \tag{11.83}$$

where 0_- means extremely close to zero but a negative number. The roots $\omega_{0\mu}$ of $J_0(\omega) = 0$ are

$$\omega_{01} = 2.4 \qquad \omega_{02} = 5.5$$
$$\omega_{03} = 8.7 \qquad \omega_{04} = 11.8$$

In conclusion, the values of Ka at the $\gamma a = 0$ cutoff are $\omega_{0\mu}$.* The characteristic curves are numbered starting with the smallest Ka cutoff values to the larger values.

The TE modes are designated by $TE_{0\mu}$, where 0 means $\nu = 0$, and μ indicates the μth root of the $\nu = 0$ modes. Thus, the cutoffs of the TE_{01}, TE_{02}, and TE_{03} modes are at $Ka = 2.4$, 5.5, and 8.7, respectively.

From Eq. (11.79), applying the limit as γa approaches zero gives the same result as for Eq. (11.78). Thus, at the cutoff $\gamma a = 0$, the TM modes become identical to the TE modes and have the same cutoff values as above. It is said that the TE and TM modes are degenerate at $\gamma a = 0$. At all other points, however, they are different and the degeneracy is removed.

Next, the behavior of the TE mode in the region far from cutoff, $\gamma a \to \infty$, is examined. As seen either from Fig. 11.15 (or more exactly by Eq. (11.101), which will appear later), $\lim_{\gamma a \to \infty}[K_1(\gamma a)/K_0(\gamma a)]$ is finite and in this region, the first term in Eq. (11.78) has to be zero:

$$J_1(Ka) = 0 \tag{11.84}$$

The roots for $J_1(P) = 0$ are

$$P_{10} = 0.0 \qquad P_{11} = 3.8$$
$$P_{12} = 7.0 \qquad P_{13} = 10$$

The first term of Eq. (11.78) does not become zero at $Ka = P_{10}$ because Ka is also zero and the value of the limit of the first term becomes

$$\lim_{Ka \to 0} \frac{J_1(Ka)}{Ka} = \frac{1}{2} \tag{11.85}$$

and P_{10} has to be excluded as an asymptote for $\gamma a = \infty$. Equation (11.79) will give the same result as Eq. (11.78) and the TM modes have the same asymptote for $\gamma a = \infty$ as the TE modes.

From the behavior near $\gamma a = 0$ and $\gamma a = \infty$, the general curve shape of the characteristic equation is drawn, as shown in Fig. 11.16. The upper and lower graphs are $J_1(Ka)$ and $J_0(Ka)$, respectively. From the roots of these functions, dotted lines are extended to the middle graph to indicate the locations of the cutoff and the asymptotes of the characteristic curve.

The values of Ka and γa for a given value of V are found from the intercepts between the characteristic curve and the circle with radius V. The propagation constant β for a particular mode is found by using Eq. (11.37) or (11.43). This procedure is the same as that of the slab optical guide treated in Section 9.3.1.

* $Ka = 0$ is excluded because of Eq. (11.85).

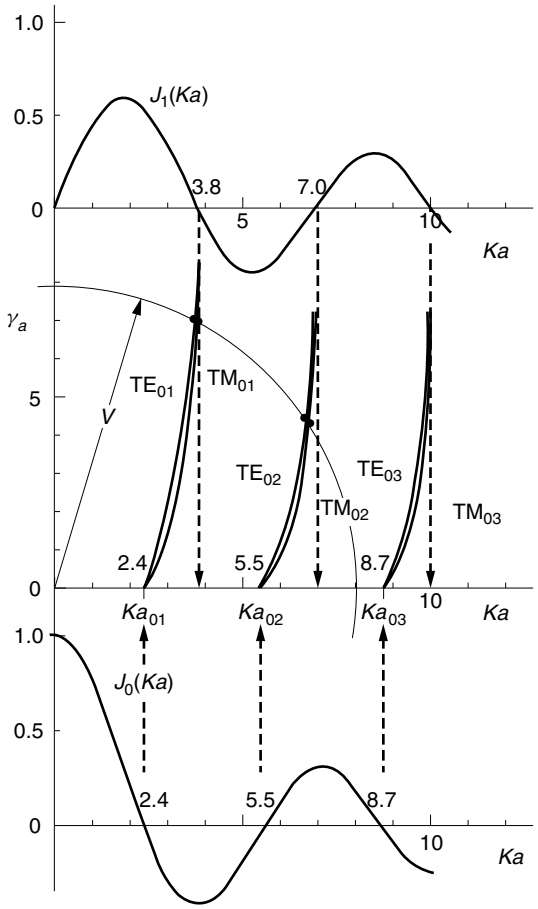


Figure 11.16 Curves of the characteristic equation of the $TE_{0\mu}$ and $TM_{0\mu}$ modes.

When Eq. (11.78) for the $TE_{0\mu}$ modes is compared with Eq. (11.79) for $TM_{0\mu}$, the only difference is that Eq. (11.79) has an extra factor $(n_2/n_1)^2$, which in most cases is $(1.005)^2$ and is close to unity. Both characteristic curves are of similar shape except that the curve for $TM_{0\mu}$ is slightly lower than that of the $TE_{0\mu}$ mode.

11.2.5.2 Skew Modes: $\nu \neq 0$

When $\nu \neq 0$, the right-hand side of Eq. (11.75) is not zero and the procedure for obtaining the solution of the characteristic equation is much more complicated. In practice, the difference in the index of refraction of the core and cladding regions is 0.3–0.5% for most optical fibers, and the approximation

$$\left(\frac{n_2}{n_1}\right)^2 \doteqdot 1 \tag{11.86}$$

is a very good one.

This approximation significantly reduces the complexity of solving Eq. (11.75), for both the $\nu = 0$ and $\nu \neq 0$ cases. The theory based on this approximation is known as the *theory with the weakly guiding approximation* [20]. This name is from a sense that the guiding capability of a fiber comes from the difference in the indices of refraction of the core and cladding, and when the difference is small, the light is weakly guided. With the weakly guiding approximation, Eq. (11.75) becomes

$$\frac{J'_\nu(Ka)}{KaJ_\nu(Ka)} + \frac{K'_\nu(\gamma a)}{\gamma a K_\nu(\gamma a)} = \pm\nu\left(\frac{1}{(Ka)^2} + \frac{1}{(\gamma a)^2}\right) \tag{11.87}$$

where the assumption

$$\beta \doteqdot n_1 k \tag{11.88}$$

was made, because in a weakly guided fiber, only the component waves with a very shallow zigzag angle can propagate.

As mentioned earlier, the case of $\nu = 0$ was rather straightforward. One of the roots was $A = 0$; hence, $E_z = 0$ and this mode was called the TE mode or sometimes the H mode. The other root was $B = 0$; hence, $H_z = 0$ and this was called the TM mode or E mode.

In the case of $\nu \neq 0$, the situation is different. With the weakly guiding approximation, Eq. (11.75) is split into two equations, as indicated in Eq. (11.87), because of the positive and negative signs. Referring to the right-hand side of Eq. (11.87), one has to realize that the Ka versus γa curve with the $+$ sign is certainly different from that with the $-$ sign, for the same value of ν. It is necessary to establish two names to distinguish the two different solutions. There are three ways of distinguishing the two solutions. One is by the sign in front of ν. The characteristic equation with the negative sign is named the characteristic equation of the *HE* modes while that with the positive sign is named the characteristic equation of the *EH* modes. The second is by the sign of $j(B/A)\eta_1$, which is more popular.

Let us derive the value of $j(B/A)\eta_1$. From Eq. (11.73) and (11.87), the ratio B/A is

$$\frac{B}{A} = -\frac{\beta}{j\omega\mu}\frac{\nu\left(\dfrac{1}{(Ka)^2} + \dfrac{1}{(\gamma a)^2}\right)}{\pm\nu\left(\dfrac{1}{(Ka)^2} + \dfrac{1}{(\gamma a)^2}\right)} \tag{11.89}$$

With Eq. (11.88), Eq. (11.89) becomes

$$j\frac{B}{A} = \mp\frac{1}{\eta_1} \tag{11.90}$$

The values of $j(B/A)\eta_1$ depend on the sign in front of ν. Snitzer [21] proposed to call a mode *HE* if the value $j(B/A)\eta_1$ is $+1$ at the cutoff, and *EH* if it is -1 at the cutoff.

Here we will accept Snitzer's proposal.

$$\begin{aligned} &\text{With negative sign in front of } \nu, \quad j\frac{B}{A}\eta_1 = +1 \quad HE \text{ mode} \\ &\text{With positive sign in front of } \nu, \quad j\frac{B}{A}\eta_1 = -1 \quad EH \text{ mode} \end{aligned} \tag{11.91}$$

Thus, it has been shown that the sign in front of ν is related to that of $j(B/A)\eta_1$.

The fact that $j(B/A)\eta_1 = \pm 1$ means that A or B alone cannot be zero, and there is no pure TE or TM mode because TE, for instance, requires $A = 0$ and $B \neq 0$. Both E_z and H_z are present when $\nu \neq 0$. Modes that contain both E_z and H_z components are called *hybrid modes*. This is the reason why the notation for hybrid modes uses both letters, E and H.

The third way of deciding which solution is to be called the HE mode and which is to be called the EH mode will be mentioned. This method dictates that, when the E_z component found by using the rigorous equations is larger than the H_z component, the mode is called an EH mode, and when it is the other way around, the mode is called an HE mode. This method, however, requires detailed information about the location, how far from the cutoff, and so forth.

Now, Eq. (11.87) will be rewritten using the following Bessel function recurrence formulas

$$J'_\nu(x) = J_{\nu-1}(x) - \frac{\nu}{x}J_\nu(x) \tag{11.92}$$

$$K'_\nu(x) = -K_{\nu-1}(x) - \frac{\nu}{x}K_\nu(x) \tag{11.93}$$

$$J'_\nu(x) = -J_{\nu+1}(x) + \frac{\nu}{x}J_\nu(x) \tag{11.94}$$

$$K'_\nu(x) = -K_{\nu+1}(x) + \frac{\nu}{x}K_\nu(x) \tag{11.95}$$

Equation (11.87) with the negative sign and Eqs. (11.92) and (11.93) lead to the characteristic equation of the HE modes, while Eq. (11.87) with the positive sign and Eqs. (11.94) and (11.95) lead to the characteristic equation of the EH modes.

$$\frac{J_{\nu-1}(Ka)}{KaJ_\nu(Ka)} - \frac{K_{\nu-1}(\gamma a)}{\gamma a K_\nu(\gamma a)} = 0, \quad HE \text{ modes} \tag{11.96}$$

$$\frac{J_{\nu+1}(Ka)}{KaJ_\nu(Ka)} + \frac{K_{\nu+1}(\gamma a)}{\gamma a K_\nu(\gamma a)} = 0, \quad EH \text{ modes} \tag{11.97}$$

First, the HE cutoffs will be found. Useful Bessel function approximations are, for $x \ll 1$,

$$\frac{K_0(x)}{K_1(x)} = x \ln\left(\frac{2}{1.782x}\right) \tag{11.98}$$

$$\frac{K_{\nu-1}(x)}{K_\nu(x)} = \frac{x}{2(\nu-1)}, \quad \nu \geq 2 \tag{11.99}$$

$$\frac{J_{\nu-1}(x)}{J_\nu(x)} = \frac{2}{x}\nu \tag{11.100}$$

and for $x \gg 1$,

$$\frac{K_{\nu\mp1}(x)}{K_\nu(x)} = 1 + \frac{1 \mp 2\nu}{2x} \tag{11.101}$$

The behavior of Eq. (11.96) near cutoff, $\gamma a \to 0$, will be investigated for $\nu \geq 2$. With Eq. (11.99), near $\gamma a \to 0$, Eq. (11.96) becomes

$$\frac{J_{\nu-1}(Ka)}{KaJ_\nu(Ka)} = \frac{1}{2(\nu-1)} \tag{11.102}$$

The recurrence formulas are from Eqs. (11.92) to (11.95),

$$J_{n+1}(x) + J_{n-1}(x) = \frac{2n}{x}J_n(x) \tag{11.103}$$

$$K_{n+1}(x) - K_{n-1}(x) = \frac{2n}{x}K_n(x) \tag{11.104}$$

Equation (11.103) is used with $n = \nu - 1$ to reduce Eq. (11.102) to

$$KaJ_{\nu-2}(Ka) = 0 \tag{11.105}$$

For $\nu \geq 2$, $Ka = 0$ has to be excluded from the solutions. This is because for $\nu \geq 2$ the right-hand side of Eq. (11.102) is finite, and the left-hand side becomes infinite from Eq. (11.100). The roots of Eq. (11.105) are the cutoffs of the HE modes. The roots are numbered by μ, and the HE mode with the μth root is designated by $HE_{\nu\mu}$. The roots of the $HE_{2\mu}$ modes are $Ka = 2.40, 5.52, 8.65, \ldots$.

In the above, the case of $\nu = 1$ was avoided because a different approximate expression applies, namely Eq. (11.98) rather than Eq. (11.99). With Eq. (11.98), the second term of Eq. (11.96) approaches infinity as $\gamma a \to 0$, and the solution becomes

$$KaJ_1(Ka) = 0 \tag{11.106}$$

On the other hand, from Eq. (11.96), when $\gamma a \to \infty$, $J_0(Ka) = 0$.

The general curve shape of the characteristic equation is drawn as shown in Fig. 11.17. The upper and lower graphs are $J_0(Ka)$ and $J_1(Ka)$, respectively,* to indicate the locations of the asymptotes and the cutoff of the characteristic curve.

The roots of Eq. (11.106) are $Ka = 0$, 3.83, 7.02, 10.2. The first root is $Ka = 0$, and the cutoff of the HE_{11} mode is zero. For this reason the HE_{11} mode is the dominant mode of the optical fiber. The cutoff condition of the next higher order mode is Eq. (11.83) for TE_{01} and TM_{01}, and Eq. (11.105) for the HE_{21} mode as shown in Fig. 11.18. Both of these cutoff conditions are the same, $J_0(Ka) = 0$. Thus, the condition to be a single-mode fiber is

$$V < 2.40 \tag{11.107}$$

Next, the cutoffs of the EH modes will be obtained. First, the cases with $\nu \geq 2$ are considered. Using the relationship of Eq. (11.99) with $\nu = \nu' + 1$, we see that

* $KaJ_1(Ka)$ and $J_1(Ka)$ have their zeros at the same locations.

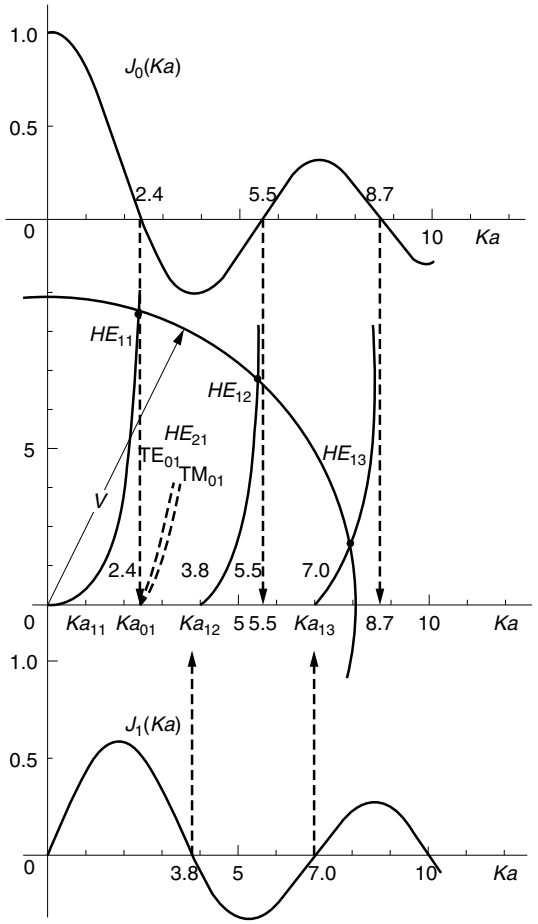


Figure 11.17 Curves of the characteristic equation of the $HE_{1\mu}$ modes.

the second term of Eq. (11.97) becomes infinite as $\gamma a \to 0$, and at first glance the solution is

$$KaJ_\nu(Ka) = 0 \tag{11.108}$$

The first root of Eq. (11.108) is $Ka = 0$. However, we have to consider this first root more carefully. An approximate expression for $J_\nu(Ka)$ for small Ka and integer ν is

$$J_\nu(Ka) = \frac{1}{\nu!}\left(\frac{Ka}{2}\right)^\nu \tag{11.109}$$

With Eq. (11.109), the first term of Eq. (11.97) becomes

$$\frac{J_{\nu+1}(Ka)}{KaJ_\nu(Ka)} = \frac{1}{2(\nu+1)} \tag{11.110}$$

which is not infinity at $Ka = 0$, and the first root has to be excluded. Next, a similar treatment with $\nu = 1$ leads to the same condition as Eq. (11.108) and the same exclusion of $Ka = 0$. The lowest cutoff of the EH modes is at $Ka = 3.83$ of the EH_{11} mode, and the cutoff of the next higher EH mode is at $Ka = 5.13$ of EH_{21} as summarized in Fig. 11.18.

From Eqs. (11.105) and (11.108), note that the cutoff conditions of $HE_{\nu+1,\mu}$ are identical to that of $EH_{\nu-1,\mu}$.

Next, it will be shown that it is not just the cutoff values of the $HE_{\nu+1,\mu}$ and $EH_{\nu-1,\mu}$ modes that are identical, but the characteristic equations themselves are identical. After inserting $\nu = m + 1$ into Eq. (11.96) and $\nu = m - 1$ into Eq. (11.97), the second terms are moved to the right-hand side. Then, both sides are inverted. The results are

$$\frac{KaJ_{m+1}(Ka)}{J_m(Ka)} = \frac{\gamma aK_{m+1}(\gamma a)}{K_m(\gamma a)} \tag{11.111}$$

$$\frac{KaJ_{m-1}(Ka)}{J_m(Ka)} = -\frac{\gamma aK_{m-1}(\gamma a)}{K_m(\gamma a)} \tag{11.112}$$

and use of the recurrence formulas, Eqs. (11.103) and (11.104), for rewriting $J_{m+1}(Ka)$ and $K_{m+1}(\gamma a)$ convert Eq. (11.111) into (11.112), thus proving that the characteristic equation of $HE_{m+1,\mu}$ is identical with that of $EH_{m-1,\mu}$ in the case of the weakly guiding approximation. Because of the identical nature of the characteristic equations, the two modes are given the same mode name, $\mathrm{LP}_{m\mu}$, which stands for linearly polarized mode [20]. For instance, the characteristic equations of the $HE_{3\mu}$, $EH_{1\mu}$ and $\mathrm{LP}_{2\mu}$ modes are all identical as far as the characteristic equations are concerned. However, the field patterns of these three modes are all different, as will be explained in detail in the next section.

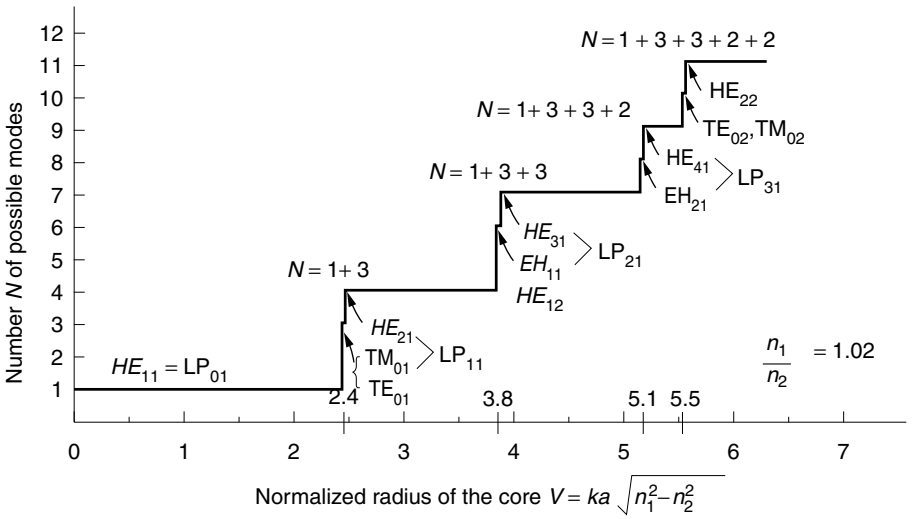


Figure 11.18 Number of possible modes in the step-index fiber as a function of the normalized radius of the core.

Example 11.1 What is the HE mode that pairs with $EH_{5\mu}$ to form a linear polarization (LP) mode? What is the corresponding $\mathrm{LP}_{m\mu}$ mode?

Solution They are $HE_{7\mu}$ and $\mathrm{LP}_{6\mu}$. □

As was done for the $\nu = 0$ case, a rough sketch of the characteristic curves for the $\nu \neq 0$ case will be made. These curves are used to find the values of Ka and γa from the intersect with a circle of constant V. The behavior of the characteristic curves has been derived near cutoff, $\gamma a = 0$. Before we can sketch the curves, the behavior of the characteristic curve near $\gamma a = \infty$ has to be investigated.

The Ka asymptote of the $HE_{\nu\mu}$ modes for $\gamma a \to \infty$ are found from Eqs. (11.96) and (11.101) as

$$\frac{J_{\nu-1}(Ka)}{KaJ_{\nu}(Ka)} = 0 \tag{11.113}$$

Thus,

$$J_{\nu-1}(Ka) = 0 \tag{11.114}$$

Finally, the Ka asymptote for the $EH_{\nu\mu}$ modes for $\gamma a \to \infty$ are found from Eqs. (11.97) and (11.101) as

$$J_{\nu+1}(Ka) = 0 \tag{11.115}$$

A summary of the various Bessel function conditions obtained in Section 11.2.5 for the limits $\gamma a \to 0$ and $\gamma a \to \infty$ are tabulated in Table 11.1.

In summary, graphs of the characteristic equation with $\nu = 0$ are drawn in Fig. 11.16. Figure 11.17 is an example of a graph with $\nu = 1$ for the $HE_{1\mu}$ modes. Another example, for the $HE_{3\mu}$ and $EH_{1\mu}$ modes, is sought in Problem 11.3. Intersections of the characteristic curves in Ka–γa coordinates with a circle of radius V, where $V^2 = (Ka)^2 + (\gamma a)^2$, are solutions of Ka and γa for the wavelength of operation in a given fiber.

As shown in Fig. 11.17, the intersection of V with the HE_{11} curve can exist from $V = 0$ to infinity. At $V = 2.4$, the intersections with the TE_{01}, TM_{01}, and HE_{21} curves start to exist. Only the HE_{11} mode is excited in the region $0 < V < 2.4$, so that for

Table 11.1 Bessel functions whose roots give Ka values at either $\gamma a \to 0$ or $\gamma a \to \infty$, for modes in a step-index fiber

	Types of Modes		
Limits	$\mathrm{TE}_{0\mu}$, $\mathrm{TM}_{0\mu}$	$HE_{\nu\mu}$	$EH_{\nu\mu}$
Near cutoff, $\gamma a \to 0$	$J_0(Ka) = 0$	$\nu = 1$, $J_1(Ka) = 0$ including $Ka = 0$ $\nu \geq 2$, $J_{\nu-2}(Ka) = 0$ except $Ka = 0$	$J_{\nu}(Ka) = 0$ except $Ka = 0$
Far from cutoff, $\gamma a \to \infty$	$J_1(Ka) = 0$	$J_{\nu-1}(Ka) = 0$	$J_{\nu+1}(Ka) = 0$

$V < 2.4$, the fiber is a single-mode fiber. Each mode can be excited over a region starting from its cutoff value V_c and extending to infinite V. Inside the range of a given mode, the larger V is, the larger γa is. A large value of γ means the field is better confined in the core and the radiation loss of the fiber due to bending decreases.

As the radius V is increased, the number N of intersections with the characteristic curves is increased. N is plotted as a function of V in Fig. 11.18.

Numerical solutions of the transcendental equations of the characteristic curves based on the weakly guiding approximation and the V circle are shown in Fig. 11.19 [19]. The solutions are shown in universal coordinates in Fig. 11.19. The vertical axis is the normalized propagation parameter given by

$$b = (\gamma a/V)^2 \tag{11.116}$$

and the horizontal axis is V. It should be noted that the (m, μ) values are for the $\mathrm{LP}_{m\mu}$ modes, but they can be converted into hybrid modes as explained in Example 11.1. An advantage of taking b as a variable is that b has a finite range from 0 to 1. This can be seen by referring to the triangle in Fig. 11.20, which is drawn from Eq. (11.76). The quantity b is expressed as $b = \sin^2\theta$. Using Eq. (11.76), the value of Ka is calculated from b as

$$Ka = \sqrt{1 - b}V \tag{11.117}$$

Using Eq. (11.37), the propagation constant can also be calculated from b as

$$\beta^2 = (n_1 k)^2[1 - 2\Delta(1 - b)] \tag{11.118}$$

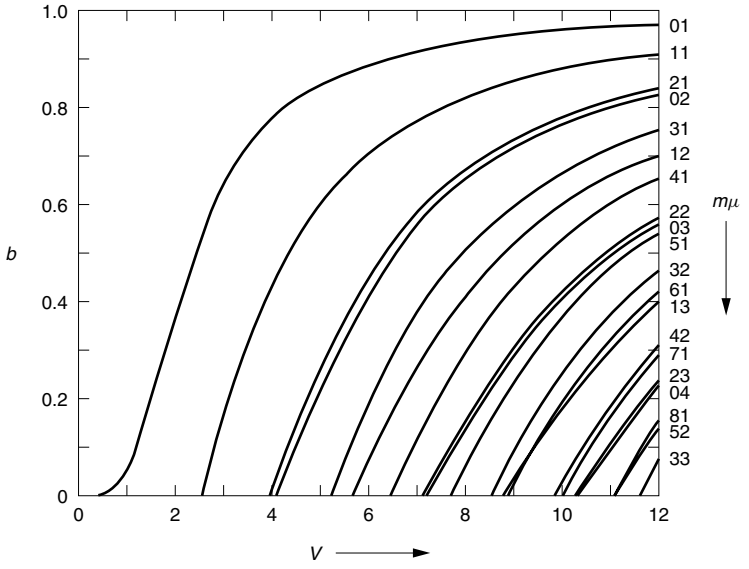


Figure 11.19 Normalized propagation parameter b as a function of the normalized frequency V for $\mathrm{LP}_{m\mu}$ modes, which consists of $HE_{m+1\mu}$ and $EH_{m-1\mu}$ for $m > 1$; and $HE_{2\mu}$, $TM_{0\mu}$, $TE_{0\mu}$ for $m = 1$; and $HE_{1\mu}$ for $m = 0$. (After D. Gloge [20].)

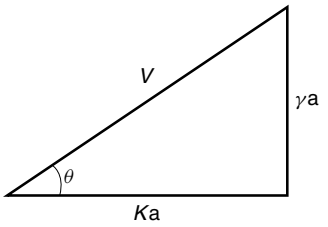


Figure 11.20 Relationship between γa, Ka, and V.

where

$$\Delta = \frac{n_1 - n_2}{n_1} \tag{11.119}$$

and

$$2n_1 \doteq n_1 + n_2$$

Applying the binomial expansion to Eq. (11.118) gives the approximation

$$\beta = n_1 k[1 - \Delta(1 - b)] \tag{11.120}$$

Example 11.2

(a) If a laser light with wavelength $\lambda = 0.63$ μm is incident into an optical fiber that was originally designed to be a single-mode fiber ($V = 2.0$, $a = 1.5$ μm, $n_1 = 1.458$) at wavelength $\lambda = 1.55$ μm, how many modes will the fiber support?

(b) List the mode(s) in hybrid mode designation. (The parameters ($m\mu$) in the figure are linear polarization designations).

(c) Conversely, if a laser light with wavelength $\lambda = 1.55$ μm is incident into an optical fiber that was orginally designed to be a single-mode fiber ($V = 2.0$, $a = 1.5$ μm, $n_1 = 1.458$) at wavelength $\lambda = 0.63$ μm, how many modes will the fiber support? What is (are) the value(s) of the propagation constant(s)?

(d) What is the value of γ, which represents the evanescent field in the cladding?

Solution

(a)

$$2.0 = \frac{2\pi}{1.55} a \sqrt{n_1^2 - n_2^2}$$

$$V = \frac{2\pi}{0.63} a \sqrt{n_1^2 - n_2^2}$$

From the ratio of the two equations, the value of V is

$$V = 4.92$$

From Fig. 11.19, four LP modes are supported.

(b) The intercepts between the mode lines and $V = 4.92$ in Fig. 11.19 give the modes. The parameters $(m\mu)$ of Fig. 11.19 are for the $\mathrm{LP}_{m\mu}$ modes and their corresponding expressions in hybrid modes for $m = 0$, and $m = 1$ are explained in the labeling of the graph. The results are tabulated as follows:

$(m\mu)$	(01)	(11)	(21)	(02)
$\mathrm{LP}_{m\mu}$ designation	LP_{01}	LP_{11}	LP_{21}	LP_{02}
Hybrid mode designation	HE_{11}	HE_{21}	HE_{31}	HE_{12}
		TE_{01}	EH_{11}	
		TM_{01}		

(c)

$$V = 0.81$$

From the graph, only one mode HE_{11} is supported with $b = 0.04$. Next, the propagation constant is

$$\beta = n_1 k[1 - \Delta(1 - b)]$$

The value of Δ has to be found to calculate β:

$$\begin{aligned} 2.0 &= \frac{2\pi}{0.63}(1.5)(1.458)\sqrt{2\Delta} \\ \Delta &= 0.0042 \\ \beta &= (1.458)\frac{2\pi}{1.55}[1 - 0.0042(1 - 0.04)] \\ &= 5.9 \text{ rad/}\mu\text{m} \end{aligned}$$

(d) Since $b = (\gamma a/V)^2$ is reduced to 0.04, the decay of the field in the cladding is small and is therefore subject to leakage. More specifically,

$$\gamma = \sqrt{b}\frac{V}{a} = (0.2)\frac{0.81}{1.5} = 0.108 \text{ rad/ } \mu\text{m} \qquad \square$$

11.3 FIELD DISTRIBUTIONS INSIDE OPTICAL FIBERS

Field expressions have already been given by Eqs. (11.62) through (11.69). These expressions, however, are too cumbersome from which to draw field patterns. Attempts will be made to simplify these expressions by some legitimate approximations [17]. With the approximations of Eq. (11.88), Eq. (11.62) at $z = 0$ can be rewritten as

$$E_r = E\left[J'_\nu(Kr) + j\frac{B}{A}\frac{\omega\mu}{n_1 k}\cdot\frac{\nu}{Kr}J_\nu(Kr)\right]e^{j\nu\phi} \tag{11.121}$$

where

$$E = \frac{jAn_1k}{K} = |E|e^{j(\pi/2)}$$

The factor involving the ratio B/A in the square brackets in Eq. (11.121) is simplified as

$$j\frac{B}{A}\frac{\omega\mu}{\omega\sqrt{\epsilon_1\mu}} = j\frac{B}{A}\eta_1 \tag{11.122}$$

From Eq. (11.91), the value of Eq. (11.122) is either $+1$ or -1 depending on whether the mode is a HE or EH mode. Thus, with Eq. (11.92), the real part of E_r for the HE modes becomes

$$E_r = |E|J_{\nu-1}(Kr)\cos\left(\nu\phi + \frac{\pi}{2}\right) \quad HE \text{ modes} \tag{1.123}$$

In the following, the absolute sign on $|E|$ and $\pi/2$ will be suppressed. And with Eq. (11.94), E_r for the EH modes becomes

$$E_r = -EJ_{\nu+1}(Kr)\cos\nu\phi \quad EH \text{ modes} \tag{11.124}$$

Next, simplified expressions for the magnetic fields will be obtained. Equation (11.64), with the approximation of Eq. (11.88), is written as

$$H_r = \frac{jn_1k}{K}\left[-\frac{\omega\epsilon_1}{n_1k}A\frac{j\nu}{Kr}J_\nu(Kr) + BJ'_\nu(Kr)\right]e^{j\nu\phi} \tag{11.125}$$

The factor in the first term in the square brackets can be rewritten with the help of Eq. (11.122) and H_r becomes

$$H_r = -jH\left[\frac{\nu}{Kr}J_\nu(Kr) + \left(j\frac{B}{A}\eta_1\right)J'_\nu(Kr)\right]e^{j\nu\phi} \tag{11.126}$$

where

$$H = \frac{E}{\eta_1}$$

With Eq. (11.91) for the HE modes and Eq. (11.92), the expression for the real part of H_r becomes

$$H_r = HJ_{\nu-1}(Kr)\sin\nu\phi \quad HE \text{ modes} \tag{11.127}$$

With Eq. (11.91) for the EH modes and Eq. (11.94), the expression for H_r is

$$H_r = HJ_{\nu+1}(Kr)\sin\nu\phi \quad EH \text{ modes} \tag{11.128}$$

The rest of the field components can be found in a similar manner. They are summarized as follows. From Eqs. (11.40), (11.41), and (11.91), E_z and H_z for both $HE_{\nu\mu}$

and $EH_{\nu\mu}$ modes are

$$\begin{aligned} E_z &= AJ_\nu(K_{\nu\mu}r)\cos\nu\phi \\ H_z &= \pm\frac{A}{\eta_1}J_\nu(K_{\nu\mu}r)\sin\nu\phi \end{aligned} \tag{11.129}$$

For $HE_{\nu\mu}$ modes only,

$$\begin{aligned} E_r &= EJ_{\nu-1}(K_{\nu\mu}r)\cos\nu\phi \\ E_\phi &= -EJ_{\nu-1}(K_{\nu\mu}r)\sin\nu\phi \end{aligned} \tag{11.130}$$

and from Eqs. (11.62) to (11.65),

$$\begin{aligned} H_r &= -\frac{1}{\eta_1}E_\phi \\ H_\phi &= \frac{1}{\eta_1}E_r \end{aligned}$$

For $EH_{\nu\mu}$ modes only,

$$\begin{aligned} E_r &= -EJ_{\nu+1}(K_{\nu\mu}r)\cos\nu\phi \\ E_\phi &= -EJ_{\nu+1}(K_{\nu\mu}r)\sin\nu\phi \\ H_r &= -\frac{1}{\eta_1}E_\phi \\ H_\phi &= \frac{1}{\eta_1}E_r \end{aligned} \tag{11.131}$$

The expressions for the $\mathrm{TE}_{0\mu}$ and $\mathrm{TM}_{0\mu}$ modes are calculated later in Eqs. (11.136) through (11.138).

11.3.1 Sketching Hybrid Mode Patterns

Let us draw a few mode patterns using the field expressions in Eqs. (11.130) and (11.131). The first drawing is the HE_{11} mode, which is the simplest and also the most important mode, as it is the only mode in a single-mode fiber. Equation (11.130) with $\nu = 1$ gives

$$\begin{aligned} E_{r1} &= EJ_0(K_{11}r)\cos\phi \\ E_{\phi 1} &= -EJ_0(K_{11}r)\sin\phi \\ H_{r1} &= -\frac{1}{\eta_1}E_{\phi 1} \\ H_{\phi 1} &= \frac{1}{\eta_1}E_{r1} \end{aligned} \tag{11.132}$$

The top row of Fig. 11.21a illustrates the HE_{11} mode. Field lines of E and H at strategic points are drawn first, in the left and center diagrams. Note that the positive

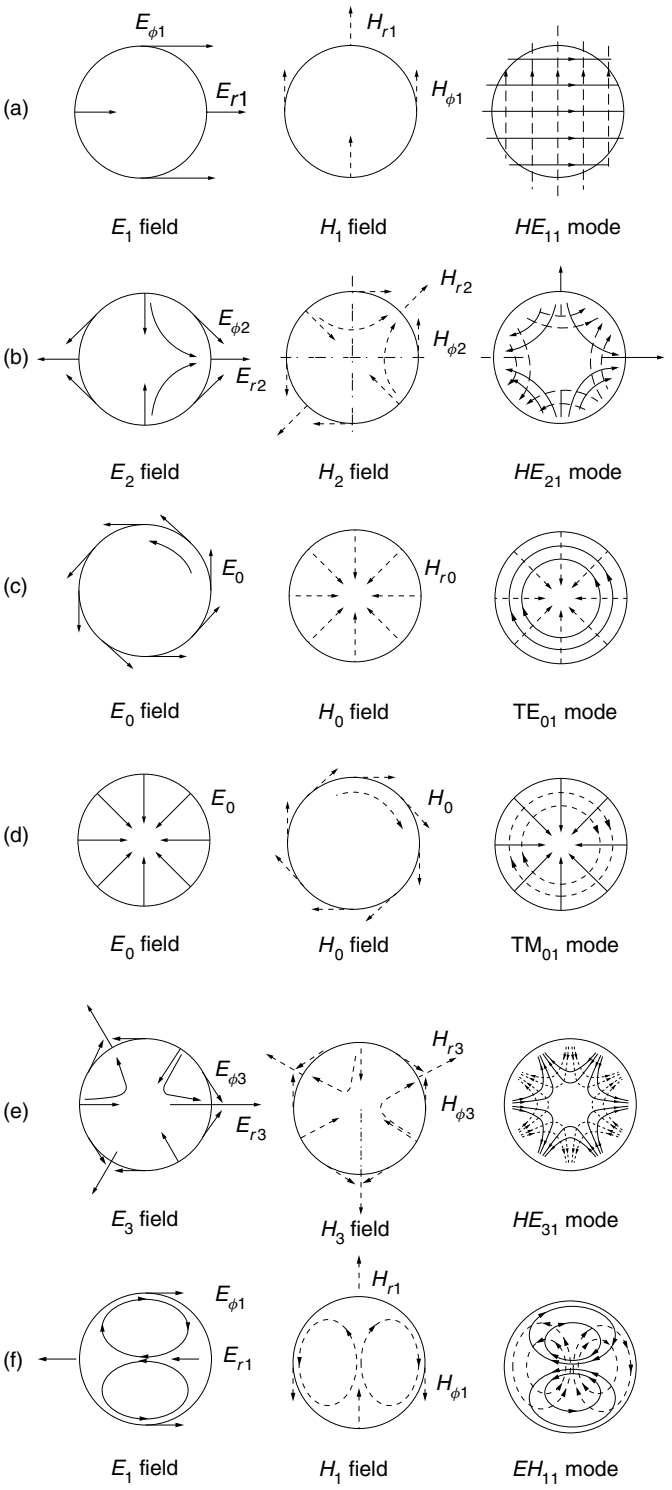


Figure 11.21 Mode patterns in a step-index fiber. The z axis is out of the page. The solid lines represent the E field and the dotted lines represent the H field.

directions of E_r and E_ϕ are in the directions of increasing r and ϕ, respectively. In the left diagram, the extrema of E_{r1} and $E_{\phi 1}$ are drawn, and in the center diagram those of H_{r1} and $H_{\phi 1}$ are drawn. The lines of the extrema are connected to complete the field pattern. The completed drawings are shown on the right-hand side. In the cylindrical coordinate system, $\hat{\mathbf{r}} \times \hat{\boldsymbol{\phi}} = \hat{\mathbf{z}}$, where $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\phi}}$ and $\hat{\mathbf{z}}$ are unit vectors in the r, ϕ, and z directions. The positive $\hat{\mathbf{z}}$ direction is out of the page. That is why all Poynting vectors $\mathbf{E} \times \mathbf{H}$ point out of the page. A special characteristic of the HE_{11} mode pattern is that the E and H field lines make a square grid. There are two extrema in the ϕ direction and there is only one extremum in the r direction.

Next, the HE_{21} mode is drawn. With $\nu = 2$ and $\mu = 1$, Eq. (11.130) becomes

$$\begin{aligned} E_{r_2} &= EJ_1(K_{21}r)\cos 2\phi \\ E_{\phi_2} &= -EJ_1(K_{21}r)\sin 2\phi \\ H_{r_2} &= -\frac{1}{\eta_1}E_{\phi_2} \\ H_{\phi_2} &= \frac{1}{\eta_1}E_{r_2} \end{aligned} \tag{11.133}$$

Drawings of the HE_{21} mode are shown in Fig. 11.21b. The HE_{21} mode has four extrema in the ϕ direction and one extremum in the r direction. By connecting these extrema smoothly, the field patterns are completed.

Next, an attempt will be made to draw the patterns for the TE_{01} and TM_{01} modes. In these cases, A or B becomes zero. Equations (11.130) and (11.131) were derived using the relationship Eq. (11.121), which cannot be used for $A = 0$. The best approach in this case is to start from the original differential equations.

With $\nu = 0$, Eqs. (11.31) and (11.32) change to

$$\Phi'' = 0 \tag{11.134}$$

$$F'' + \frac{1}{r}F' + [(nk)^2 - \beta^2]F = 0 \tag{11.135}$$

The solution of Eq. (11.134) is a constant* and that of Eq. (11.135) is the zero-order Bessel function and the solution becomes

$$\begin{aligned} E_z &= A'J_0(Kr) \\ H_z &= B'J_0(Kr) \end{aligned} \tag{11.136}$$

The other components will be found from Eqs. (11.62) through (11.65) with $A = A'$ and $B = B'$. With the TE_{01} mode, $A' = 0$ and $\nu = 0$, and the components are

$$\begin{aligned} E_{r0} &= 0 \\ E_{\phi 0} &= \frac{j\omega\mu}{K}B'J_1(Kr) \\ H_{r0} &= -\frac{j\beta}{K}B'J_1(Kr) \\ H_{\phi 0} &= 0 \end{aligned} \tag{11.137}$$

* The solution $\Phi = a\phi$ cannot satisfy Eq. (11.39) unless $a = 0$.

Both the E_ϕ and H_r components are constant with respect to ϕ, and the mode pattern is cylindrically symmetric. The pattern is shown in Fig. 11.21c.

Next, the TM_{01} mode will be calculated in a similar manner. Equation (11.136) with $B' = 0$ is put into Eqs. (11.62) through (11.65). The results are

$$\begin{aligned} E_{r0} &= -\frac{j\beta}{K} A' J_1(Kr) \\ E_{\phi 0} &= 0 \\ H_{r0} &= 0 \\ H_{\phi 0} &= -\frac{j}{K} \omega \epsilon_1 A' J_1(Kr) \end{aligned} \tag{11.138}$$

The mode pattern of the TM_{01} mode is shown in Fig. 11.21d below that of the TE_{01} mode. If E and H are exchanged, TM_{01} is obtained from TE_{01}. They are complementary patterns.

The HE_{31} and EH_{11} that generate an LP_{21} mode (see Problem 11.4) will now be drawn. These modes have the same cutoffs (see Problem 11.3) but have different field patterns. Equation (11.130) for the $HE_{\nu\mu}$ modes with $\nu = 3$ gives

$$\begin{aligned} E_r &= EJ_2(Kr)\cos 3\phi \\ E_\phi &= -EJ_2(Kr)\sin 3\phi \\ H_r &= -\frac{1}{\eta_1} E_\phi \\ H_\phi &= \frac{1}{\eta_1} E_r \end{aligned} \tag{11.139}$$

The field pattern of HE_{31} is shown in Fig. 11.21e. Equation (11.131) for the $EH_{\nu\mu}$ modes with $\nu = 1$ and $\mu = 1$, gives

$$\begin{aligned} E_r &= -EJ_2(Kr)\cos\phi \\ E_\phi &= -EJ_2(Kr)\sin\phi \\ H_r &= -\frac{1}{\eta_1} E_\phi \\ H_\phi &= \frac{1}{\eta_1} E_r \end{aligned} \tag{11.140}$$

The field pattern for EH_{11} is drawn in Fig. 11.21f at the bottom of the figure. Both HE_{31} and EH_{11} have one extremum in the r direction, but HE_{31} has six extrema in the ϕ direction, while EH_{11} has two extrema in the ϕ direction. That is, there are 2ν extrema in the ϕ direction and μ extrema in the r direction with either pattern.

11.3.2 Sketching Linearly Polarized Mode Patterns

As mentioned in Section 11.2.5.2, the combination of the $HE_{m+1\mu}$ and $EH_{m-1\mu}$ mode generates the $LP_{m\mu}$ mode.

We will demonstrate graphically that the addition of the HE_{21} and TM_{01} mode patterns in Figs. 11.21b and 11.21d generates the LP_{11} mode. While the x components of the E fields cancel each other, the y components of the fields add together and the E field of the LP_{11} mode is

$$E_x = 0$$
$$E_y = 2E$$

Similarly, the x components of H add, while the y components cancel each other, and the H field of the LP_{11} mode becomes

$$H_x = 2H$$
$$H_y = 0$$

Because $E_x = 0$ and $H_y = 0$, the new mode created by the two modes is indeed a linearly polarized (LP) mode. Figure 11.22 shows the field pattern of the LP_{11} mode. The shaded area represents the intensity of the Poynting vector.

It will now be shown that cancellation of the E_x and H_y fields takes place between $HE_{\nu+1\mu}$ and $EH_{\nu-1\mu}$ for any value of ν (excluding $\nu < 2$). Referring to Fig. 11.23, the E_r and E_ϕ components are decomposed into E_x and E_y components as

$$\begin{aligned} E_x &= E_r \cos\phi - E_\phi \sin\phi \\ E_y &= E_r \sin\phi + E_\phi \cos\phi \end{aligned} \tag{11.141}$$

The expressions for E_r and E_ϕ from Eq. (11.130) are inserted into Eq. (11.141) to obtain the fields for the $HE_{\nu+1\mu}$ mode, and likewise, those of Eq. (11.131) are inserted into Eq. (11.141) to obtain the fields for the $HE_{\nu-1\mu}$ mode. For example, the x

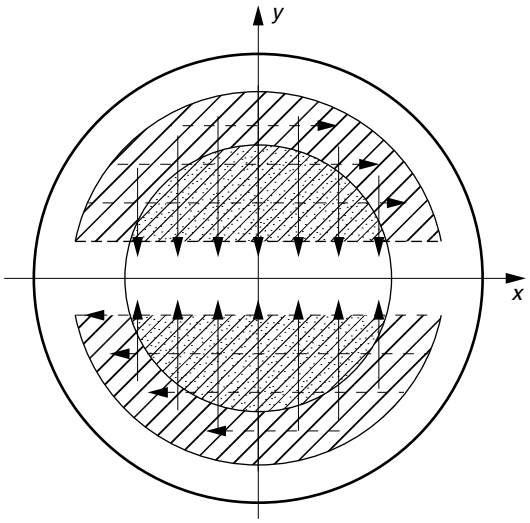


Figure 11.22 Field pattern of the LP_{11} mode.

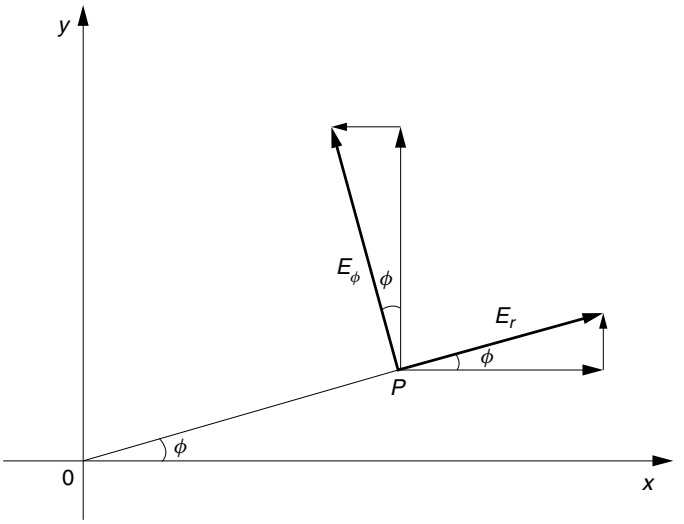


Figure 11.23 Relationship between components in cylindrical and rectangular coordinates.

component of the $HE_{\nu+1\mu}$ mode, E_x^{HE}, is given by

$$\begin{aligned} E_x^{HE} &= EJ_\nu(Kr)[\cos(\nu+1)\phi\cos\phi + \sin(\nu+1)\phi\sin\phi] \\ &= EJ_\nu(Kr)\cos\nu\phi \end{aligned} \tag{11.142}$$

A summary of all the field components is

$$\begin{aligned} E_x^{HE} &= \cos\nu\phi & E_x^{EH} &= -\cos\nu\phi \\ E_y^{HE} &= -\sin\nu\phi & E_y^{EH} &= -\sin\nu\phi \\ H_x^{HE} &= \sin\nu\phi & H_x^{EH} &= \sin\nu\phi \\ H_y^{HE} &= \cos\nu\phi & H_y^{EH} &= -\cos\nu\phi \end{aligned} \tag{11.143}$$

where $EJ_\nu(Kr)$ or $HJ_\nu(Kr)$ is suppressed. The addition of the $HE_{\nu+1\mu}$ and $EH_{\nu-1\mu}$ fields gives the $\mathrm{LP}_{\nu\mu}$ fields. From Eq. (11.143), this addition gives

$$\begin{aligned} E_x^{HE} + E_x^{EH} &= 0 \\ E_y^{HE} + E_y^{EH} &= -2EJ_\nu(Kr)\sin\nu\phi \\ H_x^{HE} + H_x^{EH} &= 2HJ_\nu(Kr)\sin\nu\phi \\ H_y^{HE} + H_y^{EH} &= 0 \end{aligned} \tag{11.144}$$

Thus, the $\mathrm{LP}_{\nu\mu}$ mode has only E_y and H_x components, and neither the E_x nor the H_y component exists. Moreover, E_y and H_x have the same distribution functions. It should be noted that the pattern $(\sin\phi)$ is antisymmetric with respect to ϕ and thus

this LP mode is called the $\mathrm{LP}_{\nu\mu}^{\mathrm{odd}}$ or $\mathrm{LP}_{\nu\mu}^{o}$ mode. Figure 11.24 shows the field pattern of the $\mathrm{LP}_{\nu\mu}$ mode.

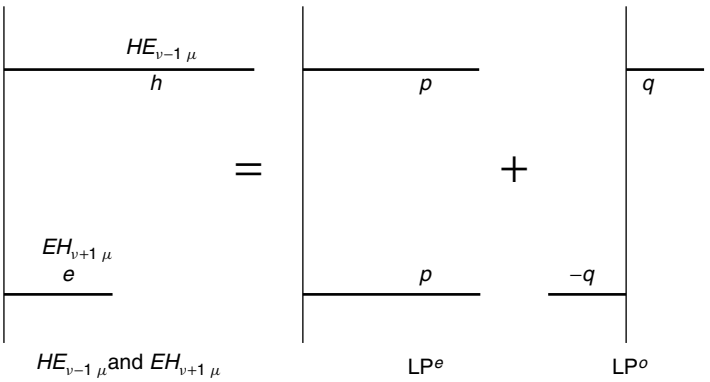
So far, it has always been assumed that both the magnitude and the phase of the $HE_{\nu+1\mu}$ and $EH_{\nu-1\mu}$ modes are identical. A different pattern would be generated, for instance, if the phase of $HE_{\nu+1\mu}$ were shifted π radians from that of $EH_{\nu-1\mu}$. In such a case, the components of the $\mathrm{LP}_{\nu\mu}$ mode from Eq. (11.143) become

$$\begin{aligned} E_x &= 2EJ_\nu(Kr)\cos\nu\phi \\ H_y &= 2HJ_\nu(Kr)\cos\nu\phi \\ E_y &= H_x = 0 \end{aligned} \tag{11.145}$$

Because $E_y = 0$ and $H_x = 0$, this mode also classifies as a linearly polarized mode, but this time the LP mode is called $\mathrm{LP}_{\nu\mu}^{\mathrm{even}}$ or $\mathrm{LP}_{\nu\mu}^{e}$ because of the symmetry of the pattern ($\cos\nu\phi$) with respect to ϕ. With an arbitrary combination of amplitude and phase, the composite modes generated from $HE_{\nu+1\mu}$ and $EH_{\nu-1\mu}$ consist of both $LP_{\nu\mu}^{\mathrm{odd}}$ and $\mathrm{LP}_{\nu\mu}^{\mathrm{even}}$ modes.

Example 11.3 Suppose that an $HE_{\nu-1\mu}$ mode and an $EH_{\nu+1\mu}$ mode are in phase but their amplitudes are different and are given by h and e, respectively. What are the amplitudes of the $\mathrm{LP}_{\nu\mu}^{\mathrm{odd}}$ and $\mathrm{LP}_{\nu\mu}^{\mathrm{even}}$ modes?

Solution The amplitudes h and e can be divided into p, which is the magnitude of LP^o, and q, which is the magnitude of LP^e as



From the diagram

$$\begin{aligned} h &= p + q \\ e &= p - q \end{aligned}$$

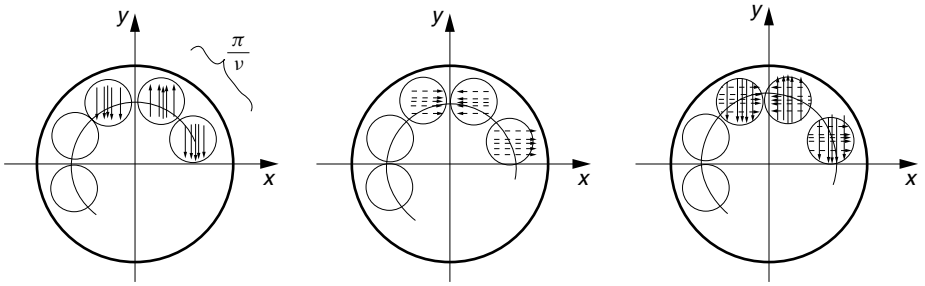


Figure 11.24 Field pattern of $LP_{\nu\mu}$ mode.

Solving for p and q, the magnitudes of LP^e and LP^o are given by

$$p = \frac{h+e}{2}$$

$$q = \frac{h-e}{2} \qquad \square$$

11.4 DUAL-MODE FIBER

A dual-mode fiber transmits both the dominant LP_{01} mode and the next higher order LP_{11} mode. The dual-mode fiber is made by setting the normalized radius V within the range of

$$2.405 < V < 3.832$$

which corresponds to the values between the first zero of $J_0(Ka)$ and the second zero of $J_1(Ka)$. The mode pattern of the LP_{01} mode (HE_{11} mode) is shown in Fig. 11.21a, and the LP_{11} mode is shown in Fig. 11.22.

The two modes in a dual-mode fiber behave as if two separate waves are propagating in one fiber. This feature is especially useful in a fiber sensor interferometer (explained below) because the interferometer is more stable when the two waves propagate in the same fiber rather than in two separate fibers.

Figure 11.25 shows the geometry of a dual-mode fiber strain gauge. A short section (20–30 cm) of dual-mode fiber is spliced to a single-mode fiber at both ends. The relative phase between the LP_{01} and LP_{11} modes shifts as the modes propagate due to the difference in the propagation constants. In order to excite both the LP_{01} and LP_{11} modes, the splices between the single-mode fiber and the dual-mode fiber are off center. At the input end, the E fields of the LP_{01} and LP_{11} modes in the encircled input portion of Fig. 11.25 are in phase, but after a distance of $L_B/2$, say, the E fields of the LP_{01} and LP_{11} modes in the encircled output portion of Fig. 11.25 are out of phase. When this occurs, the output from the single-mode fiber is at a minimum. After another $L_B/2$ of propagation, the output pattern becomes the same as the input pattern.

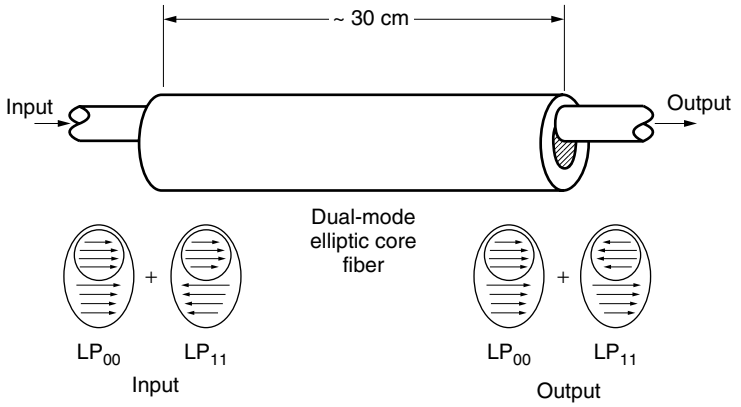


Figure 11.25 A strain gauge made out of a dual-mode fiber. (After F. Blaha.)

The length L_B is called the *coupling length* and is expressed as

$$L_B = \frac{2\pi}{\beta_{01} - \beta_{11}} \tag{11.146}$$

where β_{01} and β_{11} are the propagation constants of the LP_{01} and LP_{11} modes.

For instance, if the length of the dual-mode fiber sensor section is 30 cm and $L_B = 270$ μm, the percentage change in fiber length due to stretching that would be required to vary the light output from the minimum to the maximum is

$$\frac{L_B/2}{L} = \frac{0.135}{300} = 4.5 \times 10^{-4}$$

If 1% of the change from the maximum to minimum is measurable, then the sensitivity of such a sensor is 4.5×10^{-6}. The sensor employs an elliptic core dual-mode fiber, because, as mentioned in Section 11.1.5, the elliptic core fiber preserves the state of polarization of the light. In this way, fluctuations of the output power due to fluctuations in the relative directions of polarization of the LP_{01} and LP_{11} modes can be prevented.

Next, a tunable optical coupler constructed from dual-mode fiber will be described. Often, a slight but adjustable amount of light needs to be tapped off a main trunk line with minimum disturbance. The trunk line is single-mode fiber. Figure 11.26 shows the geometry. The coupler consists of a mode converter and a coupling section. In the mode converter section, either a set of teeth or rods creates a periodic microbending disturbance whose period is equal to the coupling length L_B. A portion of the energy of the LP_{01} mode in the trunk line is converted into that of the LP_{11} mode depending on the degree of the disturbance. Both LP_{01} and LP_{11} modes enter the coupling section. Even though the tap-off fiber is a normal dual-mode fiber, the main trunk line is gradually tapered from the dual-mode into a single-mode fiber in the coupling section. The tap-off of a particular wavelength is possible by using a longer converter section [22].

Finally, an application of the dual-mode fiber to a dispersion compensator will be described. Another feature of the dual-mode fiber is that the LP_{11} mode has an exceptionally large negative dispersion parameter when it is operated near its cutoff,

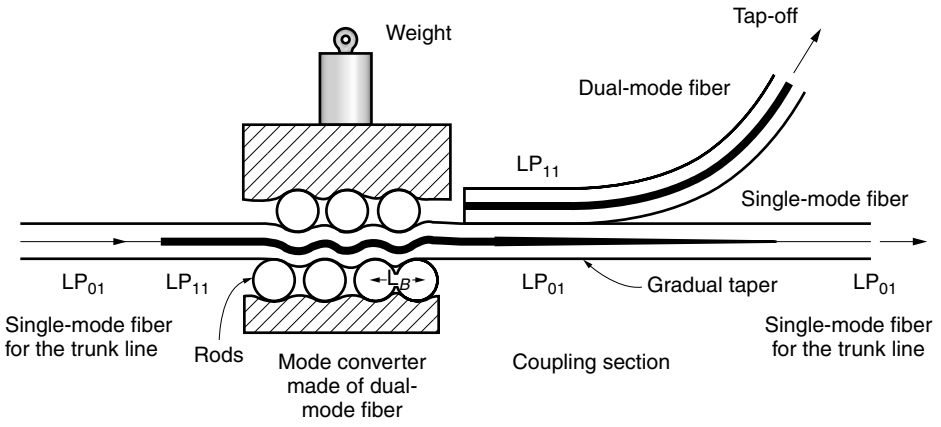


Figure 11.26 Tunable optical coupler for tapping off the trunk line.

and this can be used as a dispersion compensator. A dispersion compensator made out of a dual-mode fiber can provide the same amount of dispersion as a step-index single-mode fiber but with opposite sign, thereby establishing a zero-dispersion light transmission system. Figure 11.27a shows an experimental setup for demonstrating the effectiveness of the dual-mode fiber dispersion compensator. One of the 3-dB coupler output ports monitors the pulse shape after transmission through 50 km of single-mode step-index fiber with dispersion parameter $D = +15.8$ ps/(km · nm). The other output port of the 3-dB coupler monitors the signal that has made a round trip through 720 m of elliptic core dual-mode fiber with $D = -600$ ps/(km · nm) over and above the 50 km of step-index single-mode fiber. A mode converter is inserted between the single-mode and dual-mode fibers to convert the LP_{01} mode into the LP_{11} mode.

Near the reflecting mirror, a 90° polarization rotator is installed so as to remove the effect of the birefringence of the compensator's elliptic core fiber. As shown in Fig. 11.18, the LP_{11} mode consists of TE_{01}, TM_{01}, and HE_{21} modes. Their propagation constants are very close but slightly different from each other and could be a source of mode dispersion. This problem, however, is removed by using the elliptic core mode converter as in this system.

Figure 11.27b illustrates the effectiveness of the dispersion compensator [23]. The top figure shows the waveform of the input light pulses. The middle figure shows the waveform distorted by the dispersion of the 50-km step-index single-mode fiber. The pulse shape is unrecognizable. The bottom figure shows the waveform recovered by going through the 720×2-m dual-mode dispersion compensator fiber. A more quantitative description of the dispersion compensator will be provided in Section 11.6 and Section 11.7.

11.5 PHOTOIMPRINTED BRAGG GRATING FIBER

When a germanium-doped (GeO_2-doped) fiber is exposed to a high-intensity ultraviolet light, the glass is bleached and its index of refraction is permanently altered. The amount of change in the refractive index varies from 10^{-5} to 10^{-3} [24,25]. Such a phenomenon is called the *photosensitivity* of the GeO_2-doped fiber.

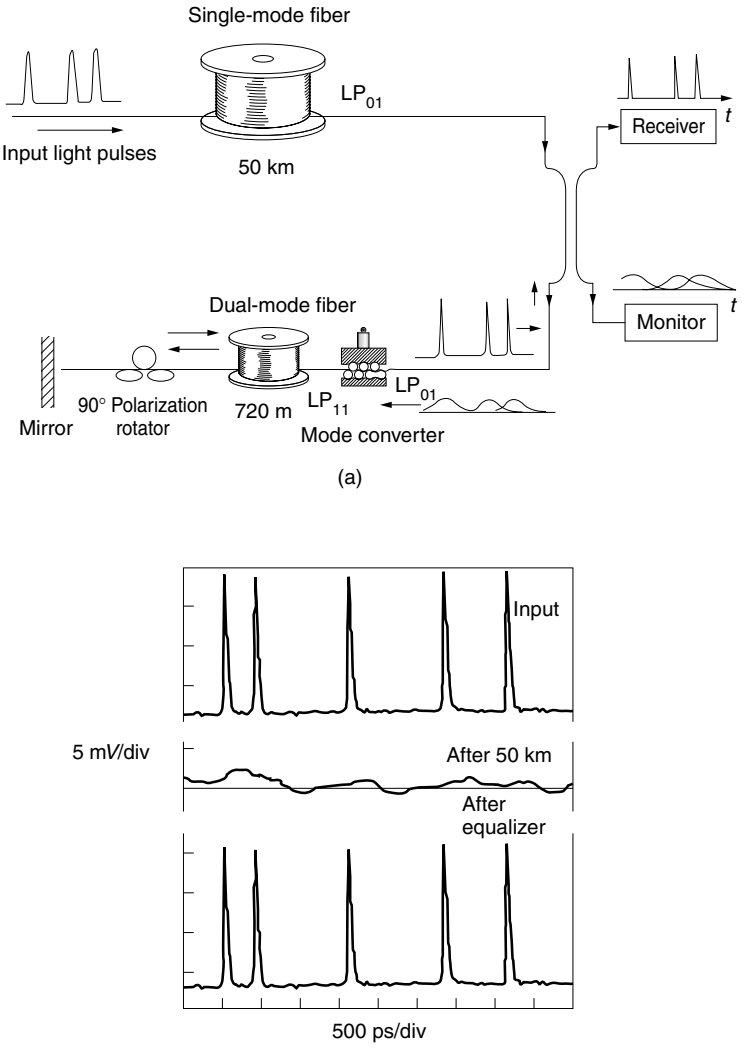


Figure 11.27 Dispersion compensation by a dual-mode fiber. (a) Experimental setup to demonstrate the effectiveness of the dual-mode fiber dispersion compensator. (b) Pulse shapes before and after the compensation. (Courtesy of C. D. Poole et al. [23].)

Photobleaching, however, does not take place in pure silica glass. A fiber consisting of a germanium-doped core surrounded by pure fused silica cladding is ideal for writing a Bragg grating reflector inside the fiber core. Various applications of the fiber Bragg grating have been explored in the fields of fiber lasers, optical communication devices, and fiber sensors.

11.5.1 Methods of Writing Photoinduced Bragg Gratings in an Optical Fiber

The Bragg gratings can be written either by internal or external illumination of the bleaching light.

11.5.1.1 Internal Writing

A high-power ($\sim$GJ/cm^2) visible range laser light, for example, argon ion laser light at 488 nm, is injected into a single-mode germanium-doped core fiber terminated with a mirror, as shown in Fig. 11.28. The standing wave inside the optical fiber photoimprints a Bragg grating with the same pattern as the standing wave. The merit of this method is its simplicity, but the drawbacks are that the same pattern is generated throughout the fiber and the period of the grating is changed only by changing the source. Moreover, it requires a high-intensity source.

11.5.1.2 Holographic Writing

A UV light interference pattern [24] externally illuminates a germanium-doped optical fiber, as shown in Fig. 11.29. The interference pattern is formed by an excimer laser operating at a wavelength of 249 nm. The geometry of the arrangement is designed to confine the interference pattern to the plane of the fiber. The UV-absorbing polymer coating of the optical fiber has to be stripped and recoated. Alternatively, it is also possible to write the grating as the fiber is drawn in the tower just before the polymer coating is applied if the laser source is pulsed with a pulse width of about 20 ns [26].

The required power for writing by means of UV light reduces to 1 kJ/cm^2, which is one-millionth of what is required for writing by means of visible light.

The photosensitivity can be augmented by loading the fiber with H_2. The fiber is placed in pressurized H_2 gas at an elevated temperature (250–400°C) for weeks. The change in the index of refraction reaches as high as $\Delta = 0.5\%$, which is comparable to the refractive index step Δ between the core and cladding of a step-index fiber [27]. Alternatively, the so-called flame brushing method [28], which repeatedly brushes the fiber by a flame fueled by hydrogen, augments the photosensitization almost as much as the H_2 gas immersion method. Even though the loss of the fiber is increased by hydrogen loading, the length of the fiber grating is normally less than a few centimeters, and the loss due to hydrogen exposure is not significant.

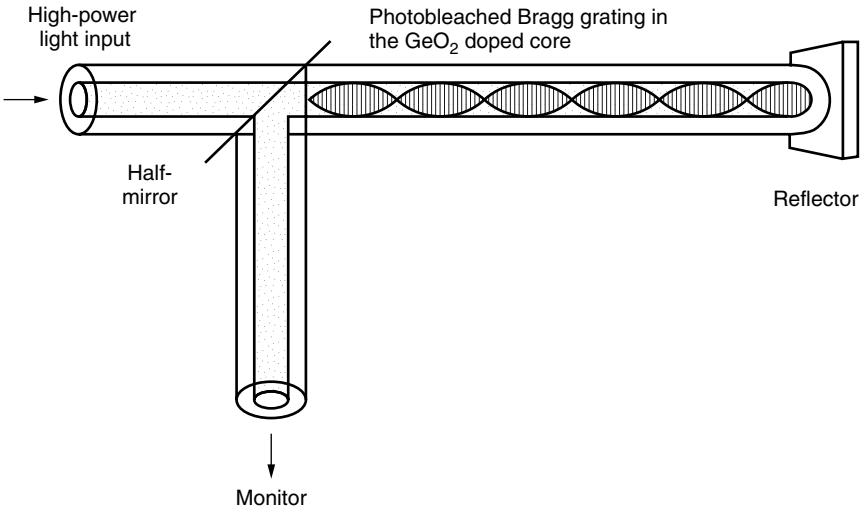


Figure 11.28 Arrangement for internally writing a Bragg grating in the fiber core by generating a standing wave in the core.

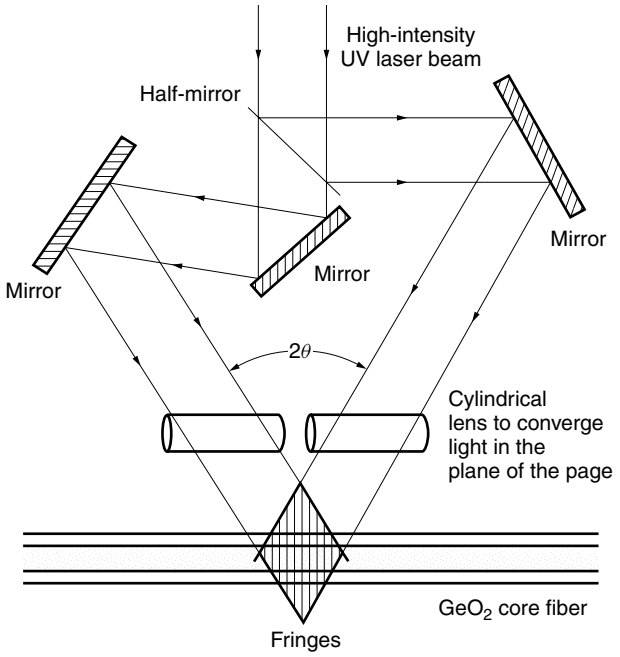


Figure 11.29 An interferometric arrangement for externally writing a Bragg grating in the fiber core.

11.5.1.3 Point-by-Point Writing

A UV laser beam is focused onto the photosensitive optical fiber. After each irradiation, the fiber is translated in the axial direction of the fiber by a computer controlled microtranslation stage. The grating is written one fringe at a time. The advantages of this method are that a laser beam of high intensity is obtainable because the beam is focused; and there is freedom in the selection of the shape, intensity, and period of the written pattern [29].

11.5.1.4 Writing by a Phase Mask

A prefabricated phase mask is used to imprint the pattern onto the optical fiber either in contact or near contact with the fiber. The normal incident UV laser beam bleaches the optical fiber according to the phase mask pattern. The advantages of this method are the flexibility in the patterns of the written grating and the easier handling of the writing apparatus. Figure 11.30 shows a photograph of a photoinduced Bragg grating written on the core of a D-fiber [30].

11.5.2 Applications of the Photoinduced Bragg Gratings in an Optical Fiber

The photoinduced Bragg grating acts as a wavelength-selective component in some fiber lasers, such as the one shown in Fig. 14.67d in Chapter 14. Besides the simplicity of the construction, a special feature of this fiber laser is that the wavelength of the laser can be shifted by simply stretching the fiber laser including the intracore Bragg reflector by means of a piezoexpander. This is possible because the change in the resonance wavelength of the Bragg reflector tracks that of the laser cavity [31].

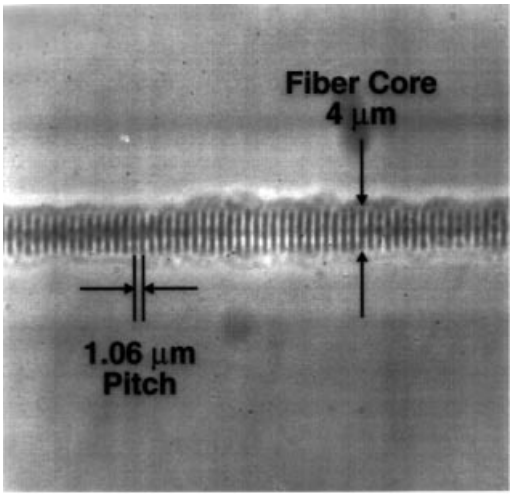


Figure 11.30. A photograph of a photoinduced Bragg grating written on the core of a D fiber. A phase contrast microscope was used. (Courtesy of B. Malo et al. [30].)

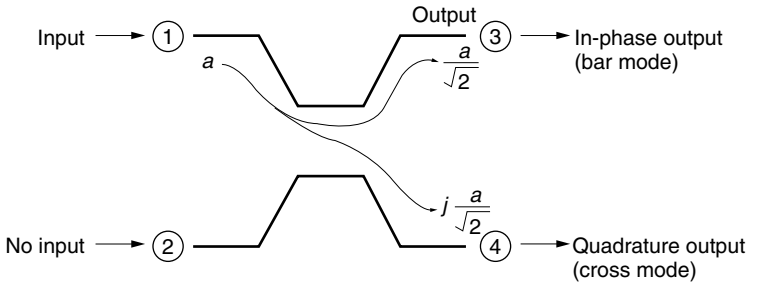
Another application of the photoinduced Bragg grating is as a band-stop optical filter. The rejection bandwidth of the apodized (see Section 1.4.2) Bragg reflector is approximately 0.2 nm with a rejection of 40 dB [32]. The transmission loss outside the stop band is less than 0.01 dB.

The demand for band-pass filters is higher than for band-stop filters. However, by combining the Bragg grating fiber with other optical devices, the band-stop characteristics can be converted into the complement characteristics of a band-pass filter.

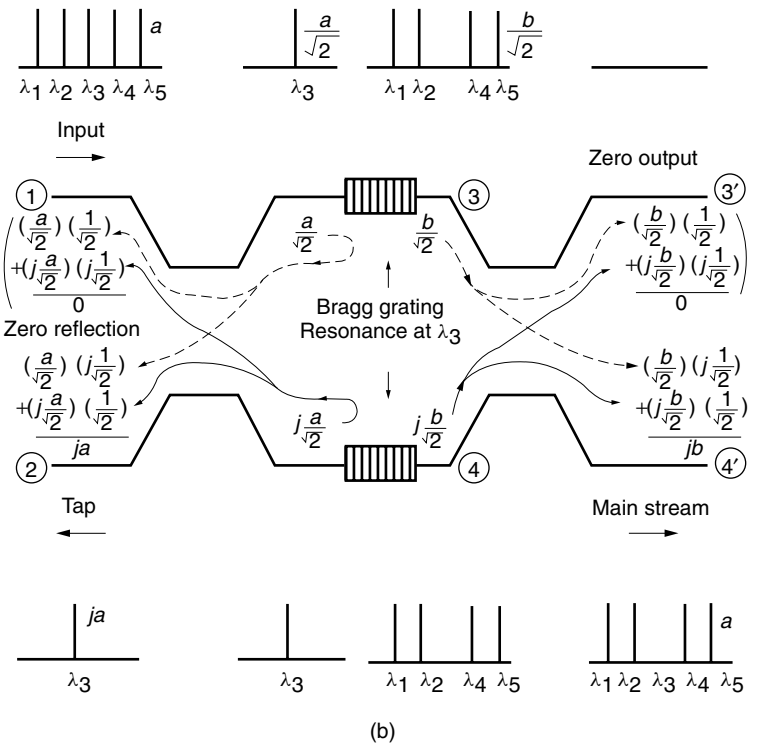
For example, one way of achieving this conversion is by adding a 50–50 (or 3 dB) Mach–Zehnder fiber coupler [33]. A demultiplexer system containing a Mach–Zehnder coupler and photoinduced Bragg gratings is shown in Fig. 11.31. As this is an interesting application of fiber gratings, the function of this system will be explained in detail, starting with the ideal Mach–Zehnder coupler shown in Fig. 11.31a. Note that the power ratio is $1/2$ but the amplitude ratio is $1/\sqrt{2}$.

Let us say the amplitude of the incident light is a and is applied only to port ①. The light power is equally split between ports ③ and ④, and the amplitudes are $a/\sqrt{2}$ each. In addition to the light being split, there is a $\pi/2$-radian phase difference between outputs ③ and ④. If output ③ that has travelled from port ① to port ③ taking the route straight through the coupling section (called the bar mode) is $a/\sqrt{2}$, then output ④ that has traveled from port ① to port ④ taking the route across the coupling section is $ja/\sqrt{2}$ (called the cross mode).

Next, a 100% Bragg grating reflector at λ_3 is installed in both the ③ and ④ arms, as shown in Fig. 11.31b. Wavelength division multiplexed (WDM) light containing wavelengths λ_1, λ_2, λ_3, λ_4, and λ_5 with identical amplitudes a is incident into port ①. Only the λ_3 component is reflected back toward ports ① and ②. The light reflected back toward port ① is considered first. The light reflected from port ③ to ① is the bar mode and $(a/\sqrt{2})(1/\sqrt{2}) = a/2$. Reflection from port ④ to ① is the cross mode and $(ja/\sqrt{2})(j/\sqrt{2}) = -a/2$. Thus, reflections from ③ and ④ cancel each other in port ① and there is no reflected light at port ①.



(a)



(b)

Figure 11.31 Fiber demultiplexer using photoimprinted Bragg reflectors. (a) Principle of the Mach–Zehnder interferometric fiber coupler. (b) Fiber demultiplexer.

Next, the light reflected back toward port ② is considered. The reflected light from ③ to ② is the cross mode and $(a/\sqrt{2})(j/\sqrt{2}) = ja/2$. The reflected light from ④ to ② is the bar mode and $(ja/\sqrt{2})(1/\sqrt{2}) = ja/2$. The reflections from ③ and ④ interfere constructively in port ② and the output from port ② is ja. In conclusion, the light power at λ_3 is fully reflected back to ② but none to ①.

Next, the paths of the rest of the spectra that were not reflected by the Bragg grating reflectors in arms ③ and ④ are considered. Additional 3-dB Mach–Zehnder couplers

are connected to ports ③ and ④ as shown in Fig. 11.31b. The contribution to ③′ from ③ is the bar mode and $(b/\sqrt{2})(1/\sqrt{2}) = b/2$; and that from ④ to ③′ is the cross mode and $(ja/\sqrt{2})(jb/\sqrt{2}) = -b/2$. Thus, the contributions to ③′ cancel each other and no output light is present at ③′. Similarly, the contribution from port ③ to ④′ is the cross mode and $b/\sqrt{2}(j/\sqrt{2}) = jb/2$; and that from ④ to ④′ is the bar mode and $(jb/\sqrt{2})(1/\sqrt{2}) = jb/2$. Thus, contributions from ③ and ④ add constructively and the output from ④′ is jb. Full power comes out of port ④′ and no power comes out of port ③′.

In conclusion, the Bragg grating fiber combined with the Mach–Zehnder interferometer performs the function of wavelength-selective tapping in the WDM stream. The same setup can be also used for a wavelength-selective insertion into a WDM channel (see Problem 11.8).

If the couplers attached to the output of ports ③ and ④ are replaced by light absorbers, then the device becomes an ordinary band-pass filter.

Another important application of the photoimprinted Bragg grating is in the field of fiber sensing. The intracore Bragg grating can be used for sensing strain, vibration, and temperature. The presence of strain or vibration will perturb the period of the Bragg grating and will cause a shift of the resonant wavelength. A perturbation occurs in the presence of a temperature fluctuation, which alters the refractive index of the Bragg grating and shifts the resonant wavelength. One advantage of the Bragg grating sensor over other fiber sensors is that the quantity being measured is a wavelength shift rather than a change in the received power level, and significantly better accuracy and signal-to-noise level are obtainable.

Another advantage is the ease in multiplexing the sensors. Figure 11.32 shows schemes for multiplexing sensors. Sensors of different resonant wavelengths are

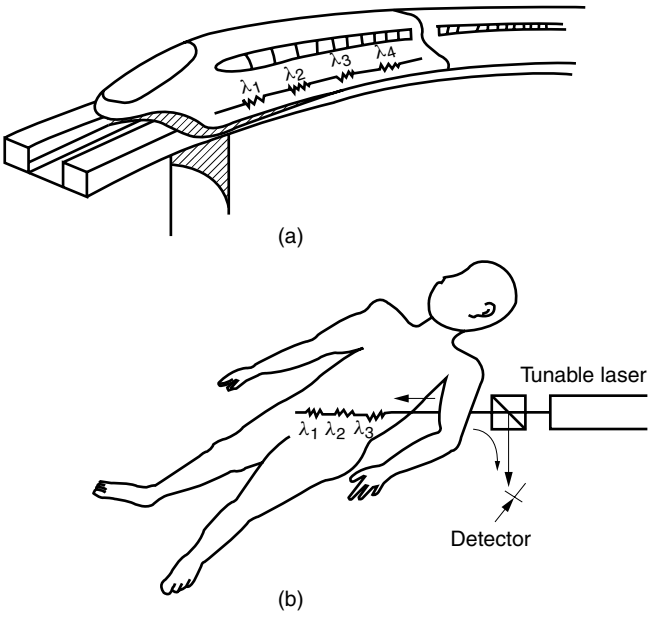


Figure 11.32 Multiplexed fiber Bragg grating sensors for (a) strain and (b) temperature.

imprinted along an optical fiber. A change in a particular resonant wavelength indentifies which sensor is being affected [34]. The number of multiplexed sensors can be large because the transmission loss through a sensor at the off-resonance wavelengths is of the order of 0.01 dB. Strain sensors are multiplexed in Fig. 11.32a. Figure 11.32b shows a similar arrangement for temperature sensors. The latter is well suited to measuring temperature profiles, and it is the profile of the temperature rather than an isolated point temperature that is essential to tumor or cancer diagnosis and treatment [35].

Another application as a device in fiber-optic communication is the chirped aperiodic Bragg grating used to compensate for the dispersion of the optical fiber transmission line. This, however, will be explained in the next section, together with other types of dispersion compensators.

11.6 DEFINITIONS ASSOCIATED WITH DISPERSION

Before going into the design of the dispersion-shifted fiber, a few definitions are presented.

11.6.1 Definitions of Group Velocity and Group Delay

The speed of propagation of the envelope of the modulated light is the *group velocity*. The modulated light contains a band of frequency components centered around the carrier frequency. The shape of the envelope of the received light can be calculated by decomposing it into the component sinusoidal waves by the Fourier transform method. The envelope of the received light can be reconstructed by summing up the spectral components, each of which has propagated by its own phase velocity to the receiver. The time for the modulated light to reach the receiver determines the group velocity [36].

A simpler method can provide the same value as obtained by the Fourier transform method when the spread of the frequency spectrum due to the modulation is moderate. As a modulated wave, the simple method uses a beat wave composed of only two frequency components: $\omega + \delta\omega$ and $\omega - \delta\omega$, where ω is the carrier angular frequency. The beat wave is

$$e^{-j(\omega+\delta\omega)t+j(\beta+\delta\beta)z} + e^{-j(\omega-\delta\omega)t+j(\beta-\delta\beta)z} = 2\cos(\delta\omega \cdot t - \delta\beta \cdot z)e^{-j\omega t+j\beta z}$$

The first term is the envelope and the second term, the carrier. The velocity of the envelope can be found in the same way that the phase velocity was found in Section 1.1. The peak of the envelope appears whenever the value of the bracket is zero. At $t = 0$, the peak is at $z = 0$. At $t = \Delta t$, the peak appears at $\Delta z = (\delta\omega/\delta\beta)\Delta t$; hence, the velocity v_g of the envelope is

$$v_g = \frac{d\omega}{d\beta} \tag{11.147}$$

Note that $\delta\beta$ contains not only the change $\delta\omega$ but also the frequency dependence of the refractive index n, for instance, in an unbound medium $\beta = n\omega/c$. Another noteworthy

point is that when the received waveform is too distorted, Eq. (11.147) for the group velocity is no longer valid.

We will now define a quantity related to the group velocity called the group delay. The time it takes for the envelope (or a pulse) to travel a unit distance (for instance, 1 km) is called the *group delay* τ. The group delay is the inverse of the group velocity and

$$\tau = \frac{d\beta}{d\omega} \tag{11.148}$$

Using the relationship $k = \omega/c$, τ can be written in terms of k as

$$\tau = \frac{1}{c}\frac{d\beta}{dk} \tag{11.149}$$

11.6.2 Definition of the Dispersion Parameter

If every frequency spectrum in the pulse has the same group delay, they all arrive with the same delay, and the received envelope is not distorted. On the other hand, if some spectra arrive earlier than others, the received pulse is distorted. In other words, if the derivative of τ with respect to λ or ω is not zero, there is a distortion. The degree of distortion due to the dispersion is expressed by

$$D = \frac{d\tau}{d\lambda} \tag{11.150}$$

and D is called the *dispersion parameter*.

11.7 DISPERSION-SHIFTED FIBER

A rudimentary description of a dispersion-shifted fiber was given in Section 11.1.4. The amount of waveguide dispersion is adjusted such that cancellation with the material dispersion takes place at $\lambda = 1.55$ μm, where the transmission loss of the optical fiber is lowest. Here, we undertake a more quantitative approach to shifting the disperison-free wavelength. The calculation starts with the expression for the propagation constant β from which the group delay $\tau = d\beta/d\omega$ is obtained. In turn, the group delay τ provides the dispersion parameter $D = d\tau/d\lambda$. Finally, the condition that makes $D = 0$ at the desired wavelength $\lambda = 1.55$ μm is found. Since the derivation is rather lengthy, the group delay and the dispersion parameter are covered separately.

11.7.1 Group Delay in an Optical Fiber

Inserting Eq. (11.120) into (11.149), the group delay in an optical fiber is found to be

$$\tau = \frac{1}{c}(N_1 - N_w) \tag{11.151}$$

where

$$N_1 = \frac{d(n_1 k)}{dk}$$

or

$$N_1 = n_1 + k\frac{dn_1}{dk} \tag{11.152}$$

and where

$$N_w = \frac{d}{dk}[n_1 k\Delta(1-b)]$$

or

$$N_w = (n_1 - n_2)\left[1 - \frac{d(kb)}{dk}\right] \tag{11.153}$$

In the above derivation as well as subsequent derivations, assumptions such as

$$\begin{aligned} \frac{d}{dk}(n_1 - n_2) &\doteqdot 0 \\ \frac{d}{dk}\sqrt{n_1 - n_2} &\doteqdot 0 \\ \frac{d}{d\lambda}\sqrt{n_1^2 - n_2^2} &\doteqdot 0 \end{aligned} \tag{11.154}$$

are made. These assumptions mean that the frequency dependence of n_1 can be assumed identical with that of n_2. This is justifiable from the curves in Fig. 11.7.

First, Eq. (11.152) is considered. N_1 is called the *group index*. N_1 is conventionally tabulated in terms of λ, and an effort will be made to convert the variable accordingly.

$$\frac{dn_1}{dk} = \frac{dn_1}{d\lambda}\frac{d\lambda}{dk} \tag{11.155}$$

$$\frac{dk}{d\lambda} = -\frac{2\pi}{\lambda^2} \tag{11.156}$$

Combining Eq. (11.156) with (11.152) gives

$$N_1 = n_1 - \lambda\frac{dn_1}{d\lambda} \tag{11.157}$$

Note that the group index N_1 has nothing to do with the dimension of the fiber but depends on the material of the core and is associated with the curve of material dispersion in Fig. 11.9.

Next, Eq. (11.153) is considered. N_w is called the *waveguide index*. N_w is usually expressed in terms of the normalized radius V given by Eq. (11.77). We have

$$\frac{dV}{dk} = a\sqrt{n_1^2 - n_2^2} \tag{11.158}$$

where Eq. (11.154) was used.

Equation (11.158) is approximated as

$$\frac{dV}{dk} \doteqdot a\sqrt{2n_1(n_1 - n_2)} \tag{11.159}$$

Using Eq. (11.159) and

$$\frac{d}{dk} = \frac{dV}{dk}\frac{d}{dV}$$

the second term in the square brackets of Eq. (11.153) becomes

$$\frac{d(kb)}{dk} = a\sqrt{2n_1(n_1 - n_2)}\,\frac{d(kb)}{dV} \tag{11.160}$$

and with Eq. (11.158)

$$\frac{d(kb)}{dk} = \frac{V}{k}\frac{d(kb)}{dV}$$

Again using Eq. (11.159), we obtain

$$\frac{d(kb)}{dk} = b + V\,\frac{db}{dV}$$

and finally,

$$\frac{d(kb)}{dk} = \frac{d(Vb)}{dV} \tag{11.161}$$

Inserting Eq. (11.161) into (11.153) gives

$$N_w = (n_1 - n_2)\,\left(1 - \frac{d(Vb)}{dV}\right) \tag{11.162}$$

The numerical value of $d(Vb)/dV$ is readily available from Ref. 37, and the curve for the dominant mode HE_{11} is shown in Fig. 11.33.

Finally, the expression for the group delay becomes, from Eqs. (11.151), (11.157), and (11.162),

$$\tau = \frac{1}{c}\left[n_1 - \lambda\frac{dn_1}{d\lambda} - (n_1 - n_2)\,\left(1 - \frac{d(Vb)}{dV}\right)\right] \tag{11.163}$$

Now, all the preparation for calculating the dispersion has been completed.

11.7.2 Dispersion Parameter of an Optical Fiber

The dispersion parameter given by Eq. (11.150) is obtained by taking the derivative of Eq. (11.163) with respect to λ. Even though the first two terms of Eq. (11.163) are differentiated by λ, the last term is normally differentiated by V. Noting that

$$V = \frac{2\pi}{\lambda}a\sqrt{(n_1^2 - n_2^2)}$$

the derivative of V with respect to λ is

$$\frac{dV}{d\lambda} = -\frac{V}{\lambda} \tag{11.164}$$

where the assumption of Eq. (11.154) was used.

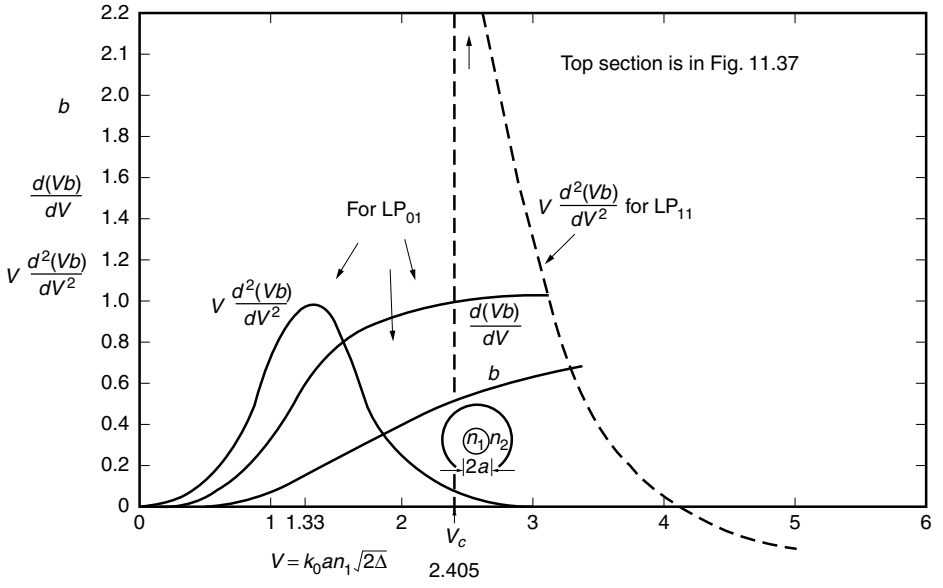


Figure 11.33 Values of b, $d(Vb)/dV$ and $Vd^2(Vb)/dV^2$. These values are used in the design of dispersion-shifted fibers and dispersion compensators. The graph of $Vd^2(Vb)/dV^2$ is shown for both the LP_{01} and LP_{11} modes. (After D. Gloge [37].)

Thus,

$$\frac{d}{d\lambda} = \frac{dV}{d\lambda}\frac{d}{dV} = -\frac{V}{\lambda}\frac{d}{dV} \tag{11.165}$$

Finally, with Eqs. (11.150), (11.154), (11.163), and (11.165), the dispersion parameter is

$$D = -\frac{\lambda}{c}\frac{d^2 n_1}{d\lambda^2} - \frac{n_1\Delta}{c\lambda}V\frac{d^2(Vb)}{dV^2} \tag{11.166}$$

The first term of Eq. (11.166) has nothing to do with the structure of the fiber: it is associated with the core material and is called the *material dispersion*. The second term depends on the physical structure of the fiber and is called the *waveguide dispersion*.

The values of the first and second terms in Eq. (11.166) are of opposite signs in the region $\lambda > 1.27$ μm. If the magnitude of the waveguide dispersion is increased, the cancellation wavelength shifts to a longer wavelength. The dispersion-shifted fiber is designed such that the cancellation takes place at $\lambda = 1.55$ μm.

The waveguide dispersion D_w is given by the second term of Eq. (11.166). The general approaches to increasing D_w in an ordinary step-index configuration are the following:

1. Choose a value of V near $V = 1.33$, where Vd^2Vb/dV^2 is a maximum, as shown in Fig. 11.33.
2. Increase the difference $n_1\Delta = n_1 - n_2$ in the indices of refraction between the core and cladding glasses. Excessive doping of the core in order to

increase the index of refraction is undesirable because of the increased Rayleigh scattering loss. Figure 11.34 shows the refractive index distribution of a low-loss dispersion-shifted fiber [38]. A combination of raising the core refractive index by doping with GeO_2 and lowering the cladding refractive index by doping with fluorine creates a low-loss dispersion-shifted fiber of 0.2 dB/km.

By alternating higher and lower indices of refraction radially in the cladding glass as shown in Fig. 11.35, it is also possible to fabricate an optical fiber whose dispersion is small for the entire wavelength region from $\lambda = 1.3$ to 1.55 μm. A disadvantage of the dispersion-shifted fiber is the larger bending loss because of the smaller value of $V(V = 1.33)$ mostly, $2.0 < V < 2.4$. Bending loss increases with decreasing V because γ is decreased and the confinement of the core light becomes poorer. The principle for

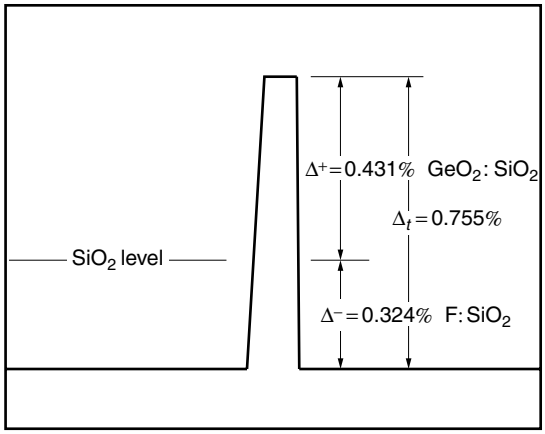


Figure 11.34 Profile of 1.55 μm zero-dispersion fiber. (After M. Ogai et al. [38].)

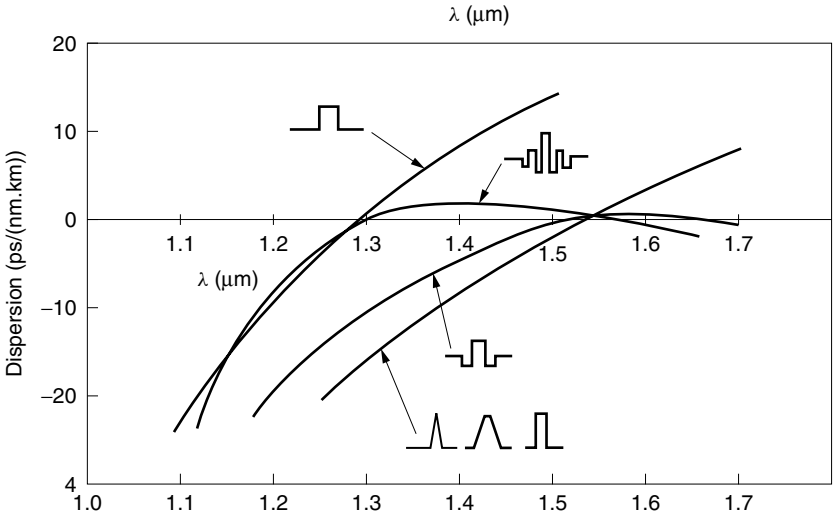


Figure 11.35 SM fiber profile and dispersion. (After H. Murata [39].)

designing a dispersion-shifted fiber will be explained using the example profile shown in Fig. 11.34.

Example 11.4 Design the radius of a dispersion-shifted fiber and its cladding refractive index n_2 using the diagrams in Figs. 11.9 and 11.33. The refractive index of the core is assumed to be $n_1 = 1.458$.

Solution First, we will find D_w. Then, we will select $V = 1.33$, where $Vd^2(Vb)/dV^2$ is a maximum, and determine Δ and a with the constraint of $V = 1.33$. From the second term of Eq. (11.166), the waveguide dispersion parameter D_w is

$$D_w = -\frac{n_1 \Delta}{c\lambda} V \frac{d^2(Vb)}{dV^2} \tag{11.167}$$

From Fig. 11.9, the value of waveguide dispersion that is needed to cancel the material dispersion at $\lambda = 1.55$ is

$$D_w = -28\text{ps/(km} \cdot \text{nm)}$$

From the first approach mentioned earlier for increasing D_w, the point of operation in Fig. 11.33 is selected at

$$\begin{aligned} V &= 1.33 \\ V\frac{d^2(Vb)}{dV^2} &= 1 \end{aligned} \tag{11.168}$$

One has to be careful about the units because D_w is defined in terms of ps, km, and nm.

$$c = \frac{3 \times 10^8 \times 10^{-3}}{10^{12}} = 3 \times 10^{-7}\text{ km/ps}$$

$$\lambda = 1550\text{ nm}$$

Even though both km and nm are distance dimensions, km belongs to only the distance of propagation, whereas nm is used for the wavelength of the light and the radius of the core. Thus, Eq. (11.167) becomes

$$28 = (1)\frac{n_1 \Delta}{(3 \times 10^{-7})(1550)}$$

$$n_1 \Delta = 0.0130$$

$$\Delta = 0.89\%$$

Hence,

$$n_2 = 1.445$$

The value of a is found from

$$V = 1.33$$

and

$$1.33 = \frac{2\pi}{1.55}(1.458)\sqrt{2 \times 0.0089}a$$

$$a = 1.69 \ \mu\text{m}$$

The profile therefore becomes tall and slender, similar to the profile shown in Fig. 11.34. □

In the above example, the value of $Vd^2(bV)/dV^2 = 1$ at $V_{\max} = 1.33$ was used to design the physical parameters of the dispersion-shifted fiber, but as the bit rate of the pulse modulation of the light becomes increasingly high, say, 100 Gb/s ($\Delta\lambda \doteqdot 1$ nm or $\Delta V = 10^{-3}$), then $Vd^2(bV)/dV^2$ can no longer be considered constant. A correction factor associated with

$$\frac{d}{dV}\left(V\frac{d^2(Vb)}{dV^2}\right)$$

has to be taken into consideration. This is called the third-order disperison effect [40]. This effect cannot be tolerated in long-haul transmission with high bit rates. As a countermeasure, compensators that create a dispersion equal and opposite to the third-order dispersion effect are installed after a specified distance of transmission in the dispersion-shifted fiber. These compensators cancel the third-order dispersion effect. The next section is devoted to the explanation of the dispersion compensator in general, not just a compensator for the third-order dispersion.

11.8 DISPERSION COMPENSATOR

Even though the dispersion-shifted fiber is an attractive solution to the dispersion problems of an optical fiber, replacing all the existing fiber transmission lines with dispersion-shifted fiber would be a very expensive undertaking. Enough single-mode step-index fiber has already been installed to wrap around the globe more than several times. If we try to make use of already installed single-mode fiber, serious limitations are encountered at high bit rates. For example, if 10 Gb/s of RZ coded pulses were to be transmitted, the maximum span distance of an uncompensated single-mode fiber is less than 100 km.

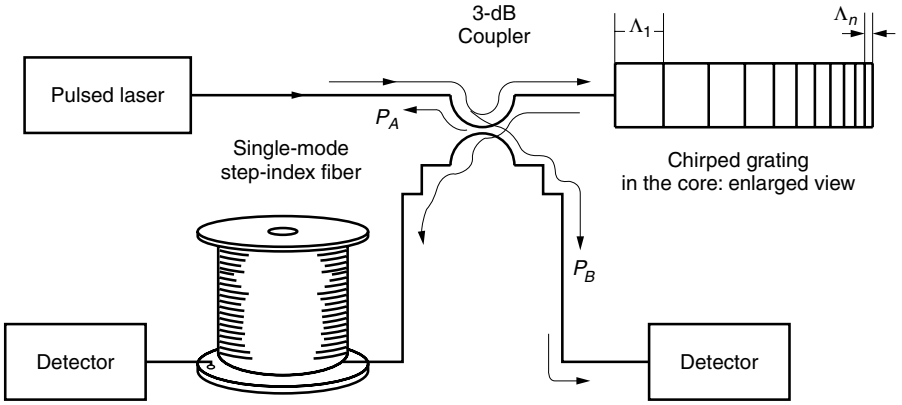
Dispersion compensators are used to break the bit rate barrier in already installed optical fibers [41]. Three different kinds of dispersion compensators are described next.

11.8.1 Phase Conjugation Method

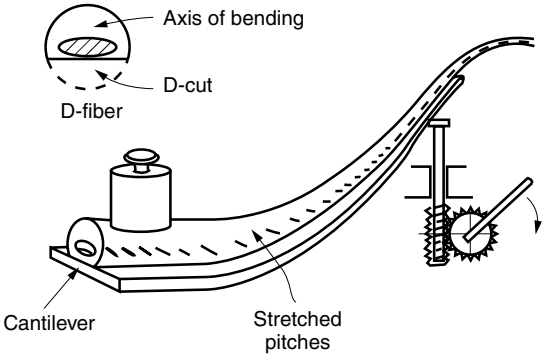
As mentioned in Section 8.12, if a phase conjugate mirror is installed midway through the transmission fiber, a close approximation of the original pulse shape is recovered at the receiver.

11.8.2 Bragg Grating Method

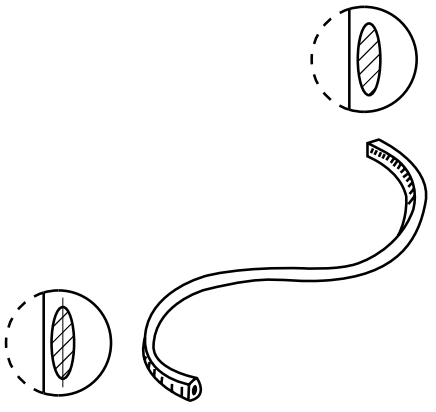
The Bragg grating method uses a chirped Bragg grating reflector photoimprinted in the core of the fiber [42]. As shown in Fig. 11.36, the pitch of the Bragg grating is



(a)



(b)



(c)

Figure 11.36 Dispersion compensation using a chirped Bragg reflector photoimprinted in the fiber. (a) Experimental setup. (b) Chirping achieved by single bending. (c) Chirping obtained by double bending.

decreased (chirped) with distance. When this grating is used as a reflector, the longer wavelength components penetrate shallower into the grating, and the turnaround time is shorter. Thus, the group delay τ decreases with wavelength and the waveguide dispersion parameter D_w becomes a negative quantity; the dispersion parameter of the step-index fiber, which is a positive quantity, can be canceled out.

The total length of the chirped Bragg reflector is calculated for the example of 10 Gb/s of RZ coded pulses in a single-mode fiber. The spread of the arrival time is $\Delta t = 35$ ps. The distance that the light pulse travels in a $n = 1.55$ fiber core in 35 ps is $(c/n)\Delta t = 6.8$ mm and the required length of the Bragg reflector is 3.4 mm.

With the arrangement shown in Fig. 11.36a, the pulse goes around the chirped reflector first before entering the single-mode step-index fiber; but the order can be reversed. Because of the 3-dB fiber coupler in this configuration, half the power is lost when entering the chirped reflector, and another half of the power is lost getting out of the chirped reflector. This power loss associated with the coupler can be avoided by modifying the Mach–Zehnder interferometer shown in Fig. 11.31. The modification consists of replacing the two ordinary gratings with two chirped gratings.

Besides using a chirped photomask, a chirped fiber grating can be fabricated by simply bending a flexible cantilever, onto which a D-fiber written by a regular pitch grating is bonded. The axis of the bending of the D-fiber is laterally displaced from the center of the circular cross section because of the D-cut. If the fiber is bent with the D-cut facing away from the center of the bending, the pitch of the grating is expanded. If, however, the fiber is bent in the opposite way, the pitch will be shortened. In Fig. 11.36b [43], the fiber is bent in only one direction, while in Fig. 11.36c [44], there are two bends in opposite directions.

11.8.3 Dual-Mode Fiber Method

Dispersion compensation using a dual-mode fiber has already been mentioned briefly, but a more detailed explanation will be given here as to how a large negative waveguide dispersion parameter is obtainable from a dual-mode fiber.

According to Eq. (11.167), the waveguide dispersion parameter is influenced by the value of $Vd^2(Vb)/dV^2$. This value for the LP_{11} mode has been calculated by Poole et al. [45] and is shown in Fig. 11.37. The value asymptotically increases as the cutoff is approached and can provide a large negative waveguide dispersion parameter.

It is instructive to give a physical interpretation of why the LP_{11} mode displays such a large negative waveguide dispersion parameter. As indicated by Eq. (11.150), the variation of τ with respect to λ is the dispersion parameter D. Thus, D can be known by finding how τ changes with an increase in λ. From Eq. (11.77) as λ is increased, V decreases. From Fig. 11.17, a decrease in V causes a decrease in γ. The value of γ controls the field in the cladding, which is approximately expressed as $E_0 e^{-\gamma r}$. The decrease in γ means more light energy propagates in the cladding whose index of refraction is smaller than that of the core. Consequently, a decrease in γ increases the group velocity, if the cladding is assumed to be an unbounded medium, and $v_g = c/n$, and hence decreases the group delay τ. In short, an increase in λ decreases τ and the waveguide dispersion parameter is a negative quantity.

Next, the reason for the large magnitude near the cutoff is considered. The shift of the energy into the cladding is controlled by $E_0 e^{-\gamma r}$. For instance, the field at $r = a$ is expressed as $E_0 e^{-\gamma a}$. The rate of change with γ, $d(E_0 e^{-\gamma a})/d\gamma$, is largest when $\gamma = 0$. This is when the operating point is near or at the cutoff.

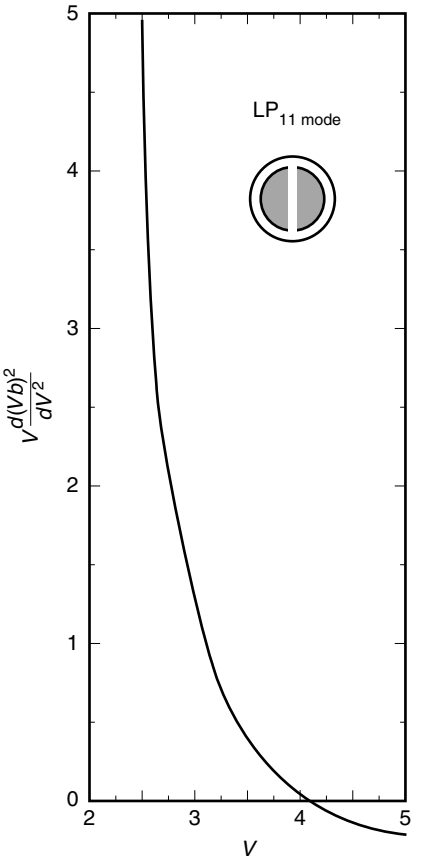


Figure 11.37 $Vd^2(Vb)/dV^2$ of LP_{11} versus V for calculating the dispersion parameter. (After C. D. Poole et al. [45].)

The actual values of D_w for the LP_{11} near the cutoff will be calculated in the next example.

Example 11.5 Determine the value of the waveguide dispersion parameter of a dual-mode fiber with the following physical parameters:

Wavelength of operation, $\lambda = 1.550$ μm
Core radius, $a = 2.1$ μm
Refractive index of the core, $n_1 = 1.460$
Relative core index step, $\Delta = 0.02$

Solution

$$V = kan_1\sqrt{2\Delta}$$

$$= \frac{2\pi}{1.55}(2.1)(1.460)\sqrt{2(0.02)} = 2.49 > 2.40$$

From Eq. (11.167) and Fig. 11.37,

$$D_w = -\frac{n_1\Delta}{c\lambda}V\frac{d^2(Vb)}{dV^2}$$
$$= \frac{-(1.46)(0.02)(5)}{(3\times 10^5\times 10^{-12})(1550)} = -314 \text{ ps/(km}\cdot\text{nm)}$$

The material dispersion D_m is 28 from Fig. 11.9. The dispersion parameter D is

$$D = D_m + D_w = 28 - 314 = -286 \text{ ps/(km}\cdot\text{nm)} \qquad \square$$

Example 11.6 After 100 km of transmission through a step-index single-mode fiber, the dual-mode fiber specified in Example 11.5 is used as a dispersion compensator. What is the required length of the dual-mode fiber?

Solution The length of the dual-mode fiber required to cancel the total dispersion of the single-mode fiber is found. From Fig. 11.9 the dispersion constant D for a single mode step index fiber at $\lambda = 1.55$ μm is $D = 17$ ps/(km · nm). We have

$$17\times 100 = 286x$$
$$x = 5.9 \text{ km} \qquad \square$$

11.9 RAY THEORY FOR GRADED-INDEX FIBERS

The refractive index of the core of a graded-index fiber varies with radius. This space dependence of n makes it very difficult to solve the wave equations (11.14) and (11.15). Ray theory [46,47], which is sometimes called geometrical optics, is often used instead of wave theory. Ray theory is approximate and assumes that the lightwave in a fiber is a plane wave that is changing its course according to the spatial variation of the refractive index. In ray theory, the procedure for solving the wave equation starts with reducing the wave equation to the eikonal equation. The eikonal equation is an approximate but simpler differential equation to solve for the spatial distribution of the phase. After finding the phase distribution, the path of light or wavenormal is found from its gradient. This approach assumes that any physical dimensions such as the fiber core and the fiber length are much larger than the wavelength of the light. In other words, ka is much larger than unity. The accuracy of the ray theory deteriorates as the dimensions approach the order of a wavelength.

Knowledge of the light path, however, leads to such important quantities as the propagation constant, modes, group delay, and dispersion of the multimode graded-index fiber.

11.9.1 Eikonal Equation

When the index of refraction n becomes *space dependent*, the exact solution of the wave equation

$$\nabla^2 u + [kn(r,\phi,z)]^2 u = 0 \qquad (11.169)$$

is very difficult to obtain. As a matter of fact, solutions can be found in closed form for only a limited number of functions $n(r, \phi, z)$.

The procedure begins by assuming a solution of the wave equation of the form

$$u = e^{jkL(r,\phi,z)} \tag{11.170}$$

We will seek $L(r, \phi, z)$ that satisfies the wave equation. Before putting Eq. (11.170) into (11.169), $\nabla^2 u$ will be calculated. A useful identity for $\nabla^2 u$ is

$$\nabla^2 u = \boldsymbol{\nabla} \cdot \boldsymbol{\nabla} u \tag{11.171}$$

where $\boldsymbol{\nabla} u$ in cylindrical coordinates is

$$\boldsymbol{\nabla} u = \hat{\mathbf{r}} \frac{\partial u}{\partial r} + \hat{\boldsymbol{\phi}} \frac{1}{r} \frac{\partial u}{\partial \phi} + \hat{\mathbf{k}} \frac{\partial u}{\partial z}$$

From Eq. (11.170), ∇u is expressed as

$$\boldsymbol{\nabla} u = \underbrace{jke^{jkL}}_{f} \underbrace{\boldsymbol{\nabla} \mathbf{L}}_{A} \tag{11.172}$$

Equation (11.172) is the product of the scalar function

$$f = jke^{jkL} \tag{11.173}$$

and the vector function

$$\mathbf{A} = \boldsymbol{\nabla} L \tag{11.174}$$

The second operation $\boldsymbol{\nabla}\cdot$ in Eq. (11.171) is performed using the identity

$$\boldsymbol{\nabla} \cdot (f\mathbf{A}) = f \boldsymbol{\nabla} \cdot \mathbf{A} + (\boldsymbol{\nabla} f) \cdot \mathbf{A} \tag{11.175}$$

With Eqs. (11.170) to (11.175), the wave equation of Eq. (11.169) is written as

$$\nabla^2 u + (kn)^2 u = \{jk\nabla^2 L - k^2[(\boldsymbol{\nabla} L)\cdot(\boldsymbol{\nabla} L) - n^2]\} e^{jkL} = 0 \tag{11.176}$$

If k is much larger than unity, the k term is much smaller than the k^2 term and the first term in the bracket can be ignored. Thus, Eq. (11.176) becomes

$$|\boldsymbol{\nabla} L|^2 = n^2 \tag{11.177}$$

Equation (11.177) is called the eikonal equation (meaning "image" in Greek). $|\boldsymbol{\nabla} L|$ is the magnitude of the vector $\boldsymbol{\nabla} L$. L is called the eikonal and gives the specific phase distribution of the light.

Next, the eikonal of the light inside a graded-index fiber will be found. The eikonal equation, Eq. (11.177), in cylindrical coordinates is

$$\left(\frac{\partial L}{\partial r}\right)^2 + \left(\frac{1}{r}\frac{\partial L}{\partial \phi}\right)^2 + \left(\frac{\partial L}{\partial z}\right)^2 = n^2(r) \tag{11.178}$$

where the distribution of the index of refraction $n(r)$ is assumed cylindrically symmetric. The differential equation, Eq. (11.178), will be solved by separation of variables. Let the solution be of the form

$$kL = R(r) + \Phi(\phi) + Z(z) \tag{11.179}$$

Both sides of Eq. (11.178) are first multiplied by k^2 and Eq. (11.179) is inserted to obtain

$$\begin{array}{lll} \underbrace{(R')^2 - n^2(r)k^2 + \dfrac{1}{r^2}(\Phi')^2}_{-\beta^2} & + & \underbrace{(Z')^2}_{\beta^2} = 0 \\ & + & = 0 \end{array} \tag{11.180}$$

Since the last term is the only variable in z, the condition that the equation be satisfied for any value of z in space is that the last term has to be constant, as designated by β^2; so

$$Z' = \pm\beta \tag{11.181}$$

and

$$Z = \pm\beta z \tag{11.182}$$

With Eq. (11.181), Eq. (11.180) becomes

$$\begin{array}{lll} \underbrace{r^2[R'^2 - (nk)^2 + \beta^2]}_{-\nu^2} & + & \underbrace{\Phi'^2}_{\nu^2} = 0 \\ & + & = 0 \end{array} \tag{11.183}$$

For the same reason as before, Φ'^2 has to be a constant ν^2 and

$$\Phi = \pm\nu\phi \tag{11.184}$$

From Eq. (11.183)

$$R' = \pm\sqrt{(nk)^2 - (\nu/r)^2 - \beta^2} \tag{11.185}$$

Only the plus signs will be taken up for now, and with the plus sign, Eq. (11.179) becomes

$$kL = \int \sqrt{(kn)^2 - (\nu/r)^2 - \beta^2}\,dr + \nu\phi + \beta z \tag{11.186}$$

Finally, the expression for the lightwave in the fiber is

$$u = \exp\left[j\int \sqrt{(nk)^2 - (\nu/r)^2 - \beta^2}\,dr + j\nu\phi + j\beta z\right] \tag{11.187}$$

Now that the phase distribution has been found, the direction of the wave normal or the light path can be calculated.

11.9.2 Path of Light in a Graded-Index Fiber

The light path can be found from the gradient of the phase contour; namely, $\nabla(kL)$ indicates the direction of the wavenormal. In order to make the wavenormal direction a unit vector $\hat{\mathbf{s}}$, the quantity $\nabla(kL)$ has to be divided by its magnitude $|\nabla(kL)|$:

$$\hat{\mathbf{s}} = \frac{\nabla(kL)}{|\nabla(kL)|} \tag{11.188}$$

From Eq. (11.177), the unit vector $\hat{\mathbf{s}}$ is written as

$$\hat{\mathbf{s}} = \frac{\nabla L}{n} \tag{11.189}$$

One more expression is needed to find the path of light; that is, the expression for the light path using the position vector. As shown in Fig. 11.38, the position vector $\mathbf{R}$ is generated by connecting a point on the light path with the origin of the coordinate system. The cord $\Delta\mathbf{R}$ in Fig. 11.38 is an approximation to the path of light. The shorter $\Delta\mathbf{R}$ is, the better $\Delta\mathbf{R}$ matches the path of light. As soon as $\Delta\mathbf{R}$ exactly matches, the magnitude becomes zero and there is no vector. This situation can be remedied by expressing the path by

$$\hat{\mathbf{s}} = \lim_{s \to 0} \frac{\Delta\mathbf{R}}{\Delta s} \tag{11.190}$$

where Δs is the length measured along the actual light path. The limit operation leads to

$$\hat{\mathbf{s}} = \frac{d\mathbf{R}}{ds} \tag{11.191}$$

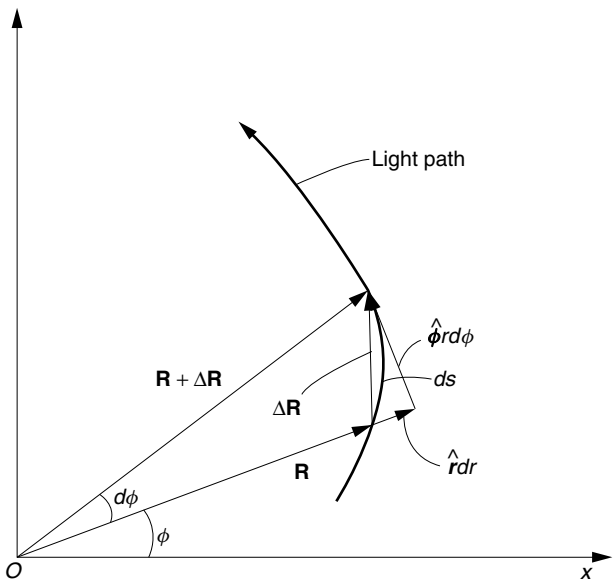


Figure 11.38 Expression for the light path in terms of the position vector.

where, from Fig. 11.38, $d\mathbf{R}$ is given by

$$d\mathbf{R} = \hat{\mathbf{r}}dr + \hat{\boldsymbol{\phi}}rd\phi + \hat{\mathbf{k}}dz \tag{11.192}$$

The combination of Eqs. (11.189) with (11.191) gives the expression for the light path

$$n\frac{d\mathbf{R}}{ds} = \nabla L \tag{11.193}$$

Insertion of Eqs. (11.186) and (11.192) into Eq. (11.193) gives

$$nk\frac{dr}{ds} = \sqrt{(nk)^2 - (\nu/r)^2 - \beta^2} \tag{11.194}$$

$$nkr\frac{d\phi}{ds} = \frac{1}{r}\nu \tag{11.195}$$

$$nk\frac{dz}{ds} = \beta \tag{11.196}$$

Division of Eq. (11.196) by (11.194) and division of Eq. (11.195) by (11.194) finally gives

$$z = \int \frac{\beta}{\sqrt{[n(r)k]^2 - (\nu/r)^2 - \beta^2}}dr \tag{11.197}$$

$$\phi = \int \frac{\nu}{r^2\sqrt{[n(r)k]^2 - (\nu/r)^2 - \beta^2}}d\phi \tag{11.198}$$

Calculation of Eqs. (11.197) and (11.198) for a given distribution of $n(r)$ gives the expression for the light ray path.

Here we will consider the special case of a fiber with a quadratic distribution of $n^2(r)$ as

$$n^2(r) = n_c^2(1 - \alpha^2 r^2) \tag{11.199}$$

where α^2 is the refractive index gradient constant squared. A graded-index fiber with such a refractive index distribution is known as a Selfoc fiber.

After some manipulation, inserting Eq. (11.199) into (11.197) leads to

$$z = \frac{\beta}{2kn_c\alpha}\int \frac{dt}{\sqrt{q^2 - (t-p)^2}} \tag{11.200}$$

where

$$t = r^2 \tag{11.201}$$

$$p = \frac{(n_c k)^2 - \beta^2}{2(n_c k\alpha)^2} \tag{11.202}$$

$$q^2 = p^2 - \left(\frac{\nu}{n_c k\alpha}\right)^2 \tag{11.203}$$

The integral of Eq. (11.200) gives

$$t = p + q \sin \frac{2n_c k\alpha}{\beta}(z - z_0) \tag{11.204}$$

From Eqs. (11.201) and (11.204), r^2 wobbles with z. This suggests that the path is a helix with an elliptic cross section, which will be verified later. The maximum radius is $r = \sqrt{p+q}$ and the minimum is $r = \sqrt{p-q}$.

Next, the integral of Eq. (11.198) will be calculated. It is quite similar to Eq. (11.197) and it becomes

$$\phi = \frac{v}{2n_c k\alpha} \int \frac{dt}{t\sqrt{q^2 - (t-p)^2}} \tag{11.205}$$

After the change of variable

$$x = t - p \tag{11.206}$$

and with the help of the following relationship from a table of integrals,

$$\int \frac{dx}{(c^2x + ab)\sqrt{b^2 - c^2x^2}} = \frac{1}{bc\sqrt{a^2 - c^2}} \sin^{-1} \frac{acx + bc}{c^2x + ab} \tag{11.207}$$

the integral of Eq. (11.205) is

$$\phi - \phi_0 = \frac{1}{2} \sin^{-1} \left(\frac{p(t-p) + q^2}{qt} \right) \tag{11.208}$$

where Eq. (11.203) was also used.

With Eq. (11.201), Eq. (11.208) becomes

$$\frac{1}{r^2}(p^2 - q^2) = p - q \sin 2(\phi - \phi_0) \tag{11.209}$$

Equation (11.209) will be rewritten in a more familiar form. Noting that

$$\sin 2(\phi - \phi_0) = \cos 2(\phi - \phi_0 - \pi/4)$$

we have

$$q \sin 2(\phi - \phi_o) = q \cos^2(\phi - \phi_o - \pi/4) - q \sin^2(\phi - \phi_o - \pi/4) \tag{11.210}$$

$$p = p \left[\cos^2(\phi - \phi_0 - \pi/4) + \sin^2(\phi - \phi_0 - \pi/4) \right] \tag{11.211}$$

Eq. (11.209) is rewritten as

$$1 = \frac{r^2 \cos^2(\phi - \phi_0 - \pi/4)}{p+q} + \frac{r^2 \sin^2(\phi - \phi_0 - \pi/4)}{p-q} \tag{11.212}$$

Changing into rectangular coordinates by noting

$$\begin{aligned} X &= r\cos(\phi - \phi_0 - \pi/4) \\ Y &= r\sin(\phi - \phi_0 - \pi/4) \end{aligned} \tag{11.213}$$

Eq. (11.212) becomes

$$\frac{X^2}{A^2} + \frac{Y^2}{B^2} = 1$$

with

$$\begin{aligned} A &= \sqrt{p+q} \\ B &= \sqrt{p-q} \end{aligned} \tag{11.214}$$

Now it has been verified that the cross section of the skew ray is an ellipse such as shown in Fig. 11.39.

Only the solutions with the positive sign in Eqs. (11.181), (11.184), and (11.185) have been considered, but the solutions with negative signs have to be added. The solutions with negative signs are again skewed but wound in the opposite sense. These two skew rays spiraling in opposite directions form a standing wave, which is precisely what the mode patterns are as will be mentioned in Section 11.9.5.

As ν in Eq. (11.203) approaches zero, the ellipse shrinks into a line and the solution becomes that of a meridional ray.

As the skew ray encircles the fiber axis once, r^2 goes through a maximum and minimum twice. The advance in the z direction after one complete turn is the pitch of

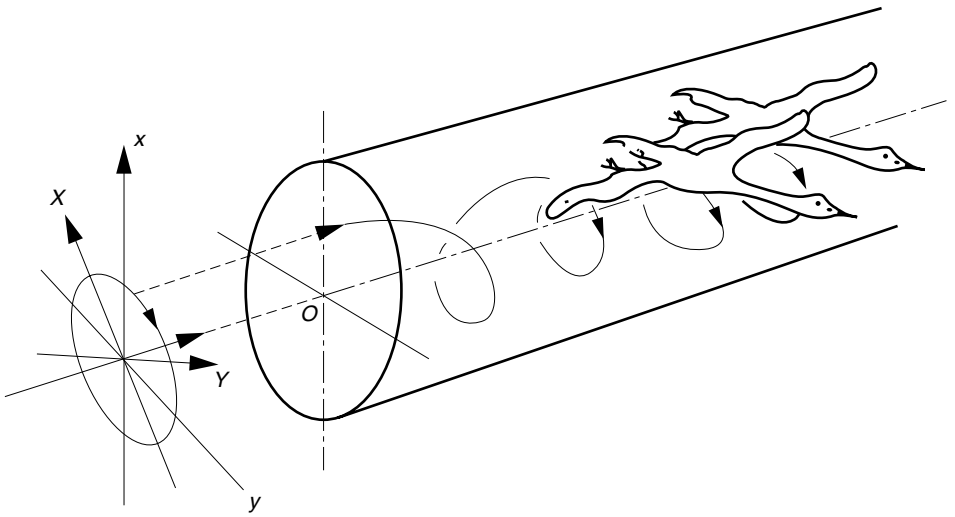


Figure 11.39 Path of a skew ray in a graded-index fiber.

the helix. From Eq. (11.204), the pitch ρ is found:

$$\frac{2n_c k\alpha}{\beta}\rho = 4\pi \tag{11.215}$$

$$\rho = 2\pi\frac{\beta}{n_c k\alpha} \tag{11.216}$$

For lower order modes, $\beta \doteqdot n_c k$ and

$$\rho \doteqdot \frac{2\pi}{\alpha} \tag{11.217}$$

Inside a Selfoc fiber with $\alpha = 0.01\ \mu\text{m}^{-1}$, the pitch is about half a millimeter.

11.9.3 Quantization of the Propagation Constant in a Graded-Index Fiber

As the dimensions of the fiber are reduced, problems of quantization become important. Suppose that a light ray has a finite width with a taper in the direction transverse to the direction of propagation, like the wings of a bird in Fig. 11.39. A portion of the width (the wing of one bird) of the light interferes with the width (the wing of another bird) of the light of the previous turn. If the phases of the two waves are different, they destructively interfere and the amplitude decreases each time the ray encircles the fiber axis. Since the skew ray goes around the fiber axis more than 10^3 times in 1 meter, even a slight decrease in amplitude accumulates, and no propagation takes place. It is therefore necessary that the phase match completely each time ϕ advances by 2π, and hence ν in Eq. (11.184) has to be an integer:

$$\nu = 0, 1, 2, 3, \ldots \tag{11.218}$$

There is another quantization. That is in the r direction. The skew ray waves toward and away from the axis between $r_{\min} = p - q$ and $r_{\max} = p + q$ according to Eq. (11.204). This wavy path can be pictured as a zigzag path in a curved slab guide wrapped around the fiber axis, where the width of this slab is $r_{\max} - r_{\min}$. If we interpret the wavy ray path as a zigzag path in a curved slab guide, then the optical width of such a slab guide has to be an integral multiple of π radians, as discussed in Section 9.7.4. Thus,

$$\mu\pi = \int_{r_{\min}}^{r_{\max}} \sqrt{(nk)^2 - (\nu/r)^2 - \beta^2}\,dr \tag{11.219}$$

The integrand of Eq. (11.219) has a square root, but the square root can be removed by converting the variable from r to z using Eqs. (11.194), (11.196), and (11.216):

$$\mu\pi = \frac{1}{\beta}\int_0^{\rho/4} \left[(nk)^2 - (\nu/r)^2 - \beta^2\right] dz \tag{11.220}$$

For convenience, the integral I of Eq. (11.220) is separated into two parts:

$$I_1 = \frac{1}{\beta}\int_0^{\rho/4} [(nk)^2 - \beta^2]\, dz \tag{11.221}$$

$$I_2 = \frac{1}{\beta}\int_0^{\rho/4} \left(\frac{\nu}{r}\right)^2 dz \tag{11.222}$$

where

$$I = I_1 - I_2 \tag{11.223}$$

Equation (11.204) is rewritten in terms of Eq. (11.216) as

$$r^2 = p + q\sin\frac{4\pi}{\rho}(z - z_0)$$

Setting $z_0 = \rho/8$ the above simplifies as

$$r^2 = p + q\cos\left(\frac{4\pi}{\rho}z\right) \tag{11.224}$$

With Eqs. (11.199), (11.202), and (11.224), after some manipulation the calculated integral I_1 becomes

$$I_1 = \frac{n_c k\alpha\pi}{2}p \tag{11.225}$$

The second integral I_2 can be calculated using the following formula from an integral table:

$$\int \frac{dx}{p + q\cos bx} = \frac{1}{b\sqrt{p^2 - q^2}}\cos^{-1}\left(\frac{q + p\cos bx}{p + q\cos bx}\right) \tag{11.226}$$

The result of integral I_2 is

$$I_2 = \frac{\pi}{2}\nu \tag{11.227}$$

From the results of I_1 and I_2, Eq. (11.220) becomes

$$\mu\pi = \frac{n_c k\alpha\pi}{2}p - \frac{\pi}{2}\nu$$

Using Eq. (11.202), p is rewritten and the quantization formula becomes

$$2n_c k\alpha(\nu + 2\mu) = n_c^2 k^2 - \beta_{\nu\mu}^2 \tag{11.228}$$

Finally, the propagation constant $\beta_{\mu\nu}$ of a graded-index fiber is

$$\beta_{\nu\mu} = n_c k\sqrt{1 - \frac{2\alpha(\nu + 2\mu)}{n_c k}} \tag{11.229}$$

The mode number

$$N = \nu + 2\mu \tag{11.230}$$

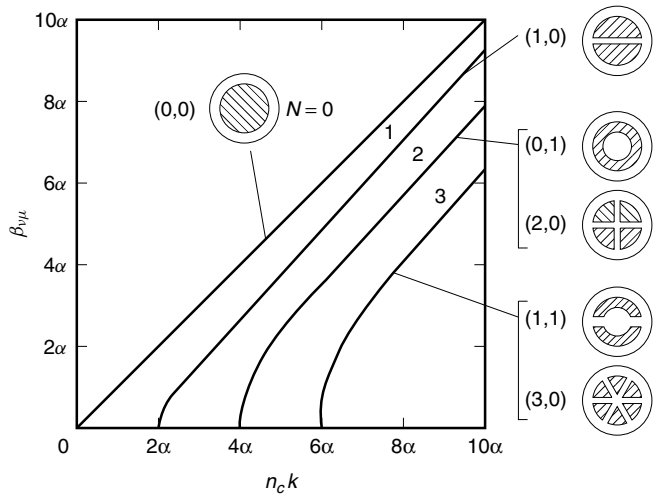


Figure 11.40 $\beta_{\nu\mu}$ versus $n_c k$.

is called the *principal mode number* of a graded-index fiber. The propagation constant $\beta_{\nu\mu}$ can be the same for different combinations of ν and μ. Modes with the same N are degenerate modes. Figure 11.40 shows a graph of $\beta_{\nu\mu}$ versus $n_c k$ with N as a parameter. Both axes are graduated by α. $N = 0$ corresponds to the mode $(\nu, \mu) = (0, 0)$; $N = 1$ to $(1, 0)$; $N = 2$ to the degenerate modes of $(2, 0)$ and $(0, 1)$; and so forth. The condition for a single-mode graded-index fiber is

$$\frac{n_c k}{\alpha} < 2 \tag{11.231}$$

11.9.4 Dispersion of Graded-Index Fibers

Now that the expression for $\beta_{\nu\mu}$ has been obtained, it is possible to calculate the group delay. Taking the derivative of Eq. (11.228) with respect to k, the group delay Eq. (11.149) for the Selfoc fiber is obtained as

$$\tau = \frac{N_c}{c}\frac{n_c k}{\beta_{\nu\mu}}\left(1 - \alpha\frac{(\nu + 2\mu)}{n_c k}\right) \tag{11.232}$$

where

$$N_c = \frac{d(n_c k)}{dk} \tag{11.233}$$

and where N_c is the group index at the fiber axis.

Next, the material and waveguide dispersion of a single-mode graded-index fiber will be calculated. For a single-mode fiber, Eqs. (11.229) and (11.232) with $\mu = \nu = 0$ give

$$\tau = \frac{N_c}{c} \tag{11.234}$$

Analogous to the derivation of Eq. (11.157), N_c is expressed as

$$N_c = n_c - \lambda\frac{dn_c}{d\lambda} \tag{11.235}$$

By taking the derivative of Eq. (11.235), the dispersion parameter of the graded-index fiber therefore becomes

$$D = -\frac{\lambda}{c}\frac{d^2 n_c}{d\lambda^2} \tag{11.236}$$

Thus, unlike a step-index single-mode fiber, a graded-index single-mode fiber has only material dispersion and no waveguide dispersion.

Next, the mode dispersion of a multimode graded-index fiber will be calculated. The factor $(\nu + 2\mu)$ can be eliminated from Eq. (11.232) by expressing it in terms of $\beta_{\nu\mu}$ using Eq. (11.228) as

$$\tau = \frac{N_c}{c}\frac{n_c k}{2\beta_{\nu\mu}}\left[1 + \left(\frac{\beta_{\nu\mu}}{n_c k}\right)^2\right] \tag{11.237}$$

If the ray of the lowest mode propagates straight without any zigzag or encircling of the fiber axis and Eq. (11.234) is satisfied, the group delay τ_c is

$$\tau_c = \frac{N_c}{c} \tag{11.238}$$

From Eqs. (11.237) and (11.238), the mode dispersion, which is the difference $(\tau - \tau_c)$ in the group delays, is

$$\Delta\tau = \frac{N_c}{c}\frac{(1 - \beta/n_c k)^2}{2(\beta/n_c k)} \tag{11.239}$$

where subscripts $\nu\mu$ were suppressed.

Referring to Fig. 11.41, β for the (ν, μ) mode makes an angle δ with the fiber axis, and

$$\cos\delta = \frac{\beta}{n_c k} \tag{11.240}$$

The angle δ can be expressed in terms of the numerical aperture of the fiber using Snell's law at the entrance:

$$\mathrm{NA} = n_c \sin\delta \tag{11.241}$$

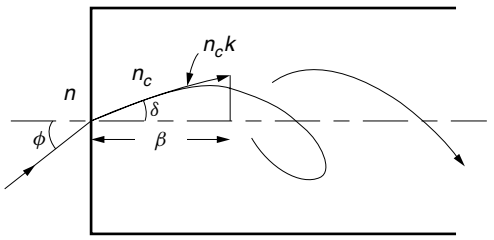


Figure 11.41 Relationship between NA and δ.

Finally, using Eqs. (11.240) and (11.241) for a small value of δ, Eq. (11.239) can be approximated as

$$\Delta\tau = \frac{(\text{NA})^4}{8cn_c^3} \tag{11.242}$$

where $N_c \doteqdot n_c$ was assumed.

Let us now compare the information-carrying capacity of a multimode step-index fiber with that of a multimode graded-index fiber. Comparison between Eqs. (11.7) and (11.242) leads to the conclusion that the graded-index fiber capacity is $4n_c^3/n_1(\text{NA})^2$ times greater, which is normally a few hundred times that of a step-index fiber.

11.9.5 Mode Patterns in a Graded-Index Fiber

A few words about approximate field distributions obtained from ray theory will be mentioned. As the relevant dimensions decrease, the accuracy of ray theory deteriorates. Mode patterns calculated from ray theory cannot be expected to be accurate when the dimensions are of the order of a wavelength. We will calculate the fields using ray theory and compare these with our earlier results. Thus far only the positive signs of Eqs. (11.184) and (11.185) have been retained and the negative signs were discarded. The negative signs generate the skew rays encircling in the opposite direction. Including the terms with the negative sign, two oppositely propagating skew rays form a standing-wave pattern that is the mode pattern of the fiber. Equation (11.187) indicates that such standing waves, $u_0(e^{j\nu\phi} + e^{-j\nu\phi}) = 2u_0 \cos \nu\phi$, have 2ν maxima in every complete turn in circumference.

Next, we will consider the radial direction. From Eq. (11.187)

$$\begin{aligned} u &= u_0 \left[e^{j\int_{r_{\min}}^{r} \sqrt{(nk)^2-(\nu/r)^2-\beta^2}\,dr} + e^{-j\int_{r_{\min}}^{r} \sqrt{(nk)^2-(\nu/r)^2-\beta^2}\,dr} \right] \\ &= 2u_0 \cos \int_{r_{\min}}^{r} \sqrt{(nk)^2 - (\nu/r)^2 - \beta^2}\,dr \end{aligned}$$

The phase advances with r. Eq. (11.219) indicates that by the time r reaches $r_{\max}$ from $r_{\min}$, the phase advances $\mu\pi$ radians. During this phase advance the cosine function goes through μ extrema.

In conclusion, the (ν, μ) mode generates 2ν extrema in the circumference and μ extrema in the radial direction. Examples of the mode patterns are shown in Fig. 11.40.

Thus, the field expressions of the graded-index fiber obtained by ray theory match those of the $\text{LP}_{\nu\mu}$ modes of the step-index fiber, except that the r dependence of the $\text{LP}_{\nu\mu}$ mode is a Bessel function, while the fields obtained by ray theory are sinusoidal functions. The results of the two theories agree reasonably well.

Example 11.7 Answer the following questions for a Selfoc fiber whose physical parameters are as follows:

Refractive index at the center of the fiber, $n_c = 1.54$

Wavelength, $\lambda = 1.3$ μm

Effective radius (i.e., the largest possible major axis of the light path), $r_{\text{eff}} = 10$ μm

Numerical aperture, NA = 0.2

(a) What is the acceptance angle of the fiber?
(b) What is the refractive index gradient constant α?
(c) What is the largest value of the principal mode number N?
(d) Approximately how many modes are excited?
(e) How many of the modes are meridional rays?

Solution

(a) From Eqs. (11.3) and (11.4),

$$\phi_c = 11.5^\circ$$

(b) The major axis A of the ellipse is, from Eq. (11.214),

$$A = \sqrt{p+q}$$

With Eqs. (11.202) and (11.203), A has the largest when $\nu = 0$ and is $\sqrt{2p}$.

$$r_{\text{eff}} = \sqrt{2p} = \frac{\sqrt{(n_c k)^2 - \beta^2}}{n_c k \alpha}$$
$$= \frac{\sqrt{1 - (\beta/n_c k)^2}}{\alpha} \doteq \frac{\delta}{\alpha}$$

where Eq. (11.240) was used. From the above equation and Eq. (11.241), α is

$$\alpha = \frac{\text{NA}}{n_c r_{\text{eff}}} = 0.013\ \mu\text{m}^{-1} \tag{11.243}$$

The result is interpreted as follows. A larger α means a larger focusing power of the incident light to the fiber, and hence a larger NA.

(c) From Eqs. (11.229) and (11.230), $N_{\max}$ is

$$1 = \frac{2\alpha N_{\max}}{n_c k}$$
$$N_{\max} = \frac{n_c k}{2\alpha} = \frac{1.54 \times 2\pi/1.3}{2 \times 0.013} = 286$$

(d) Note that

$$\nu + 2\mu \le 286$$
$$\nu = 286 - 2\mu$$

which is the equation of a line in (ν, μ) coordinates, intercepting the axes at $\nu = 286$ and $\mu = 143$, as shown in Fig. 11.42. Each point inside the triangular area is a mode, so that the total number m of modes is

$$m = \tfrac{1}{2} \times 286 \times 143 = 20{,}449$$

including degenerate modes.

(e) Meridional rays have $\nu = 0$ and there are 143 such modes. □

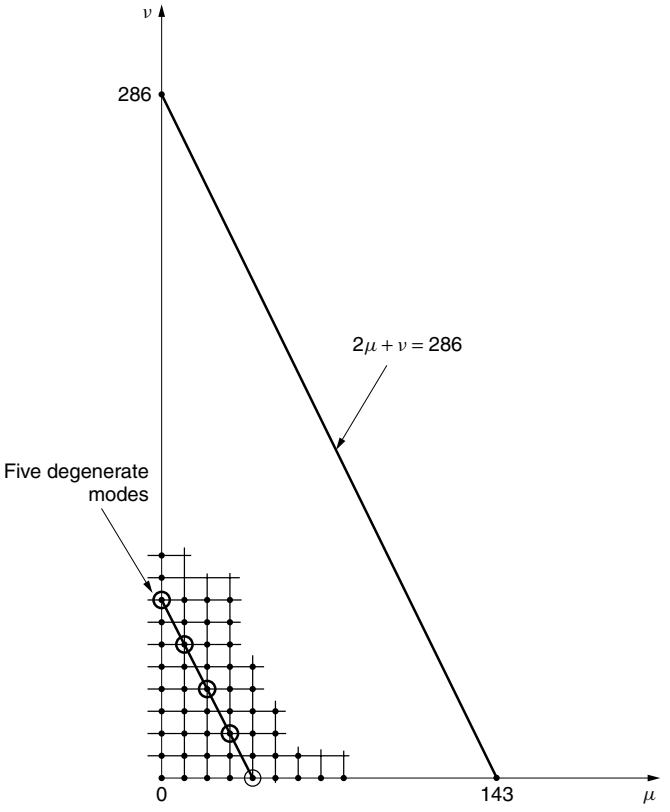


Figure 11.42 Total number of modes in a Selfoc fiber.

Example 11.8 Figure 11.43 shows an excerpt from a Corning Glass Works specification sheet for their 50/125 μm LDF^{TM} CSB4 multimode optical fiber. Verify (a) the numerical aperture and (b) the bandwidth at $\lambda = 1300$ nm listed in the optical specifications of Fig. 11.43 using the parameters found in the characterization data. Assume a quadratic distribution of the index of refraction.

Solution

(a) In order to use Eq. (11.243), the value of α has to be found first. From the given refractive index profile

$$\Delta_{\max} = 0.0125$$

The index of refraction n_c at the axis as specified is

$$n_c = 1.4598$$

Hence, from $\Delta_{\max}$, n_2 at the cladding is

$$n_2 = 1.4416$$

2.0 Optical Specifications

Bandwidth:

Standard Bandwidth Cells:

Bandwidth Cells [MHz-km]	
850 nm	1300 nm
≥400 - ≥1000	≥400 - ≥1500

Special bandwidth cells available upon request.

Core Diameter 50.0 ± 3.0 μm

Numerical Aperture: 0.200 ± 0.015

6.0 Characterization Data

Characterized parameters are typical values.

Core Material Index of Refraction (Peak): 1.4655 at 850 nm
1.4598 at 1300 nm

Refractive Index Profile (typical fiber):

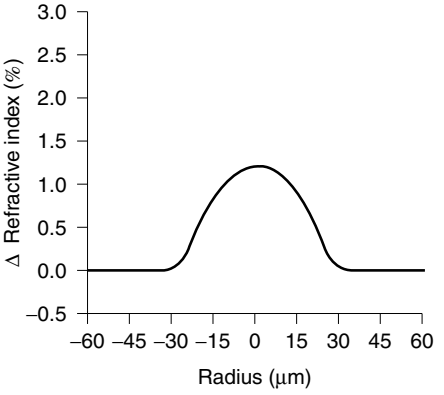


Figure 11.43 Excerpt from Corning's specification sheet.

If a quadratic distribution is assumed, then

$$n^2 = n_c^2(1 - \alpha^2 r^2)$$

$$1 - \left(\frac{n}{n_c}\right)^2 = \alpha^2 r^2$$

The index of refraction in the graph reaches n_2 at $r = 27$ μm.

$$\alpha = \frac{\sqrt{1 - 0.9752}}{27} = 0.00583 \ \mu\text{m}^{-1}$$

From the previous example, Eq. (11.243) gives

$$\text{NA} = n_c r_{\text{eff}} \alpha$$

assuming $r_{\text{eff}} = 25$ μm from the specification for the core diameter, the NA is calculated as

$$\begin{aligned} \text{NA} &= 1.4598 \times 25 \times 0.00583 \\ &= 0.213 \end{aligned}$$

which matches with the value $\text{NA} = 0.200 \pm 0.015$ in the specification sheet.

(b) Next, the maximum bandwith will be found. From Eq. (11.242),

$$\begin{aligned} \Delta\tau = \frac{(\text{NA})^4}{8cn_c^3} &= \frac{(0.21)^4}{(8)(3 \times 10^5)(1.4598)^3} \\ &= 260 \text{ ps/km} \end{aligned}$$

As explained in more detail in Section 16.6.5, in the field of digital communication, the most often used criterion for the required rise time ΔT (time required for the level to rise from 10% to 90% of the steady-state signal level) is that the rise time be kept less than 70% of the bit period when NRZ coding is used and 35% when RZ coding is used, as shown in Fig. 11.44.

Using the 70% criterion, the rise time ΔT should satisfy $\Delta T \leq 0.7T$, where T is the pulsewidth. If the rise time is assumed to be equal to the pulse spread $\Delta\tau$, then the bit rate $1/T$ using the 70% criterion is

$$R = \frac{1}{\Delta\tau} \times 0.7 = 2.695 \times 10^9 \text{ (bits/s)} \cdot \text{km}$$

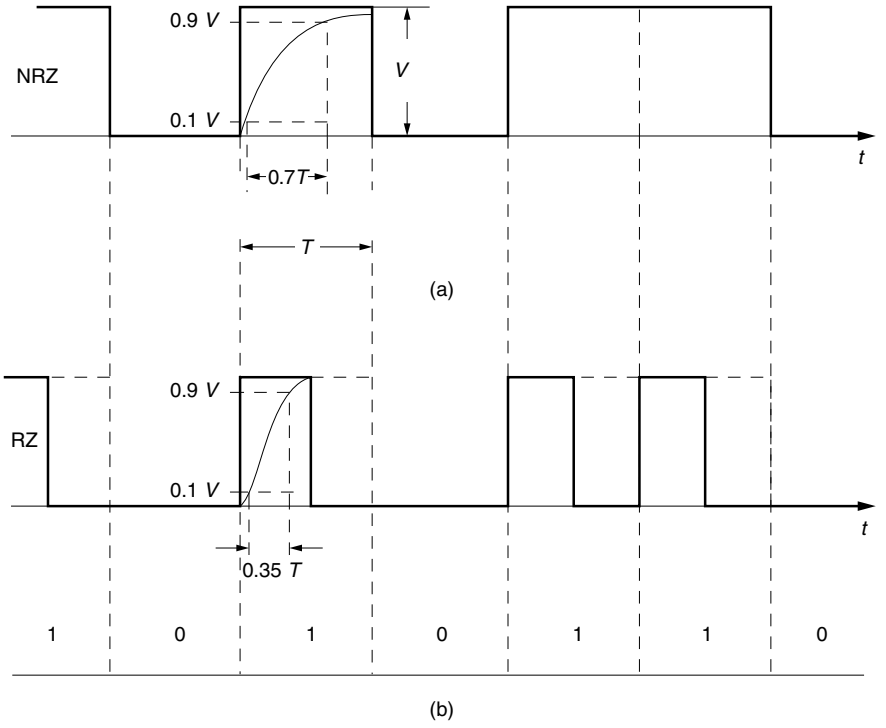


Figure 11.44 Time space T needed for pulses with finite rise time ΔT for (a) NRZ and (b) RZ coding.

Since the specification sheet is given in terms of frequency, a pulsewidth in the time domain has to be converted into the frequency domain. As explained in Section 16.6.5, the pulsetrain at the bit rate R needs a frequency bandwidth B given by

$$B = mR$$

$$m = 0.5 \quad \text{for NRZ code}$$

and

$$B = 1.348 \text{ GHz}$$

This value agrees well with the specification sheet. □

11.10 FABRICATION OF OPTICAL FIBERS

Fabrication of optical fibers may broadly be categorized into one-stage processes and two-stage processes. In the one-stage process, the fiber is drawn directly from the liquefied materials, and in the two-stage process, a highly purified preform rod is fabricated at the first stage, and then the fiber is drawn out of the preform in the second stage. Although the one-stage process has such advantages as simplicity of fabrication, unlimited fiber length, and a faster rate of fabrication, it is the two-stage process that can produce fibers with lower loss and better uniformity.

11.10.1 Fabrication of a Fiber by the One-Stage Process

Figure 11.45 shows a diagram of the double-crucible method, which belongs to the category of one-stage processes. There are inner and outer crucibles with a nozzle at the bottom. The inner crucible is filled with the material of higher index of refraction and the outer with that of the lower index of refraction. The fiber is drawn from the bottom. The nozzle of the inner crucible for the core is much smaller than that of the cladding. The molten core glass is exposed to the molten cladding glass. Diffusion of the molten glasses takes place, and the profile of the graded-index fiber is established. The drawing and the temperature of the heater control the fiber dimension as well as the refractive index profile formed by diffusion.

The diameter of the drawn fiber is monitored, and the monitor output controls the speed of the capstan that pulls the fiber. Immediately after drawing, the fiber is coated with a UV-curable resin, which is then cured inside a confocal reflector UV light cavity, and the coated fiber is reeled onto a bobbin.

An advantage of the double-crucible method is that the fiber can be drawn continuously. A disadvantage is that the double-crucible method can only be used with a lower melting point glass such as multicomponent glass, which has a higher loss than pure silica glass. Another disadvantage associated with the double-crucible method is that product uniformity is more difficult to control.

With a small variation, a plastic-clad silica (PCS) fiber can be fabricated. As the core of the fiber is drawn from the crucible of molten glass, a plastic material that can act as the cladding layer of the fiber is coated onto the glass. Even though fabrication is simpler and costs less, the transmission loss of the resulting fiber is high so that the PCS fiber is only useful for short-distance transmission. Short-distance transmission means a larger dispersion can be tolerated, and the NA values of the PCS fiber are

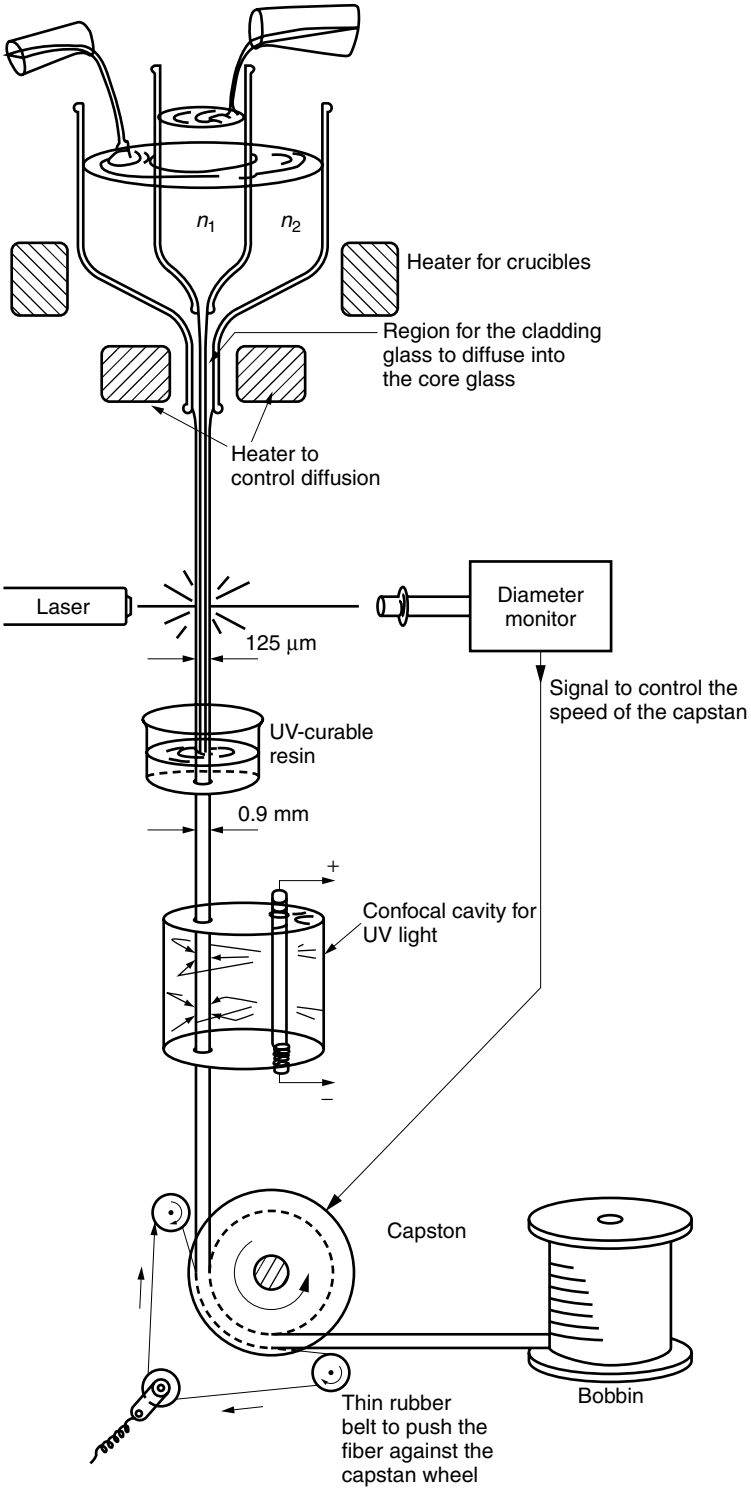


Figure 11.45 Fabrication of optical fiber by the double-crucible method.

usually made larger (NA = 0.35) than that of the glass fiber (NA = 0.2) for easier coupling.

11.10.2 Fabrication of a Fiber by the Two-Stage Process

Fabrication of a fiber by the two-stage process consists of fabricating a preform and then drawing a fiber from the preform. Since the preform can be made out of highly purified silica with a precise refractive index profile, a fiber with low loss and low dispersion can be fabricated. A disadvantage of this fabrication process, however, is the finiteness of the fiber length and the necessity of the additional stage in the process.

11.10.2.1 Fabrication of Preforms

A silica-based core with a purity better than 0.1 ppb (parts per billion 10^{-9}) can be achieved. In order to appreciate this high degree of purity, let us make an analogy with the world's population. In a population of approximately 5 billion people, 0.1 ppb corresponds to $\frac{1}{2}$ a person.

It was the chemical vapor deposition (CVD) method that made it possible to attain such a high degree of purity. The CVD method was originally developed in semiconductor industries. It is an extremely effective method. The method uses the difference in the vapor pressures of the liquid metal halides of the material and those of the impurities. Figure 11.46 shows the vapor pressures of various metal halides as a function of temperature. Metal halides such as BCl_3, $SiCl_4$, and $POCl_3$ are used for the material, and VCl_4 and $FeCl_3$ are the impurities to be removed. Halides whose vapor pressures are above the container pressure will evaporate into the container while those

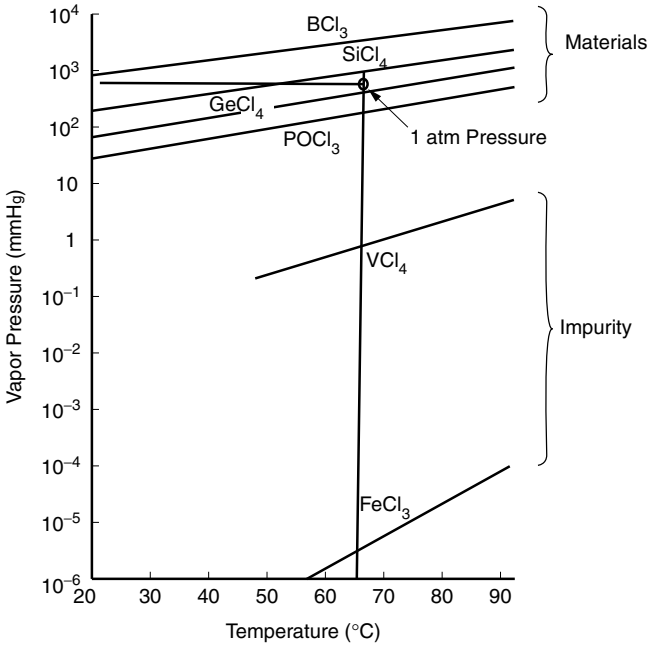


Figure 11.46 Relation of vapor pressures for metal halide additives and potential impurities. (After P. C. Schultz [48].)

below the container pressure will stay as liquid. For instance, if O_2 is bubbled into the $SiCl_4$ container at 65°C with 1 atmosphere pressure, $SiCl_4$ bubbles out as vapor but impurities such a VCl_4 or $FeCl_3$ stay in the liquid. Thus, $SiCl_4$ gas that is free from impurity is obtained. As seen in Fig. 11.46, the difference between the vapor pressures of the materials and that of the impurities is so large that purification is achievable without the need for tight temperature control.

The next step is to fabricate the preform out of these material gases. There are a number of approaches for doing this, which can be classified into three major groups depending on the way the deposit of the material is made.

In the first group, layers of material are deposited inside a silica tube, and depending on the method of heating the tube, they are further subdivided into the following three methods.

1. Modified chemical vapor deposition (MCVD) method.
2. Plasma chemical vapor deposition (PCVD) method.
3. Plasma-enhanced modified chemical vapor deposition (PMCVD) method.

The outside vapor deposition (OVD) method belongs to the second group, which deposits the layers of material on the outside of a mandrel. The mandrel is removed after the formation of the layers.

The third group is the vapor-phase axial deposition (VAD) method, which deposits the layer in the direction of the axis of the preform and therefore does not have length limitations.

Let us start with the MCVD method. The MCVD arrangement is shown in Fig. 11.47. The material gases are prepared by the CVD method. Material halides in liquid form, such as $SiCl_4$, $GeCl_4$, and $POCl_3$, are stored in separate reservoirs. Purified

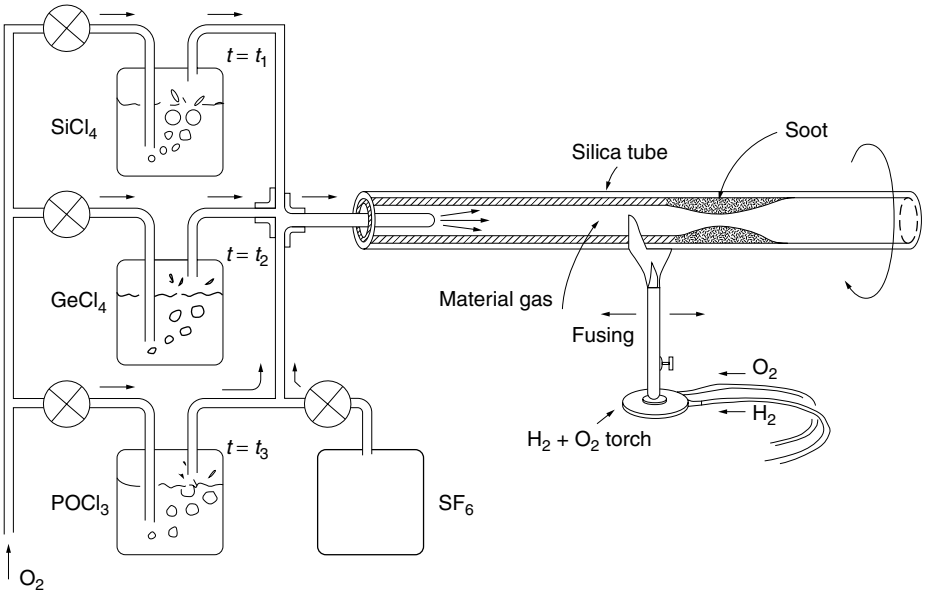


Figure 11.47 Modified chemical vapor deposition (MCVD) method.

oxygen gas is bubbled into these reservoirs at a prescribed pressure and temperature so that only the pure material gas can bubble out. The bubbled out material gases are mixed in prescribed proportions and led into the inside of a silica tube. The mixture proportions are manipulated by a computer-controlled mass flow manifold.

A chemical reaction takes place in vapor form inside the silica tube with the aid of an oxyhydrogen flame outside the tube. The metal halide vapors are oxidized, and small glass particles, which look like chalk powder, are deposited on the inside wall of the silica tube. The small glass particles are called "soot" and are formed by the following chemical reactions:

$$SiCl_4 + O_2 \rightarrow SiO_2 + 2Cl_2 \tag{11.244}$$

$$GeCl_4 + O_2 \rightarrow GeO_2 + 2Cl_2 \tag{11.245}$$

$$4POCl_3 + 3O_2 \rightarrow 2P_2O_5 + 6Cl_2 \tag{11.246}$$

Soot deposited ahead of the flame is fused to the glass wall as the flame passes. Both dopants GeO_2 and P_2O_5 raise the index of refraction of the glass, while F lowers the index of refraction. Core and cladding layers are formed by programmed control of the mass flow manifolds. The maximum achievable increase or decrease Δ in the index of refraction is limited because, with excessive amounts of dopant, the fiber cracks due to the difference in the heat expansion coefficient between the dopants and the silica glass. The maximum values of Δ that can be tolerated before cracking becomes a problem are $\Delta = 3.5\%$ for GeO_2-doped SiO_2, $\Delta = 1.2\%$ for P_2O_5-doped SiO_2, and $\Delta = 0.7\%$ for F-doped SiO_2 [49].

The preform of a graded-index fiber is completed after about 30 repetitions of the process of fusing the layers. The tube with the completed layer formation is then collapsed into a rod by heating.

Next, let us take a look at the PCVD method. The PCVD method uses the plasma generated in a microwave cavity instead of the oxyhydrogen flame as a source of heating. The cavity is cylindrical in shape and is resonated at a few gigahertz. Two holes are made on the sides of the cavity so that it can slide coaxially with the silica tube. Because of the high temperature of the plasma generated inside the cavity, the material gas becomes direct-deposited new glass on the tube wall, instead of by way of soot. Advantages of the PCVD method over the MCVD method are that, in the PCVD method, the deposition efficiency and speed are higher, there is greater control over profiles because the temperature can be more precisely controlled, and the tube need not be rotated. The disadvantage of the PCVD method is that the entire apparatus has to be under vacuum so that plasma can be generated inside the cavity, and the apparatus of the PCVD is more complex. Because of the high temperature of the plasma, the silica has to be preheated and this also creates additional complexity.

Next, the PMCVD method will be explained. PMCVD is quite similar to PCVD except that it uses an induction heater operated at a few megahertz. The advantage of the induction heater over the microwave cavity is that it does not require a vacuum. The temperatures obtainable with the induction heater are not high enough to fuse the glass directly to the tube wall, so that the process has to be assisted by an additional oxyhydrogen flame. The advantage of the PMCVD method over the original MCVD method is related to the advantage of heating from the inside (induction heating) over heating from the outside (flame). PMCVD has a much faster deposition rate, and larger

preforms can be made. The disadvantage of the PMCVD method is that the profile is more difficult to control than the MCVD and PCVD methods.

Next, the OVD method, which deposits the soot on a rotating carbon or ceramic rod, is explained.

Figure 11.48 shows a diagram of the OVD method. The material gases are prepared again by the CVD method. Material gases such as $SiCl_4$, $GeCl_4$, and $POCl_3$ are bubbled out with purified O_2. The material gases are brought into the torch region, fueled by H_2 and O_2. The soot is formulated by the following chemical reactions:

$$GeCl_4 + O_2 + 2H_2 = GeO_2 + 4HCl \tag{11.247}$$

$$SiCl_4 + O_2 + 2H_2 = SiO_2 + 4HCl \tag{11.248}$$

Note the difference from the case of MCVD where there is no chance that the material gases meet hydrogen. After the starting rod is pulled out, the soot is fused into a preform in a furnace. Cl_2 and helium gas are introduced into the furnace. Cl_2 removes OH ion impurities, and helium prevents bubble formation in the glass.

Compared to the MCVD method, it is possible to make larger preforms with the OVD method, which means longer continuous lengths of fiber can be drawn. The deposition speed is also faster with OVD than MCVD because of the direct interaction of material gases and flame gases that takes place in OVD. However, this direct interaction of H_2 and O_2 gases with the material gases means the likelihood of OH ion contamination in the glass is greater with OVD than MCVD, increasing the fiber loss. In OVD there is also the added step of removing the starting rod, which may produce irregularities in the refractive index profile of the core.

The last method of preform fabrication to be described is the VAD method. A diagram of this method is shown in Fig. 11.49a, and a photograph of the flame is

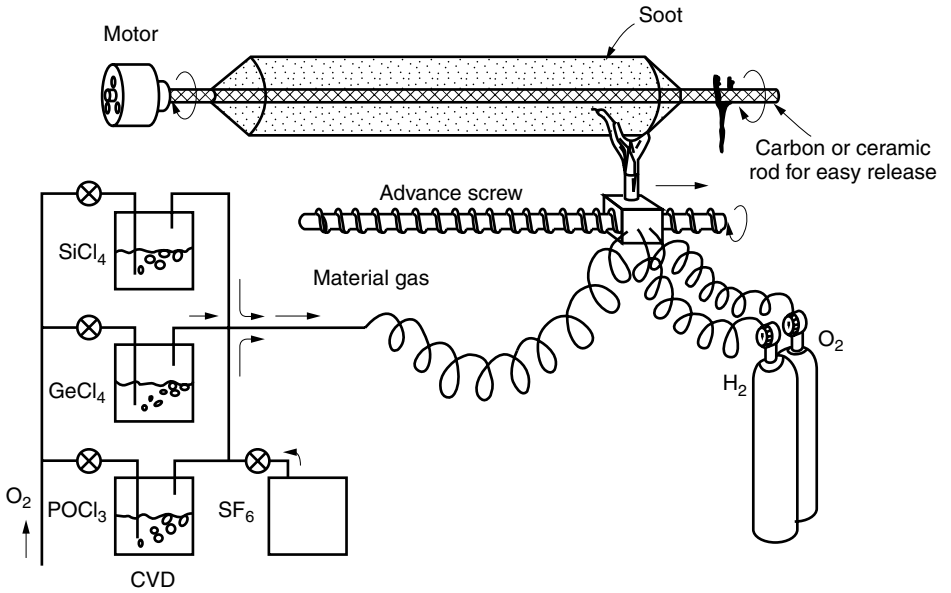
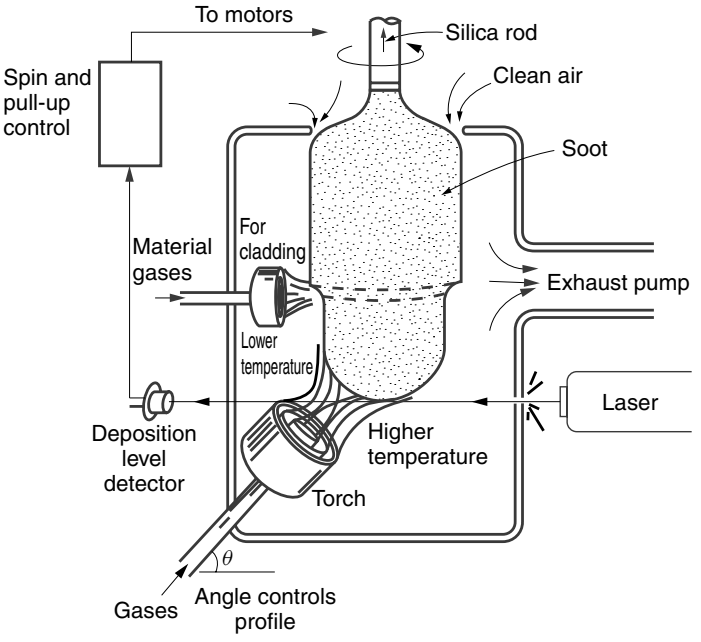
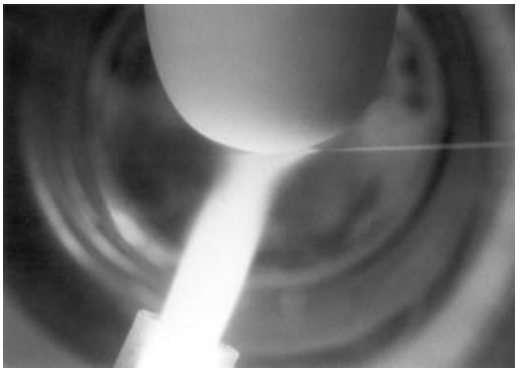


Figure 11.48 Outside vapor deposition (OVD) method.



(a)



(b)

Figure 11.49 Vapor-phase axial deposition (VAD) method. (a) Apparatus. (b) Deposit of the soot. A horizontal laser light monitors the position of the bottom of the soot. (Courtesy of H. Murata).

shown in Fig. 11.49b. The soot is deposited by the interaction of the flame with the material gases. The rod is rotated on its own axis and is slowly pulled upward as it grows. Theoretically speaking, there is no limit on the length of the soot rod, and the length of the rod can be arbitrarily long.

The torch is structured in multicoaxial layers, and the arrangement of the material and fuel gases is critical.

The flame is angled with respect to the rod so that a temperature differential is established between the center and the circumference of the rod. The temperature at

the center of the rod is highest because this spot is exposed to the flame throughout the revolution of the rod while a spot on the circumference is exposed to the flame only during a fraction of the revolution. The amount of deposition of GeO_2:SiO_2 soot is largest at the highest temperature. The desired profile of the index of refraction of the graded-index fiber can be achieved by critically adjusting the angle of the torch as well as the proportion of the material gases.

11.10.2.2 Drawing into an Optical Fiber

In the second stage of the process, an optical fiber is drawn from the preform rod. Figure 11.50 shows an apparatus for drawing an optical fiber. The lower end of the

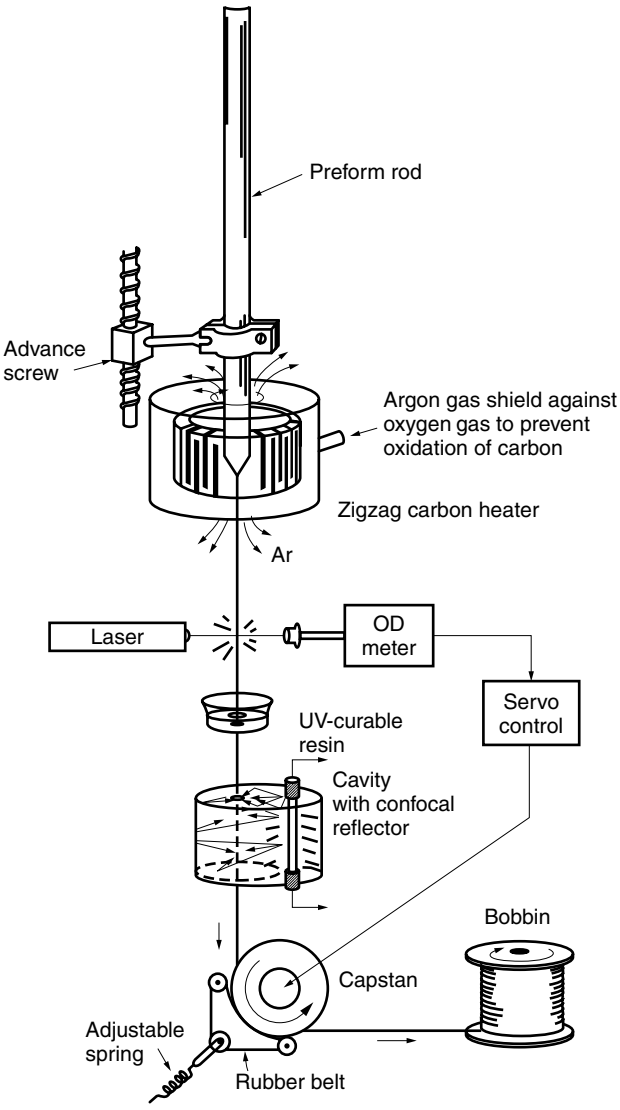


Figure 11.50 Apparatus for drawing a fiber.

preform is heated in a furnace to 2000–2100°C, which is a few hundred degrees above the melting temperature (1900°C) of the fiber. The furnace makes it possible to draw an optical fiber with a specified diameter in one step. The furnace is made out of a fat graphite filament bent like a cage surrounding the preform. Argon gas is blown into the furnace so as to protect the filament from oxidation burnout.

The fiber diameter is controlled by the speed with which the fiber is pulled. The pulling speed is adjustable by the number of revolutions per minute (rpm) of the capstan. The feedback signal of the fiber diameter monitor controls the rpm of the capstan. The fiber is pushed against the wheel of the capstan by a thin rubber belt that also rotates at the same speed as the capstan wheel. The diameter of a drawn fiber is normally 125 μm.

The drawn fiber is immediately jacketed by a UV-curable resin. This reduces the chance of foreign particles attaching themselves to the bare fiber. Foreign particles on the fiber surface become sources of small cracks, weakening the optical fiber.

The resin jacket is cured inside a confocal reflector UV light cavity. The fiber passes along one of the lines of focus of a cylindrical cavity with an elliptical cross section while a bar UV light is positioned along the other line of focus. The outer diameter of the jacketed fiber is normally either 0.25–0.35 mm or 0.8–1.0 mm.

The fiber is then reeled onto a spool either with or without going through the screen test spool. The screen test is performed by bending through a set of small wheels under tension.

The jacketed fiber is then cabled for telecommunication purposes.

11.11 CABLING OF OPTICAL FIBERS

The mechanical specifications of cabling [49,50] are that, under normal conditions of use, the elongation should be less than 0.2% and the bending radius should be no less than 3 cm. That is, the performance of the fiber is guaranteed for these specifications. As a matter of fact, when subjected to extreme tests, one piece of fiber can pick up a 5-kg weight resulting in a 5% elongation, or the fiber can be bent to curvature as small as 2–3 mm radius without breaking. The cabling specifications for the values of elongation or bending radius are far from the physical limits of the fiber.

The general principle of reducing the tension of a fiber inside a cable is stranding and is the same as that of the telephone cord shown in Fig. 11.51. A telephone cord is coiled so that even a reckless child will not rip the cord off the telephone. The difference between a telephone cord and an optical fiber cable is that the optical fiber cable is equipped with a strength member inside the cable to limit the elongation and a sheath outside to prevent damage during installation and operation.

Another important concern with optical fibers is the increase in microbending loss at lower temperatures. The temperature expansion coefficient of glass is of the order of 10^{-7}, while that of most of the jacket materials is of the order of 10^{-4}. As the temperature is lowered, the jacket shrinks one thousand times more than the glass fiber, and the optical fiber is subjected to many small bends that induce the so-called microbending loss. One solution to this problem is to introduce a buffer layer. The buffer-covered fiber has two different kinds of coatings and has minimum microbending loss. The first inner coating is a soft resin coating that acts as a slipping-buffer layer. The second coating is a hard resin coating that protects the fiber from external forces.

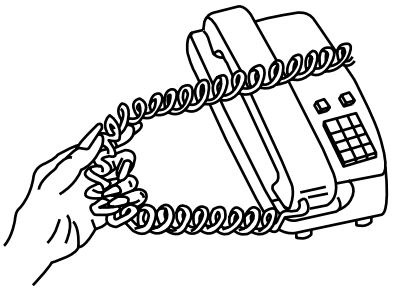


Figure 11.51 Telephone cord and cabling fibers share the same philosophy.

With this configuration, the fiber can slip with respect to the hard coating and the microbending is greatly alleviated.

The complexity of the cable design depends on the application. In laboratory experiments, a single coated fiber may suffice, whereas in a telecommunication link, the overall cable structure may house a number of individual cabled fibers. Let us now describe this hierarchy of cable design, which is broadly classified into the following three categories:

1. Element fibers.
2. Unit fiber cable.
3. Multiunit fiber cable.

The configurations for the element fibers are shown in the top row of Fig. 11.52, while those for the unit fiber cable and the multiunit fiber cable are shown in the middle and bottom rows of Fig. 11.52, respectively. The element fibers will be explained first.

The 0.25-mm-OD monocoated fiber (Fig. 11.52a) is thin with an outer diameter of only 0.25 mm. Even though it has the merit of low production cost, it has such demerits as high microbending loss and difficulty of handling and stranding.

The 0.25-mm-OD buffer-covered fiber (Fig. 11.52b) also has an outer diameter of 0.25 mm. The jacket is composed of two layers, a soft inner layer and a harder outer layer. As mentioned previously, this configuration minimizes the increase in microbending loss at lower temperatures.

The 0.9-mm-OD buffer-covered fiber (Fig. 11.52c) also has soft and hard layers, but the jacket is thicker, giving an outer diameter of 0.9 mm. This fiber has the merit of smaller microbending loss and easier handling compared to its thinner counterpart.

The loose-tube covered fiber (Fig. 11.52d) has a significantly larger buffer layer filled with either jelly or nothing. The fiber used for the inside is buffer-covered 0.25-mm-OD fiber. The microbending characteristics are excellent and no recognizable increase in the microbending is observed even at very low outside temperatures. There is, however, some degree of difficulty in removing the tube at the time of splicing because of the combination of the hard outside tube and the thin fiber inside the tube.

The middle row of Fig. 11.52 shows bundles of fibers at an intermediate stage to the final cable configurations shown in the bottom row. They are explained as follows.

For the unit cable (Fig. 11.52e), six 0.9-mm-OD buffer-covered fibers are stranded around a strength member of a copper-plated steel wire. The stranding pitch is 10 to 40 cm. Stranding provides flexible stretching of the fiber analogous to the telephone

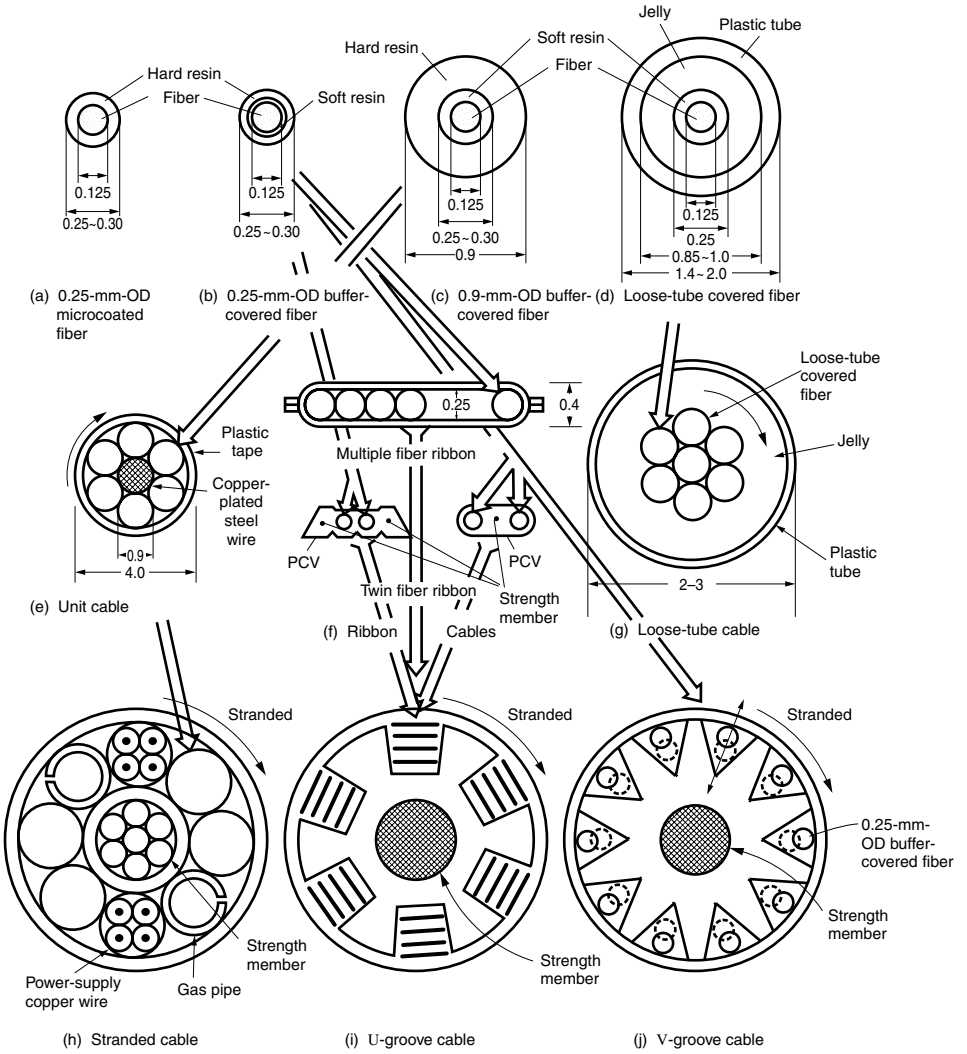


Figure 11.52 Hierarchy of optical fiber cables (dimensions in mm). (After H. Murata [49].)

cord effect mentioned earlier. The stranded fibers are held together by wrapping with plastic tape.

For the ribbon cable (Fig. 11.52f), five to twelve 0.25-mm-OD buffer-covered fibers are arranged in parallel and are sealed by adhesive plastic tape. An advantage of the fiber ribbon design is that it is possible to splice all fibers in one operation by using a mass fusion splicing machine. The twin fiber ribbons are designed for indoor use. The potential for easy speedy splicing in the field of a large number of fibers is the greatest merit of the ribbon fiber.

For the loose-tube cable (Fig. 11.52g), seven 0.25-mm-OD buffer-covered fibers are stranded into one bundle and put into a 2–3-mm-diameter plastic tube. The tube is filled with jelly and the fiber bundle moves freely inside the tube, escaping from external stresses.

Finally, the cabling structures shown in the bottom row of Fig. 11.52 will be explained.

For the stranded cable (Fig. 11.52h), several unit cables are stranded into a cable with a stranding pitch more or less the same as that of the unit cable (10–40 cm). Figure 11.50h is an example of a 24 optical fiber cable. The cable consists of four 6 fiber unit cables, two 4 copper conductor cables for the electrical power supply to the repeater stations, and a stranded copper-plated steel wire as a strength member. Stranding of the steel wire makes the cable more flexible. The last two pipes are gas pipes for keeping moisture off the inside of the cable.

For the U-groove cable (Fig. 11.52i), rectangular grooves are helically made on a polyethylene rod. Ribbon fibers are stacked inside the grooves. A special feature of this cable is its capacity for accommodating a large number of fibers in one cable. The cable shown in the figure contains twenty-four 5 ribbon fibers, and the total number of fibers is 120.

For the V-groove fiber (Fig. 11.52j), the grooves are helically made on a polyethylene rod. Each V-groove can accommodate either a single fiber or a multiple number of fibers. A strength member wire is placed in the center of the cable. The grooved rod not only makes compartments for fibers but also protects the fibers from external forces. Fibers are first put into the groove and then a jelly coating is applied.

Since the rod is helically grooved, the fibers in the grooves behave as if they are stranded. As the cable is stretched, fibers move toward the center of the cable, as indicated by the dotted lines in the figure.

The V-grooved cable can be treated as a unit V-groove cable, and a multiple number of such cables can be stranded into an even bigger size cable, thus accommodating a large number of fibers.

11.12 JOINING FIBERS

Usually, reels of optical fiber cables have to be joined in order to make a cable long enough for a communication system. Basically, there are two kinds of fiber joints. One is a permanent joint, and the other is a demountable joint. The permanent type is not intended to disconnect once installed, and an example is the installation of trunk lines connecting reels of cables. Fibers are spliced together. The demountable type of joint is intended to disconnect occasionally, and an example is the connection to the terminal equipment. Without exception, permanent joints are made by splicing fibers together, as this ensures long-life reliability. Connectors are generally used for nonpermanent joints.

11.12.1 Splicing Fibers

The most common method of splicing two fibers [51] is with an electric arc because it can provide high temperature with cleanliness, accuracy, and reproducibility. A mass fusion splicing machine whose mechanism is illustrated in Fig. 11.53 can splice five fibers in a ribbon automatically with minimal human intervention. Figure 11.53a is a step-by-step illustration of mass fusion splicing, and Fig. 11.53b illustrates the layout of the automated machine. The numbers in Fig. 11.53a indicate the steps in the following explanation.

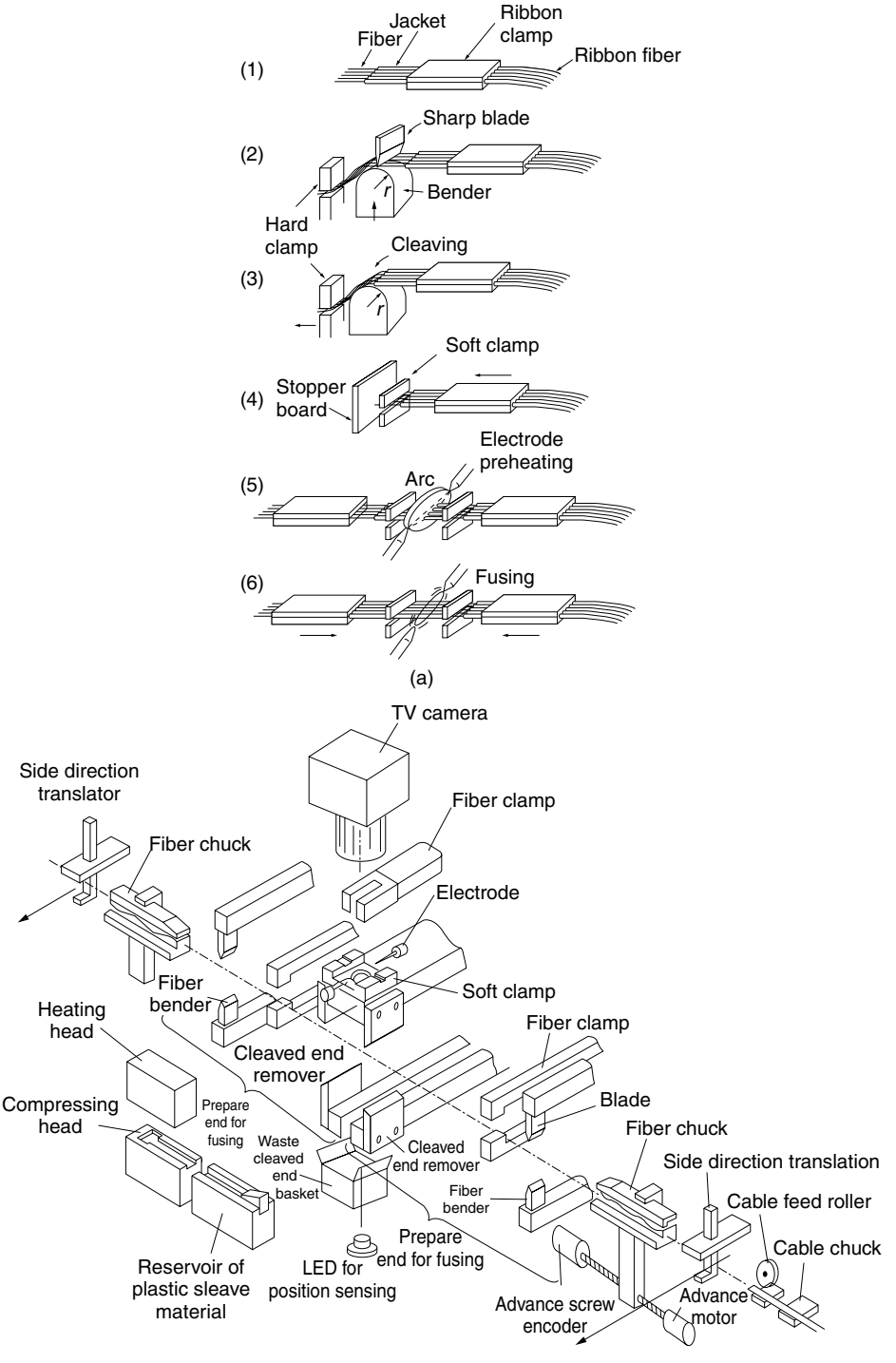


Figure 11.53 (a) Steps for mass fusion splicing. (b) Mass fusion splicing machine for five fiber ribbon. (After M. Matsumoto et al. [51].)

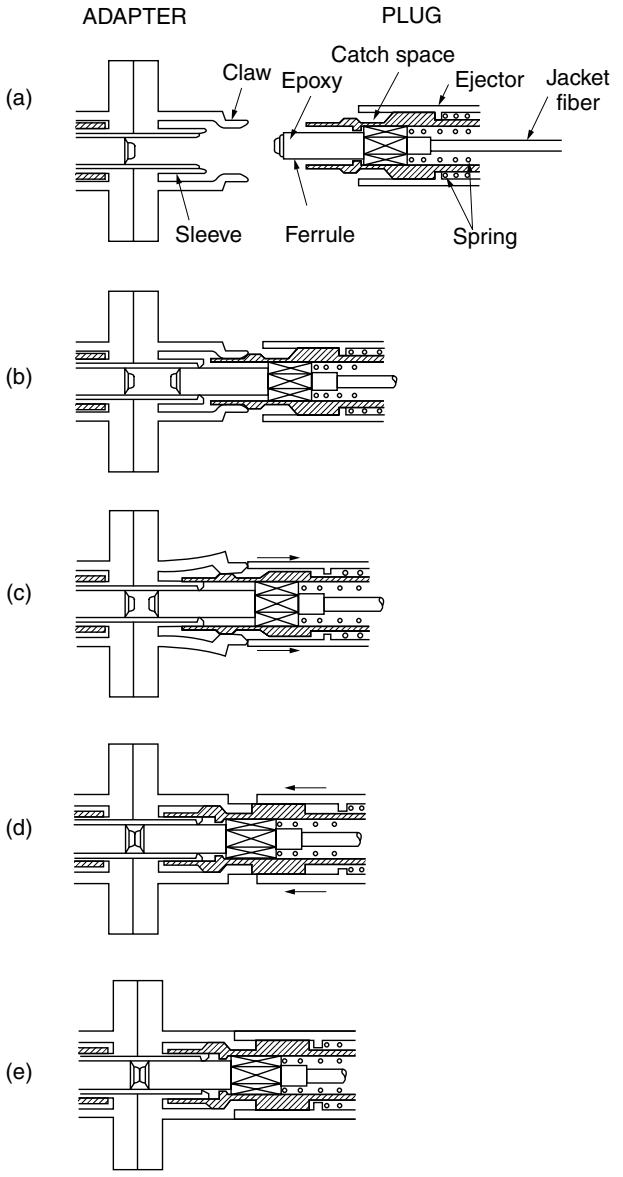


Figure 11.54 Cross section of single port optical fiber connector and steps for making the connection. (After T. Kakii et al. [53].)

(1) The ribbon fiber is first held by a fiber chuck and is cut to a predetermined length, and then its jacket is removed by a jacket stripper to expose an array of fibers. The ends of the fibers are clamped in preparation for cleaving. Cleaving action consists of scoring, bending, and breaking the fibers. Cleaving produces clean, smooth fiber ends, and the cleaved surfaces are reasonably perpendicular to the fiber axis. (2) The fibers are first individually scored by either a diamond or a high-speed steel blade. (3) The scored fibers are then pushed upward by a semicylindrical bed so that the fibers are

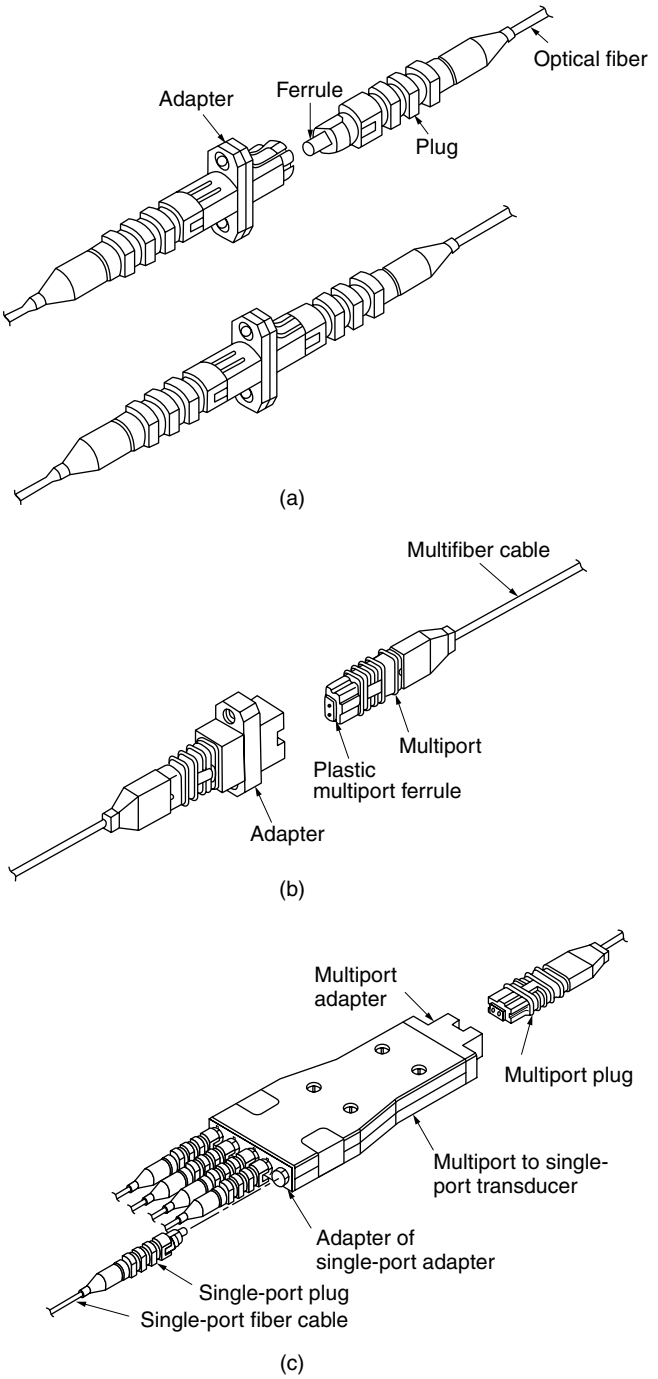


Figure 11.55 Structure of connectors of various types. (a) Single-port optical fiber connector. (b) Multiport optical fiber connector. (c) Multiport to single-port adapter (After T. Kakii et al. [53].)

stretched and bent in an arc, then further pulled, and broken off to complete cleaving. A monitoring TV image not only examines the cleaved surfaces but also aligns the left and right cores. (4) The unjacketed section of the fibers are held by soft clamps, and are pushed against the stopping plate that appears from the bottom, in order to eliminate any unevenness of the lengths of the cleaved fibers.

After the stopping plate is withdrawn to the bottom, (5) the ends of the fibers are preheated by an electric arc. After preheating the ends, the ribbon fibers are brought together until the ends of the fibers touch; then (6) the fibers are fused together by the electric arc. After the fused fibers are proof-tested by pulling apart with a predetermined tension, (7) the spliced section is reinforced by a bonding agent and the entire splicing process is completed.

Factors that influence the splicing loss are (a) the transversal offset between the two fibers being spliced, (b) the angular tilt of the two fibers, (c) the differences in the core diameters, and (d) the differences in the refractive index profiles of the index of refraction. In order to suppress the splicing loss of a single-mode fiber to less than 0.1 dB, the offset distance should be less than 0.5 μm and the tilt angle has to be less than 0.5° [52].

11.12.2 Optical Fiber Connector

Fiber connectors [53] are used instead of splices at terminal relay stations, at the locations of home subscribers, for ease of operational inspection, in new installations, and for rerouting of communication lines.

Points to be considered in the design of optical fiber connectors are the following:

1. Connectors must be able to withstand outdoor use and normal handling.
2. They should be extremely reliable and durable.
3. The design should strive to minimize connector loss.
4. For systems requiring large numbers of connectors, it is important to keep the cost of fabricating the connectors low.

The major challenge in designing optical fiber connectors is to design connectors that can be mass-produced despite severe machining tolerances.

Figure 11.54a shows the cross section of the connector, and the rest of the diagrams illustrate how the spring action interlocks the two ferrules to the center adapter, maintaining the required precision in offset as well as tilt angle between the two ferrules.

Figure 11.55a shows how the single-port fiber connector looks when the connection is completed. Figure 11.55b shows the mating of a multiport connector, in which a multiport plug is inserted into a multiport adapter. Figure 11.55c shows a transducer that converts a multiport connector into separate single-port connectors. Such a transducer is useful for dropping individual subscriber lines from a ribbon fiber cable.

PROBLEMS

11.1 A laser beam with diameter $2a = 0.8$ mm is to be focused onto the end of an optical fiber using a convex lens (Fig. P11.1). The diameter of the core is 9.5 μm.

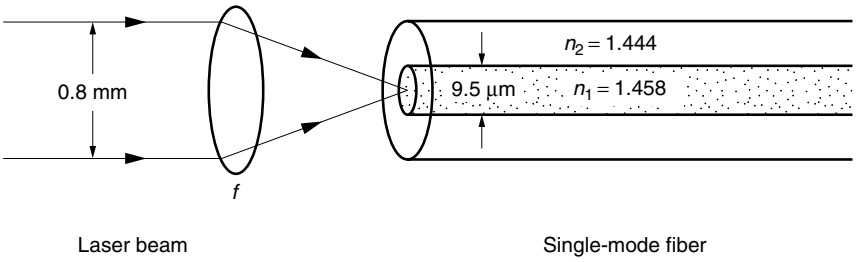


Figure P11.1 Selection of the focal length of a focusing lens.

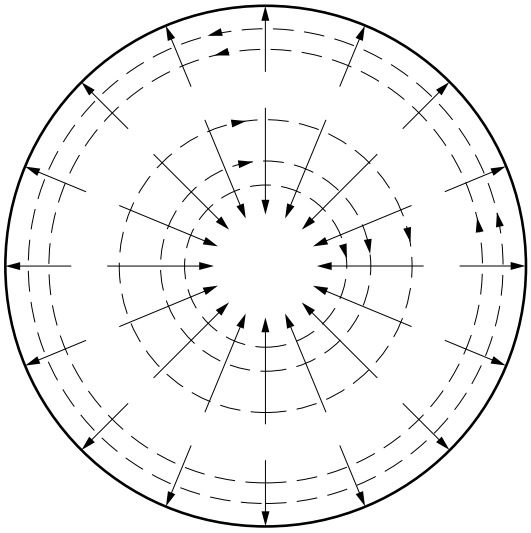


Figure P11.3 A pattern in a step-index fiber. Solid line for *E* field dotted line for *H* field.

The index of refraction of the core is $n_1 = 1.458$. The index of refraction of the cladding is $n_2 = 1.444$. The wavelength of the light is $\lambda = 1.55$ μm. What is the range of focal lengths of the lens that can be used for this purpose? Include consideration of the diffraction-limited resolution of the focused image.

11.2 Find the asymptotes of the characteristic curves of the $HE_{1\mu}$ modes.

11.3 What is the designation of the mode shown in Fig. P11.3?

11.4 **(a)** Draw a rough graph of the characteristic equations of the $HE_{3\mu}$ and $EH_{1\mu}$ modes.

(b) Verify the cutoff values of V_c indicated in Fig. 11.18 up to the seventh mode using the graphs in the text and the graph drawn for part (a) of this problem.

11.5 Graphically construct the field pattern of the LP_{21} mode from those of the hybrid modes.

11.6 Find the mode numbers of the $LP_{\nu\mu}$ modes shown in Fig. P11.6. Also find the hybrid modes that are the constituents of these $LP_{\nu\mu}$ modes.

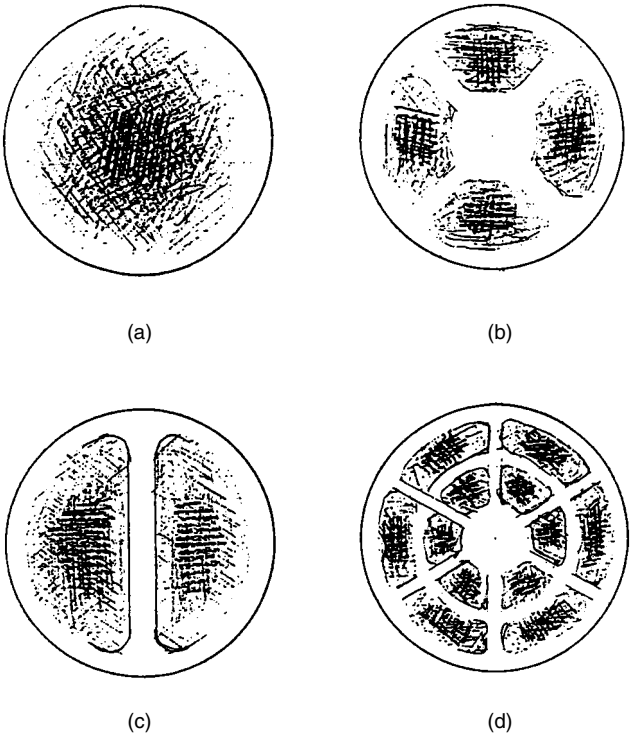


Figure P11.6 Field patterns of $LP_{\nu\mu}$ modes.

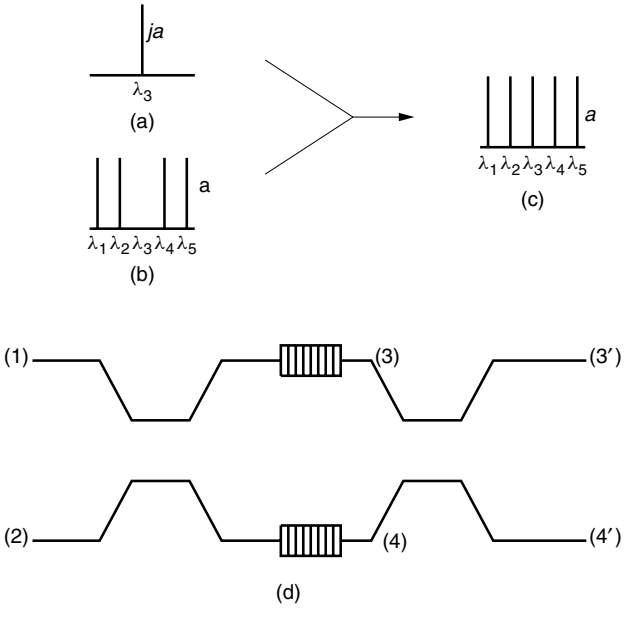


Figure P11.8 Insert the λ_3 channel in (a) into the main stream of the WDM in (b) as shown in (c) using the device shown in (d).

11.7 Draw the mode pattern of HE_{22}.

11.8 Figure P11.8 shows a multiplexer made out of a Bragg grating fiber and a Mach–Zehnder interferometer. Channel λ_3 shown in Fig. P11.8a has to be multiplexed with the rest of the WDM stream shown in Fig. P11.8b to form the filled WDM stream in Fig. P11.8c. To which ports of the device shown in Fig. P11.8d should the λ_3 signal and the rest of the WDM stream be applied?

REFERENCES

1. C. K. Kao and G. A. Hockham, "Dielectric-fiber surface waveguides for optical frequencies," *Proc. IEE (London)* **113**(7), 1151–1158 (1966).
2. F. J. Denniston and P. K. Runge, "The glass necklace," *IEEE Spectrum*, 24–27 (Oct. 1995).
3. N. S. Bergano and H. Kidorf, "Global undersea cable networks," *Optics and Photonics News*, **12**(3), 32–35 (2001).
4. T. Miya, Y. Terunuma, T. Hosaka, and T. Miyashita, "Ultimate low-loss single-mode fibre at 1.55 μm," *Electron. Lett.* **15**(4), 106–108 (1979).
5. N. J. Pitt, A. Marshall, J. Irven, and S. Day, "Long term interactions of hydrogen with single-mode fibers," 10th ECOC, pp. 308–309, Sept. 1984.
6. M. Kuwazuru, Y. Namihira, K. Mochizuki, and Y. Iwamoto, "Estimation of long-term transmission loss increase in silica-based optical fibers under hydrogen atmosphere," *J. Lightwave Technol.* **6**(2), 218–225 (1988).
7. G. Tanaka, H. Yokota, H. Kanamori, Y. Ishiguro, Y. Suetsugu, H. Takada, S. Ito, S. Tanaka, Y. Asano, M. Kyoto, M. Watanabe, and K. Yano, "Characteristics of pure silica core single mode optical fiber," *Sumitomo Electric Tech. Rev.* **26**, 43–53 (Jan. 1987).
8. S. Tanaka, and M. Honjo, "Long-term reliability of transmission loss in optical fiber cables," *J. Lightwave Technol.* **6**(2), 210–217 (1988).
9. I. Yoshida, M. Watanabe, G. Tanaka, H. Yokota, and H. Kanamori, "Temperature and dose rate dependence of radiation damage," *12th European Conference of Optical Communication (ECOC 12)*, Barcelona, Spain, p. 41, 1986.
10. H. Bueker, F. W. Haesing, F. Pfeiffer, and H.-J. Schmitz, "A fiber optic twin sensor for dose measurements in radiation therapy," *Opt. Rev.* **4**(1A), 130–132 (1997).
11. S. Kobayashi, S. Shibata, N. Shibata, and T. Izawa, "Refractive-index dispersion of doped fused silica," *1977 International Conference on Integrated Optics and Optical Fiber Communication*, Tokyo, B8-3, pp. 309–312, 1977.
12. D. B. Keck, "Fundamentals of optical waveguide fibers," *IEEE Commun. Mag.*, 17–22 (May 1985).
13. M. Sato, T. Ishigure, and Y. Koike, "Thermally stable high-bandwidth graded-index polymer optical fiber," *J. Lightwave Tech.* **18**(7), 952–958 (2000).
14. B. J. Eggleton, P. S. Westbrook, C. A. White, C. Kerbage, R. S. Windeler, and G. L. Burdge, "Cladding-mode-resonances in air-silica microstructure optical fibers," *J. Lightwave Technol.* **18**(8), 1084–1100 (2000).
15. T. M. Monro, D. J. Richardson, N. G. R. Broderick, and P. J. Bennett, "Modeling large air fraction holey optical fibers," *J. Lightwave Technol.* **18**(1), 50–56 (2000).
16. M. J. Steel and R. M. Osgood, Jr., "Polarization and dispersive properties of elliptical-hole photonic crystal fibers," *J. Lightwave Technol.* **19**(4), 495–503 (2001).
17. N. Kishi and E. Yamashita, "Optical broadband phase devices using axially nonsymmetric W-type optical fibers," *J. Lightwave Technol.* **16**(2), 301–306 (1998).

18. L. G. Van Uitert, and S. H. Wemple, "$ZnCl_2$ glass: a potential ultralow-loss optical fiber material," *Appl. Phys. Lett.* **33**(1), 57–59 (1978).
19. D. Marcuse, *Light Transmission Optics*, 2nd ed. Van Nostrand Reinhold, New York, 1982.
20. D. Gloge, "Weakly guiding fibers," *Appl. Opt.* **10**(10), 2252–2258 (1971).
21. A. Kapoor and G. S. Singh, "Mode classification in cylindrical dielectric waveguides," *J. Lightwave Tech.* **18**(5), 849–852 (2000).
22. K. O. Hill, B. Malo, D. C. Johnson, and F. Bilodeau, "A novel low-loss inline bimodal fiber tap: wavelength selective properties," *IEEE Photonics Technol. Lett.* **2**(7), 484–486 (1990).
23. C. D. Poole, J. M. Wiesenfeld, and D. J. DiGiovanni, "Elliptical-core dual-mode fiber dispersion compensator," *IEEE Photonics Technol. Lett.* **5**(2), 194–197 (1993).
24. K. O. Hill, B. Malo, F. Bilodeau, and D. C. Johnson, "Photosensitivity in optical fibers," *Annu. Rev. Mater. Sci.* **23**, 125–157 (1993).
25. W. W. Morey, G. A. Ball, and G. Meltz, "Photoinduced Bragg gratings in optical fibers," *Opt. Photonics News*, 8–14 (Feb. 1994).
26. L. Dong, J.-L. Archambault, L. Reekie, P. St. Russell, and D. N. Payne, "Single pulse Bragg gratings written during fibre drawing," *Electron. Lett.* **29**(17), 1577–1578 (1993).
27. P. J. Lemaire, A. M. Vengsarkar, W. A. Reed, and D. J. DiGiovanni, "Thermal enhancement of UV photosensitivity in H_2-loaded optical fibers," *OFC'95 Tech. Digest* **8**, 158–159 (1995).
28. F. Bilodeau, B. Malo, J. Albert, D. C. Johnson, K. O. Hill, Y. Hibino, M. Abe, and M. Kawachi, "Photosensitization of optical fiber and silica-on-silicon/silica waveguide," *Opt. Lett.* **18**(12), 953–955 (1993).
29. D. C. Johnson, F. Bilodeau, B. Malo, K. O. Hill, P. G. J. Wigley, and G. I. Stegeman, "Long-length, long-period, rocking filters fabricated from conventional monomode telecommunications optical fiber," *Opt. Lett.* **17**(22), 1635–1637 (1992).
30. B. Malo, D. C. Johnson, F. Bilodeau, J. Albert, and K. O. Hill, "Single-excimer-pulse writing of fiber gratings by use of a zero-order nulled phase mask: grating spectral response and visualization of index perturbations," *Opt. Lett.* **18**(15), 1277–1279 (1993).
31. G. A. Ball and W. W. Morey, "Continuously tunable single-mode erbium fiber laser," *Opt. Lett.* **17**(6), 420–422 (1992).
32. B. Malo, S. Thériault, D. C. Johnson, F. Bilodeau, J. Albert, and K. O. Hill, "Apodised in-fiber Bragg grating reflectors photoimprinted using a phase mask," *Electron. Lett.* **31**(3), 223–224 (1995).
33. F. Bilodeau, D. C. Johnson, S. Thériault, B. Malo, J. Albert, and K. O. Hill, "An all-fiber dense-wavelength-division multiplexer/demultiplexer using photoimprinted Bragg gratings," *IEEE Photonics Technol. Lett.* **7**(4), 388–390 (1995).
34. K. P. Koo and A. D. Kersey, "Noise and cross talk of a 4-element serial fiber laser sensor array," *OFC'96 Techn. Digest* **2**, 266–267 (1996).
35. Y. Rao, D. J. Webb, D. A. Jackson, L. Zhang, and I. Bennion, "In-fiber Bragg-grating temperature sensor system for medical applications," *J. Lightwave Technol.* **15**(5), pp. 779–785 (1997).
36. R. E. Collin, *Foundations for Microwave Engineering*, 2nd ed., McGraw-Hill, New York, 1992.
37. D. Gloge, "Dispersion in weakly guiding fibers," *Applied Optics* **10**(11), 2442–2445 (1971).
38. M. Ogai, Y. Yano, A. Iino, M. Hashimoto, K. Kokura, and K. Matsubara, "Low loss dispersion shifted fiber with fluorine-doped cladding and step like index profile," OFC'86 Post Dead Line Paper 4, 19–22 (Feb. 1986).
39. H. Murata, *Optical Fibers Cable*, Ohm-Sha, Tokyo (1987).

40. K. Takiguchi, S. Kawanishi, H. Takara, K. Okamoto, K. Jinguji, and Y. Ohmori, "Higher order dispersion equalizer of dispersion shifted fiber using a lattice-form programmable optical fiber," *Electron. Lett.*, **32**(8), 755–757 (1996).
41. A. E. Willner, "Mining the optical bandwidth," *IEEE Spectrum*, 32–41 (Apr. 1997).
42. R. Kashyap, S. V. Chernikov, P. F. McKee, and J. R. Taylor, "30 ps Chromatic dispersion compensation of 400 fs pulses at 100 Gbit/s in optical fibers using an all fiber photoinduced chirped grating," *Electron. Lett.* **30**(13), 1078–1080 (1994).
43. D. Garthe, R. E. Epworth, W. S. Lee, A. Hadjifotiou, C. P. Chew, T. Bricheno, A. Fielding, H. N. Rourke, S. R. Baker, K. C. Byron, R. S. Baulcomb, S. M. Ohja, and S. Clements, "Adjustable dispersion equalizer for 10 and 20 Gbit/s over distances up to 160 km," *Electron. Lett.* **30**(25), 2159–2160 (1994).
44. P. A. Krug, T. Stephens, G. Yoffe, F. Ouellette, P. Hill, and G. Dhosi, "270 km Transmission at 10 Gbits/s in non-dispersion-shifted fiber using an adjustable chirped 120 mm fiber Bragg grating dispersion compensator," *OFC'95, Tech. Digest* **8**, 428–431 (1995).
45. C. D. Poole, J. M. Wiesenfeld, A. R. McCormick, and K. T. Nelson, "Broadband dispersion compensation by using the higher-order spatial mode in a two-mode fiber," *Opt. Lett.* **17**(14), 985–987 (1992).
46. L. Jacomme, "A model for ray propagation in a multimode graded index-fiber," *Opt. Commun.* **14**, 134–138 (1975).
47. K. Iizuka, *Engineering Optics*, 2nd ed., Springer-Verlag, New York, 1988.
48. P. C. Schultz, "Vapor phase materials and processes for glass optical waveguides," *Fiber Optics*, edited by B. Bendow and S. S. Mitra, Plenum Press, pp. 3–31, 1979.
49. H. Murata, *Handbook of Optical Fibers and Cables*, Marcel Dekker, New York, 1988.
50. K. Hogari, S. Furukawa, K. Tomiyama, Y. Nakatsuji, S. Koshio, and K. Nishizawa, "Optical fiber cables for residential and business premises," *J. Lightwave Technol.* **16**(2), 207–213 (1998).
51. M. Matsumoto, M. Tachikura, T. Katagiri, M. Sato, A. Ide, and M. Miyauchi, "Fully automatic speed fusion splice machine for single mode optical fiber ribbons," *Proc. ECOC'87* **1**, 155–158 (1987).
52. T. Katagiri, M. Tachikura, and Y. Murakami, "Evaluation of fiber slippage and core-axis offset distribution in optical fiber mechanical splice," *J. Lightwave Technol.* **19**(2), 195–204 (2001).
53. T. Kakii, K. Kashihara, T. Komiya, S. Suzuki, Y. Iwamoto, and Y. Sakata, "Development of optical connectors for GI-type subscriber fiber optic cables," *Sumitomo-denki* **132**, 51–59 (1988).

12

DETECTING LIGHT

This chapter begins with a description of light detectors used in optics laboratories (Sections 12.1 through 12.3). Then, special purpose detectors used primarily for optical communication systems are described. These include the PIN photodiode and APD detectors (presented in Section 12.4).

The remaining sections deal with various methods of detection employed in fiber-optic communication systems. Direct detection is a simple method that measures only the intensity of light, and this method is described in Section 12.5. Sections 12.6 to 12.8 are devoted to coherent detection, which can measure not only the intensity but also the phase and frequency of the light. Lastly, Section 12.9 looks at phase and polarization jitters.

The signal to noise ratio (S/N) of the detection system will be presented in connection with optical communication systems in Chapter 16.

12.1 PHOTOMULTIPLIER TUBE

Light detectors convert light energy into an electrical signal. The photomultiplier tube (PMT) has an ultrahigh sensitivity and is widely used in optics research laboratories. The sensitivity is so high that it can almost detect a single photon.

Figure 12.1 explains the structure of the photomultiplier tube. The photomultiplier tube is a combination of the cathode, which is the photon–electron converter, and the dynodes, which act as amplifiers of the converted electrons.

The photoelectric converter section consists of the cathode and the first dynode. The cathode surface is coated with a low-work-function material, such as alkali metals. When the incident photons bombard the cathode, electrons are ejected from the surface. Emission, however, takes place only when a quantum $h\nu$ of the incident light is larger than the work function ϕ of the cathode material:

$$\phi < h\nu \tag{12.1}$$

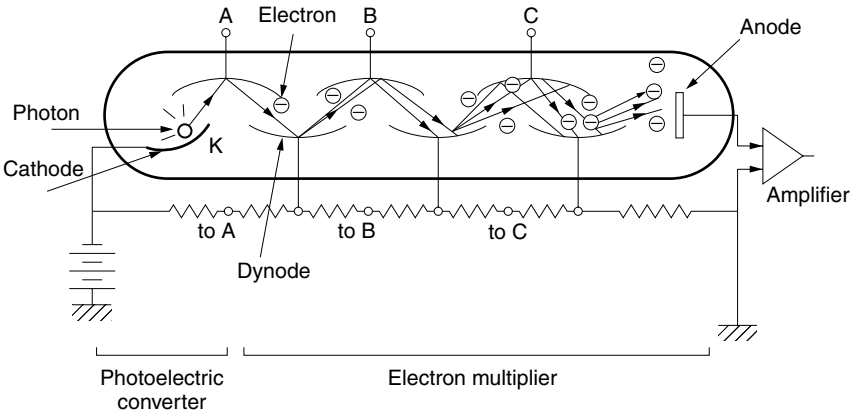


Figure 12.1 Photomultiplier tube. Points A, B, C ... at the bottom are connected, respectively, to the A, B, C, ... dynodes at the top.

where h is Planck's constant 6.63×10^{-34} J · s. Thus, the coating material of the cathode determines the range of wavelength for usable light. A number of materials are available in the visible wavelength range but, unfortunately, high-sensitivity materials have not been found in the infrared region, where the optical communication system operates.

The electrons ejected from the cathode are accelerated by the high voltage between the cathode and the first dynode and then strike the surface of the dynode with high energy. Secondary electrons are emitted from the dynode in larger numbers. With a series of dynodes operating at gradually higher voltages, the number of secondary electrons are multiplied at each dynode. The multiplied secondary electrons are collected by the anode whose potential is typically a few kilovolts higher than the cathode.

Next, the sensitivity of the photomultiplier tube will be calculated. Let W be the energy of the light entering the cathode of the photomultiplier tube. The number of photons in terms of the incident light energy is $W/h\nu$.

Let η be the quantum efficiency, which is defined as the ratio of the number of generated electrons to incident photons. Each electron carries charge e. The total amount of charge Q accumulated at the cathode is

$$Q = \eta \frac{e}{h\nu} W \tag{12.2}$$

The current from the cathode is obtained by taking derivatives of both sides of Eq. (12.2),

$$i_{\text{cathode}} = \eta \frac{e}{h\nu} P \tag{12.3}$$

where P is the incident light power.

If in each dynode stage one electron generates δ secondary electrons, the total multiplication M from the N dynodes is

$$M = \delta^N \tag{12.4}$$

Thus, the output current from the photomultiplier tube is

$$i = \eta \frac{e}{h\nu} MP \tag{12.5}$$

Finally, the responsivity R of the photomultiplier tube, which is the output current with unit incident light power, is

$$R = \eta \frac{e}{h\nu} M \text{ A/W} \tag{12.6}$$

where A/W is amperes/watt. The responsivity is used to characterize a photomultiplier tube.

Typical values are inserted into Eq. (12.6) to calculate the responsivity of the photomultiplier.

$\eta = 0.01$
$\lambda = 0.63$ μm or $\nu = 4.76 \times 10^{14}$ Hz
$\delta = 5$
$N = 10$

$$R = 0.01 \frac{(1.6 \times 10^{-19})(5^{10})}{(6.63 \times 10^{-34})(4.76 \times 10^{14})}$$
$$\div 5 \times 10^4 \text{ A/W}$$

The responsivity of a PIN photodiode is typically $R = 0.5$ A/W and the photomultiplier tube is significantly more sensitive.

The internal impedance of a photomultiplier being as high as 10 MΩ, the matched load impedance can be selected high so that a high output voltage can be achieved. For instance, for 1 pW (−90 dBm) of input light, the output voltage across the 10-MΩ load resistance R_L is

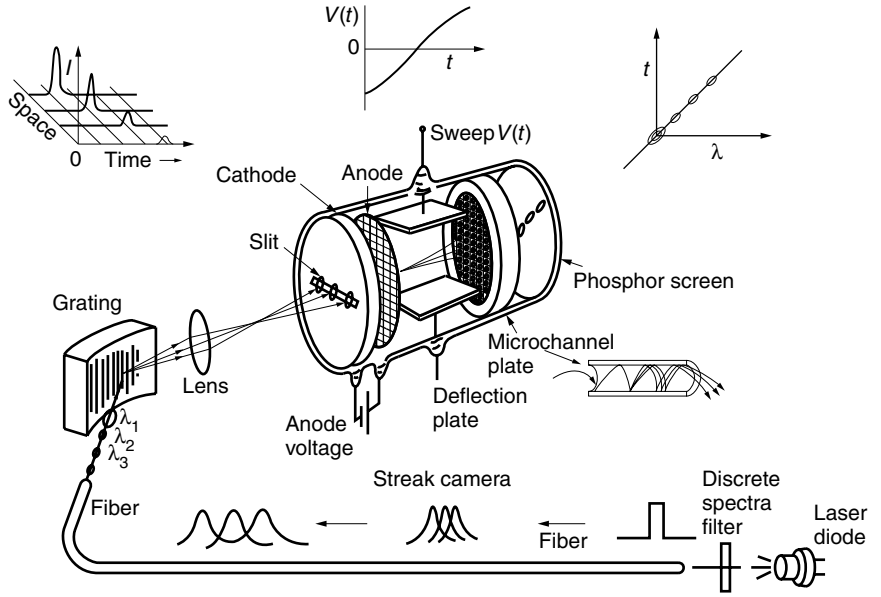
$$V_{\text{out}} = RPR_L = (5 \times 10^4)(10^{-12})(10^7) = 0.5 \text{ V}$$

It should, however, be remembered that such a large load resistance raises the value of the time constant CR_L of the detector, where C is the capacitance of the tube, and the response time of the tube becomes excessively slow.

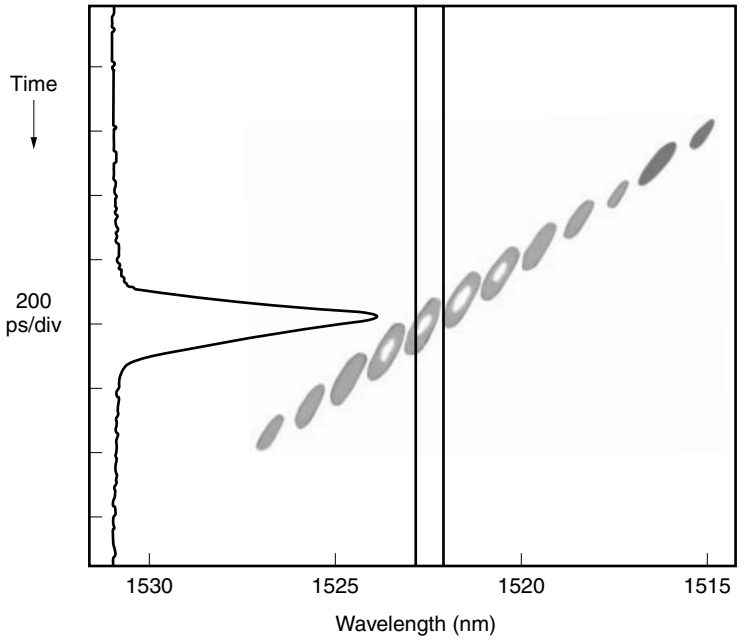
Despite the high sensitivity of the photomultiplier tube, it is not practical for use as a detector for optical communication systems. These systems operate in the $\lambda = 1.3$ or 1.55 μm regions where the photomultiplier sensitivity is low. Other disadvantages of the photomultiplier tube when used for optical communication systems are poor frequency response, large size, the necessity of a kilovolt power supply, the limited lifetime of the tube, and the cost.

12.2 STREAK CAMERA

A streak camera is used to observe light variations with respect to time of the order of picoseconds (10^{-12} s). The structure of the streak camera, which is shown in Fig. 12.2a,



(a)



(b)

Figure 12.2 Streak camera measuring (a) the dispersion during transmission in an optical fiber and (b) chromatic dispersion in single-mode optical fiber. (After Y. Tsuchiya [1].)

is quite similar to a cathode ray tube used for an oscilloscope. The major difference is that the cathode ray tube uses a steady electron beam supplied by an electron gun installed in the tube, while the streak camera uses the electrons that are generated by the bombardment of incident photons onto a low-work-function cathode and formulated as a beam by the anode.

Another feature is that even though the location of the electron gun of the cathode ray tube is fixed, the photocathode of the streak camera is sensitive along the horizontal slit, and the location of the incident light is another piece of information that the streak camera can supply. Thus, the streak camera provides simultaneously three kinds of information: the intensity of the light, the time of arrival, and the location of the incidence.

For higher speed operation the sawtooth deflector voltage is sometimes replaced by a high-frequency sinusoidal voltage.

The variation of the incident light intensity is observed as a function of time on the fluorescent screen. In order to enhance the electron density hitting the fluorescent screen, a microchannel plate is installed between the deflector plate and the fluorescent screen. The microchannel plate is made up of a honeycomb of fine dynode tubes. As the wall bounces the electron beam, the electron beam is intensified.

Figure 12.2a shows an example of using the streak camera to measure the dispersion phenomenon in an optical fiber. A pulse having different wavelengths is launched simultaneously into a long optical fiber. The difference in arrival time of the different wavelength components is measured by the streak camera. The light components arriving at a later time see a higher deflector voltage $V(t)$ and are deflected higher. A display such as the one shown in Fig. 12.2b is obtained [1].

12.3 MISCELLANEOUS TYPES OF LIGHT DETECTORS

Other types of light detectors [2–4] are summarized here. When light illuminates such materials as CdS or CdSe, the surface conductivity increases due to the photoconductive effect. The device based on this phenomenon is known as the *photoconductor cell* or *photocell*. A special feature of the CdS photocell is that its spectral response characteristics resemble that of the human eye. It is widely used as an exposure meter in photography.

Solar cells convert solar energy into electrical energy by a phenomenon known as the *photovoltaic effect*. Calibrated solar cells are also used as light meters for photography.

The phototransistor is the amalgamation of a photodiode, which will be explained extensively in the next section, and an amplifier transistor. It is like a photomultiplier tube made of semiconductor material. It can provide a large signal output, but the response time is not as high as the photodiode alone.

Calorimetric methods use the temperature rise in a heat-absorbing strip that is exposed to light. The temperature rise is measured by a thermometer with as low a heat capacity as possible. These thermometers include a thermocouple, which establishes the voltage due to the temperature difference between the ambient and the heat-absorbing strip, and a thermister or bolometer, which changes its resistance in accordance with its temperature.

As an example of a device based on the calorimetric method, Fig. 12.3 shows a diagram of a pyrometer that measures the temperature distribution inside a furnance

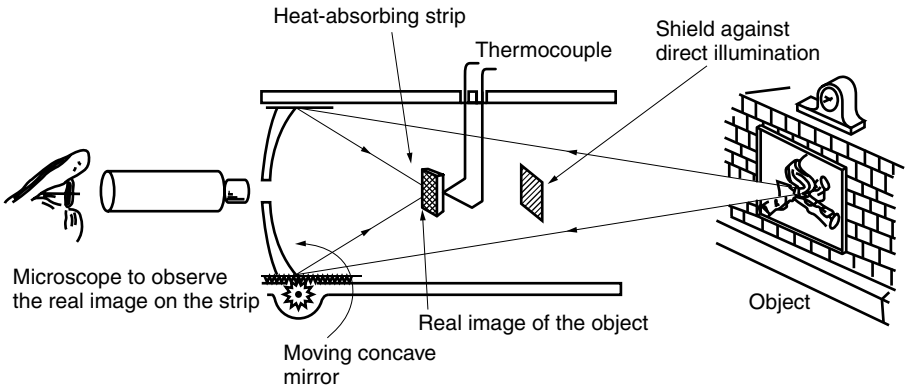


Figure 12.3 Pyrometer structure.

by measuring the radiating light intensity. The position of the concave mirror is adjustable so that the real image of the desired point of the radiation is focused onto the heat-absorbing strip. The attached microscope is used to observe the real image. The temperature rise of the strip is measured by a thermocouple. The measured intensity of light correlates with the temperature of the desired point in the furnace.

None of the devices in this section are suitable for fiber-optic communication because of the relatively slow response time. Detectors for fiber-optic communication are presented in the next section.

12.4 PIN PHOTODIODE AND APD

The basic principles of the PIN photodiode (PIN diode) and avalanche photodiode (APD) are quite similar to each other. Once the fundamental principles have been established for the PIN photodiode, the transition to understanding the avalanche photodiode will be quite easy.

Figure 12.4 shows the basic structure of the PIN and APD photodetectors. Basically, a PIN photodiode consists of three layers of semiconductor material: the p-type, the intrinsic-type, and the n-type semiconductors.

The actual physical structures of the PN and PIN photodetectors and the APD are presented in the next section; the PN and PIN photodetectors and the APD on an atomic scale are described in Appendix A of Volume II.

12.4.1 Physical Structures of PIN and APD Photodetectors

Figure 12.5a shows typical physical structures for the PIN and PN photodetectors [4,5]. There is very little difference between PN and PIN diodes. A PN diode is a diode formed by joining a p layer directly to an n layer, whereas a PIN diode is formed by joining a p layer to an intrinsic layer and then to an n layer. The p^+-type window layer is made extremely thin ($\sim$0.3 μm) because the light is incident from this layer and any charge carriers (electrons and holes) produced in the p^+ layer are subjected to a low electric field because of the high conductivity of the heavily doped p^+ layer. The lower electric field means slower response time and adds to the noise.

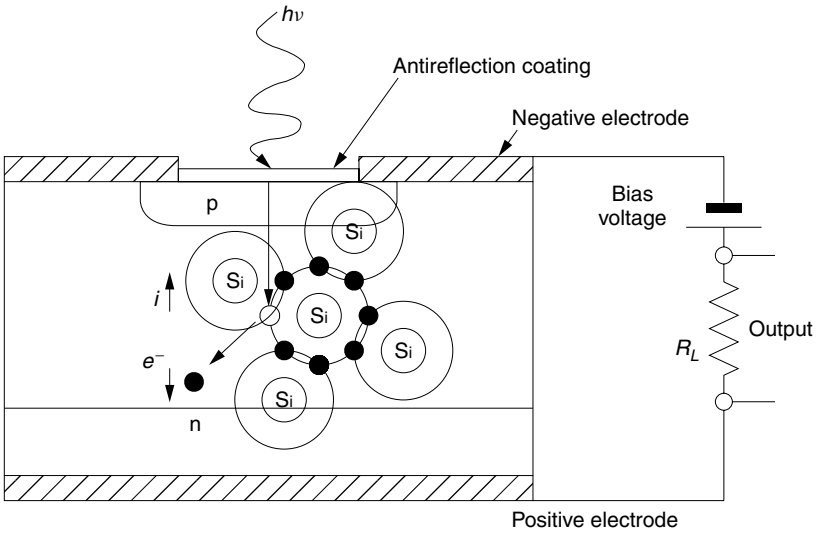


Figure 12.4 Basic structures of PN, PIN, and APD photodetectors.

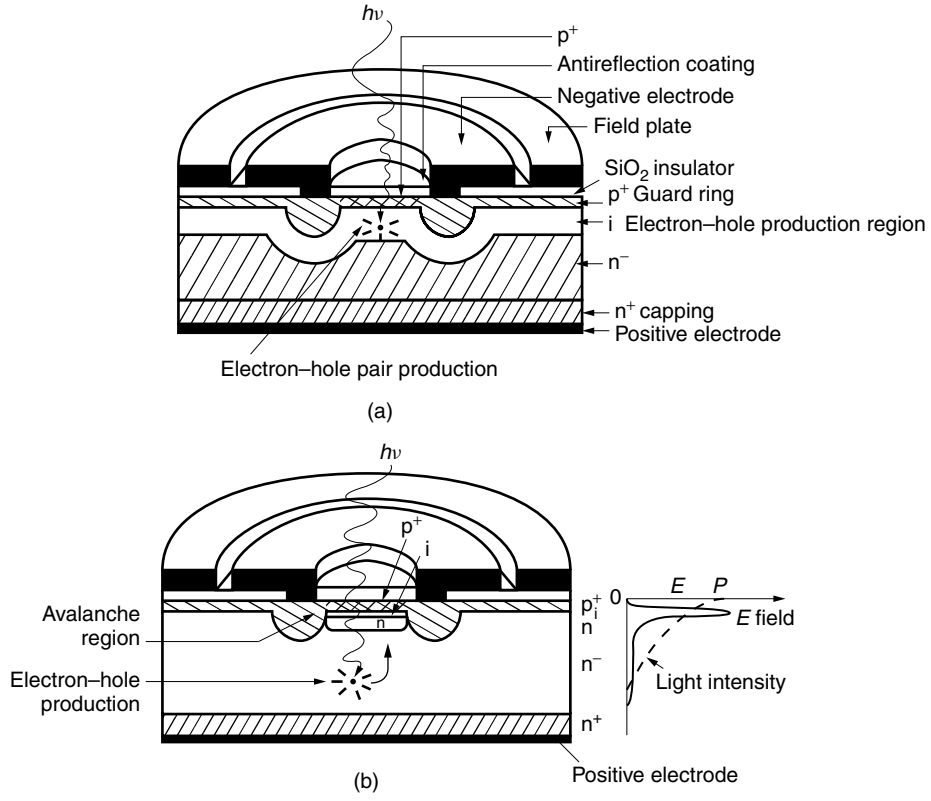


Figure 12.5 Structure of photodetectors. (a) PN and PIN photodiodes. (b) Avalanche photodiode (APD).

The light aperture is usually circular (Fig. 12.5 shows the cross section in a plane perpendicular to that of the circular aperture). A circular electrode is deposited over an insulating layer of SiO_2 and the p^+ layer guard ring. The purpose of the guard ring is to prevent dielectric breakdown by spreading the electric lines of force from the edge of the electrode. An antireflection (AR) coating is deposited over the aperture to minimize the amount of incident light that is reflected. Compared to a PN diode, the PIN diode has a thicker (a few μm thick) intrinsic layer that increases the interaction region for the incident photons with orbital electrons. An excessively long intrinsic region, however, harms the response time because the transit time for the electrons to reach the electrodes becomes too long. The diameter of the electrode is made small to minimize the capacitance. A metal field plate is deposited in order to prevent stray light entering the diode.

On the bottom side of the substrate, a highly doped n^+ capping layer is deposited to ensure better adhesion to the electrode in order to minimize the series resistance.

Figure 12.5b shows the structure of the avalanche photodiode [4] (APD). An APD has additional regions for accelerating the drift velocity of the carriers and multiplying the number of carriers by the avalanche effect. The bottom structure of the APD is identical with that of the PIN photodiode. The top p-i-n layers are multiplier layers. The resistivity of the i-type layer is made much higher than the other layers so that the potential drop across all the layers in series is concentrated in this layer.

Incident light is first converted into electric charge carriers in the n^- region. The generated carriers drift upward into the n-type layer because of the moderate electric field. The carriers then enter the highly resistive i-type layer. This i-type layer is not a separately deposited layer. Rather, it is the depletion region formed when the p^+ layer is joined to the n layer. Because of the combination of high resistivity and thinness, the field intensity in the i layer is so large that the collisions of the accelerated electrons with the crystal lattice result in the creation of new carriers. The electric field intensity reaches as high as 10^5 V/cm or even higher (the breakdown voltage of air is 3×10^4 V/cm), which is necessary to start the avalanche effect. The new carriers are also accelerated and collide with the crystal lattice in the avalanche region and generate even more carriers. The density of carriers multiplies as this snowball effect is triggered. Finally, the charge carriers exit from the top electrodes.

Thus, the bottom half of the APD has the function of converting the incident photons into electrons, and the top half has the function of amplifying the number of electrons generated. The APD performs the dual functions of photoelectric conversion and electric amplification.

12.4.2 Responsivity of the PIN Photodiode and APD

Expressions will be derived for the output current from a PIN photodiode and APD due to the incident light. The expressions are quite similar to those of a photomultiplier. Let W be the energy of the light entering the detector. The number of incident photons expressed in terms of the incident light energy is $W/h\nu$. Let η be the quantum efficiency, which is defined as the ratio of the number of generated electrons to incident photons. Each electron carries a charge e and the total amount of charge Q after multiplication due to the avalanche effect is

$$Q = \eta \frac{e}{h\nu} MW \tag{12.7}$$

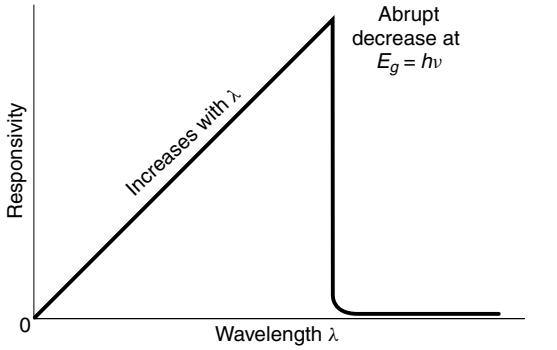
M is the multiplication factor due to the avalanche effect. The derivative of Eq. (12.7) with respect to time gives the output current from the diode

$$i_s = \eta \frac{e}{h\nu} M P_s \quad \text{or} \quad i_s = \eta \frac{e\lambda}{hc} M P_s \tag{12.8}$$

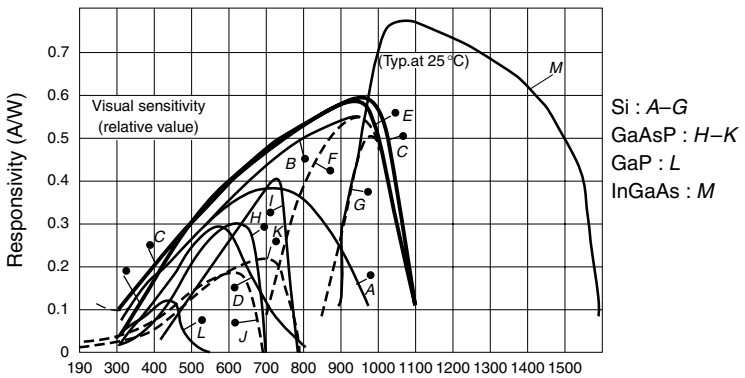
where P_s is the incident light power.

It should be noted that the above definition of η includes (1) the finite transmission of light into the surface with transmission coefficient Eq. (2.16), (2) the exponential decay of the light power inside the intrinsic region, and (3) the finite length of the intrinsic region for generating electron-hole pair production over and above the probabilistic nature of electron-hole pair production.

Equation (12.8) indicates that the output current increases linearly with the wavelength of the incident light. There is, however, an upper limit on the wavelength. For the photon to ionize the atoms in the crystal lattice, the energy quantum $h\nu$ of the incident photon has to be larger than the bandgap energy E_g of the semiconductor. The quantum efficiency η has a rather abrupt cutoff wavelength in the longer wavelength region, as shown in Fig. 12.6a. In practice, the quantum efficiency decreases faster



(a)



(b)

Figure 12.6 Spectral response of the PIN photodiode and APD. (a) Generic spectral response of the PIN photodiode and APD. (b) Spectral response of photodiodes. (After Hamamatsu Photonics catalog).

than linearly at shorter wavelengths due to scattering and deviation from the center frequency of the designed antireflection coating. Figure 12.6b shows the sensitivity of various semiconductor detectors. The wavelength dependence of η varies with the semiconductor material. The wavelength regions for optimum sensitivity are:

Ge	0.8–1.6 μm
Si	0.5–0.9 μm
InGaAs	1.0–1.6 μm

As with the photomultiplier, the responsivity is used to characterize a photodiode.

The output current with unit incident light power is the responsivity R, and R is expressed as

$$R = \eta \frac{e}{hv} M \tag{12.9}$$

or

$$i = RP \tag{12.10}$$

The responsivity of most PIN diodes is 0.5 A/W and a good rule of thumb is that every 1 watt of incident light power generates $\frac{1}{2}$ ampere of output current.

The reasons why the PIN photodiode and APD are exclusively used for fiber-optic communications are that both have (1) high photoelectric conversion efficiency, (2) fast response time, (3) high reliability, (4) compactness as well as ruggedness, (5) longevity, (6) simple electronics, (7) moderate cost, and (8) high linearity of output current over a wide range of input light power.

The rest of the chapter concentrates on how to incorporate these photodiodes in fiber-optic communication systems that are broadly categorized as direct detection systems and coherent detection systems.

12.5 DIRECT DETECTION SYSTEMS

Figure 12.7 shows a very basic circuit employing a PIN photodiode for direct detection in a fiber-optic communication system. The photodiode is back-biased. The p layer is negative, and the n layer is positive. The bias voltage is in the range of 7–15 volts. The capacitor C prevents the signal from going through the power supply. This particular circuit is connected directly to a transistor amplifier, but a more sophisticated coupling circuit is required if impedance matching, frequency response, and signal to noise ratio are all to be considered. The details for these more sophisticated circuits are presented in Chapter 16.

Let the instantaneous E field of the light incident to the photodiode be

$$E_s = S(t)\cos(\omega_c t + \theta) \tag{12.11}$$

where $S(t)$ is the modulated amplitude, ω_c is the carrier frequency of the light, and θ is either the modulated or unmodulated phase angle of the carrier. In order to calculate the detected signal current i_s using Eq. (12.8), the input power P_s has to be calculated.

The peak energy per unit time crossing a unit area is, by the Poynting vector theorem,

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \tag{12.12}$$

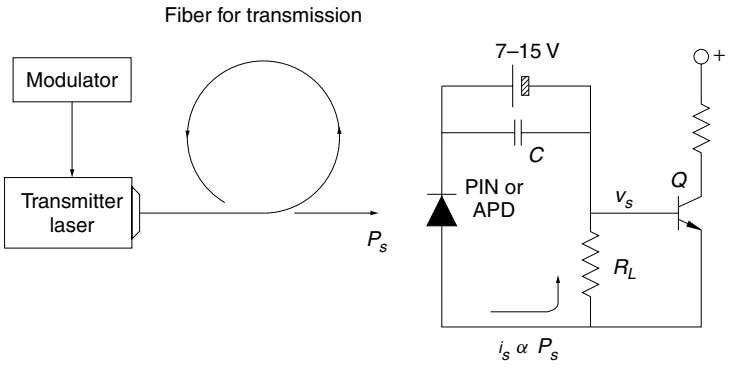


Figure 12.7 Direct detection system.

The magnitude of **E** and **H** have the relationship

$$E = \eta_0 H$$

where η_0 is the intrinsic impedance. The average energy per unit time crossing area s is

$$\begin{aligned} P_s &= \frac{s}{\eta_0}\langle E_s^2 \rangle \\ &= \frac{s}{\eta_0} S^2(t)\langle \cos^2 \omega_c t \rangle \\ &= \frac{s}{2\eta_0} S^2(t) \end{aligned} \tag{12.13}$$

where $\langle \quad \rangle$ is the time average over a length of time containing a large number of light carrier periods but not so large a number in modulation periods. The detected signal current from the avalanche photodiode can be expressed in various ways:

$$\begin{aligned} i_s &= \frac{\eta e}{h\nu} M P_s \\ &= \frac{\eta e}{h\nu} M \frac{s}{2\eta_0} S^2(t) \\ &= R \frac{s}{2\eta_0} S^2(t) \\ &= K S^2(t) \end{aligned} \tag{12.14}$$

where

$$K = \frac{e\eta}{h\nu} M \frac{s}{2\eta_0}$$

Expressions in terms of the light amplitude $S(t)$ rather than the power P_s are more convenient for explaining mixer operations. The amplitude constant K accounts for both responsivity and sensor area and directly relates the amplitude square to the output current from the diode.

Note that i_s is proportional to the light intensity and does not contain information about either the frequency ω_c of the light carrier or the phase ϕ. Thus, direct detection cannot be used for detecting frequency modulated or phase modulated signals.

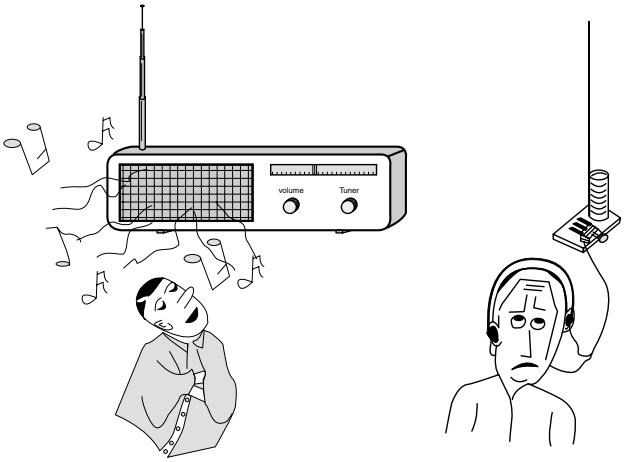
The mean square value of the direct detection signal is

$$\langle i_s^2 \rangle = \left(\frac{\eta e}{h\nu} M\right)^2 P_s^2 \tag{12.15}$$

Next, coherent detection systems that can be used not only with intensity modulated light but also with phase and frequency modulated light are presented.

12.6 COHERENT DETECTION SYSTEMS

Four more schemes besides the direct detection scheme will be explained here. These four are coherent detection schemes [6,7] and are so named because they require a local oscillator laser whose output light is either coherent or quasicoherent with the signal light. Making a comparison with a radio receiver set, direct detection is analogous to a crystal radio. The crystal radio is simple, but the sensitivity and station selectivity are poor. Direct detection is primarily used for an on–off type of modulation. Coherent optical detection is analogous to a superheterodyne radio. The superheterodyne radio is much more sophisticated, but its sensitivity and selectivity are far better than the crystal radio.



12.6.1 Heterodyne Detection

Optical heterodyne detection [8,9] uses the beat signal between the signal light and the local oscillator (LO) light.

Figure 12.8 shows a heterodyne detection circuit. The signal light from the transmitter laser passes through the fiber-optic cable transmission line and prior to being combined with the local oscillator laser light, the signal light passes through a polarization controller. The polarization controller adjusts the state of polarization of the signal light so that it is linearly polarized in the desired direction with respect to the local oscillator laser light (parallel in the case of a single mixer). The frequency difference between the local oscillator laser and the signal light has to be maintained at a constant

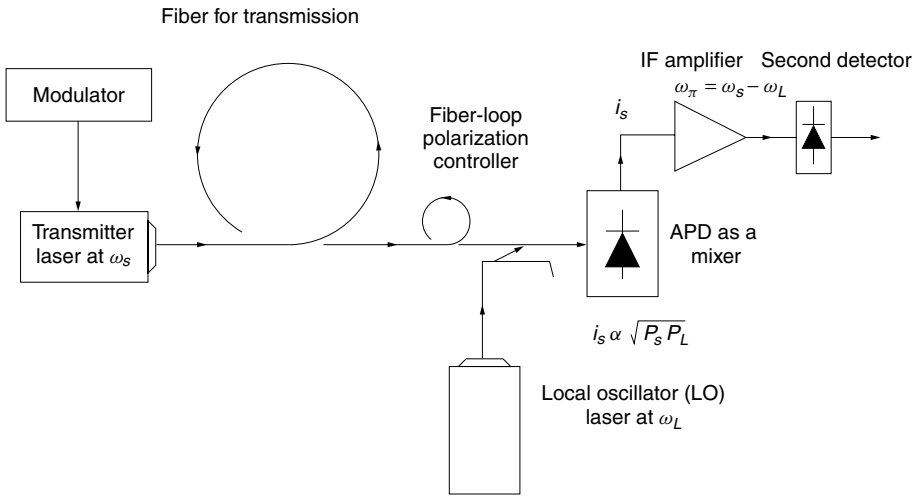


Figure 12.8 Heterodyne detection.

value. The output from the mixer is amplified by an IF amplifier and then detected a second time to recover the desired information from the IF signal.

The field at the mixer surface is the sum of the received signal field and the local oscillator laser field. Let the incident light be the same as the case of direct detection given by Eq. (12.11) and let the local oscillator laser light be

$$E_L = E_{\text{LO}} \cos(\omega_L t + \theta_L) \tag{12.16}$$

Both the signal and local laser light illuminate the APD diode used as a mixer diode. The resultant incident light power is obtained from the top equation of Eq. (12.13) replacing E_s by $E_L + E_s$.

$$P = \frac{s}{\eta_0} \langle [S(t) \cos(\omega_c t L) + E_{\text{LO}} \cos(\omega_L t + \theta_L)]^2 \rangle \tag{12.17}$$

Using Eq. (12.8) and the identity $\cos x \cos y = \frac{1}{2}[\cos(x - y) + \cos(x + y)]$, the output current from the mixer is

$$i_{\text{IF}} = \frac{\eta e}{h\nu} M \{P_s + P_L + 2\sqrt{P_s P_L} \cos[\omega_{\text{IF}} t + (\theta_c - \theta_L)]\} \tag{12.18}$$

where

$$\omega_{\text{IF}} = \omega_c - \omega_L, \qquad P_s = \frac{s}{2\eta_0} S^2(t), \qquad P_L = \frac{s}{2\eta_0} E_{\text{LO}}^2 \tag{12.19}$$

and where ω_L and ω_{IF} are the local oscillator and IF frequencies, respectively. The terms with $(\omega_c + \omega_L)$ have been ignored. The phases of the signal and local oscillator light are θ_c and θ_L, respectively. Only the third term of Eq. (12.18) contains the signal at the intermediate frequency ω_{IF}. The mean square value of the IF signal is, from Eq. (12.18),

$$\langle i_{\text{IF}}^2 \rangle = 2 \left(\frac{\eta e}{h\nu} M \right)^2 P_s P_L \tag{12.20}$$

The mean square value of the IF signal is the product of the signal and local oscillator powers, while that of the direct detection system as given by Eq. (12.15) is the square of the signal power, which for weak signals is the square of an already small quantity. Thus, the output from the heterodyne detection for the small signal case is significantly higher than that of direct detection. The selectivity of the heterodyne system is also higher than that of direct detection because the bandwidth of the heterodyne is determined by the bandwidth of the IF amplifier, which is a lower frequency and can be tailored to the required bandwidth of the signal more easily, typically in the gigahertz range. This not only improves the selectivity but enhances the rejection of the noise power. Direct detection, on the other hand, has to rely on an optical bandpass filter (whose bandwidth is normally much wider than gigahertz) for wavelength selectivity and noise power rejection.

12.6.2 Homodyne Detection

When the frequency of the local oscillator of the heterodyne detection system [10,11] is made equal to the frequency of the signal light, the system becomes homodyne. The homodyne system not only relaxes the frequency bandwidth requirement of the IF amplifier but also improves the sensitivity by 3 dB over heterodyne detection. The phase stability requirement of the local oscillator, however, is severe and an elaborate phase stabilization circuit is required for the local oscillator laser.

With

$$\omega_c = \omega_L \quad \text{or} \quad \omega_{\text{IF}} = 0$$

Eq. (12.18) becomes

$$i_m = \frac{\eta e}{h\nu} M[P_s + P_L + 2\sqrt{P_s P_L}\cos(\theta_c - \theta_L)] \tag{12.21}$$

When P_L is large in comparison to P_s, the third term is the signal component. The phase-lock loop assures $\theta_c = \theta_L$ for the steady signal. The mean square value of the desired signal is

$$\langle i_m^2 \rangle = 4\left(\frac{\eta e}{h\nu} M\right)^2 P_s P_L \tag{12.22}$$

Compare this result with the heterodyne result in Eq. (12.20). There is a factor of 2 difference. This difference comes from averaging $\langle \cos^2[\omega_{\text{IF}} t + (\theta_c - \theta_L)] \rangle$ in Eq. (12.18), which is $\frac{1}{2}$ and averaging $\langle \cos^2(\theta_c - \theta_L) \rangle$ in Eq. (12.21), which is unity for $\theta_c = \theta_L$. Thus, the signal to noise ratio and hence the sensitivity of the homodyne system is better than the heterodyne system by 3 dB.

A disadvantage of homodyne detection is the stringent requirement on the phase stability of both the transmitting and local oscillator lasers. In Eq. (12.21), if $(\theta_c - \theta_L)$ fluctuates between 0° and 90°, the output fluctuates between its maximum value and zero, and communication becomes impossible. A phase-lock loop (PLL) can extract the carrier frequency component from the amplitude modulated carrier. The extracted carrier is used to lock the oscillation of the local oscillator signal.

Figure 12.9 shows a block diagram of the Costas PLL [12]. Inside the 90° optical hybrid, the local oscillator signal is divided into two components, one of which is phase

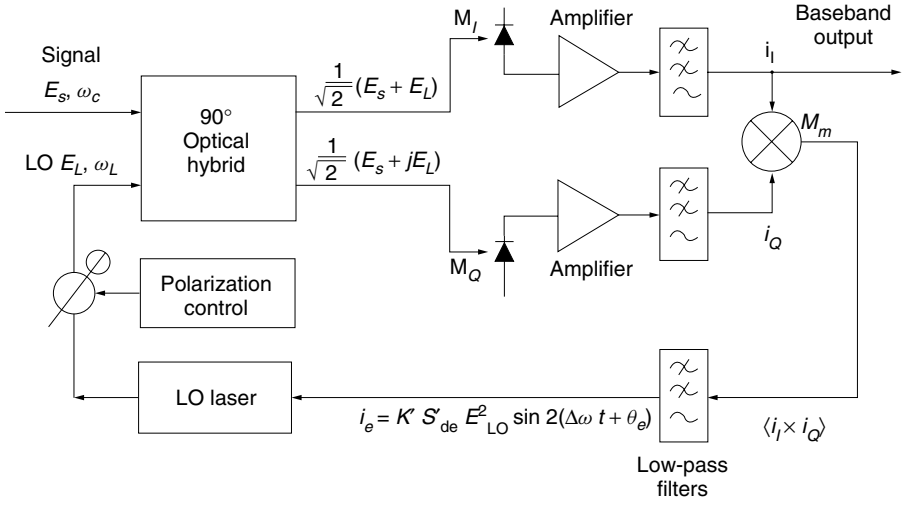


Figure 12.9 Homodyne detection with a phase-lock loop.

shifted by 90°. The component with the same phase as the original local oscillator signal and that with a 90° phase shift are expressed as

$$\begin{aligned} E_{L_I} &= \frac{1}{\sqrt{2}} E_{\mathrm{LO}} \cos(\omega_L t + \theta_L) \\ E_{L_Q} &= \frac{1}{\sqrt{2}} E_{\mathrm{LO}} \sin(\omega_L t + \theta_L) \end{aligned} \tag{12.23}$$

The local oscillator power P_L is assumed to be equally divided into the two components. The subscripts I and Q stand for in-phase and quadrature, respectively. The two components of the local oscillator are separately mixed with the received signal at mixers M_I and M_Q. When E_{L_I} is fed into M_I with the received signal, the output from M_I is similar to Eq. (12.18). The IF component of the output current is i_I. Similarly, the output from M_Q is i_Q. The $\Delta\omega$ components of the outputs from M_I and M_Q are called the in-phase and quadrature components of the IF signal. They are given by

$$\begin{aligned} i_I &= \sqrt{2} K S(t) E_{\mathrm{LO}} \cos(\Delta\omega\, t + \theta_e) \\ i_Q &= \sqrt{2} K S(t) E_{\mathrm{LO}} \sin(\Delta\omega\, t + \theta_e) \end{aligned} \tag{12.24}$$

where

$$\Delta\omega = \omega_c - \omega_L, \qquad \theta_e = \theta_c - \theta_L$$

and K is given by Eq. (12.14). Now, both i_I and i_Q are fed into another mixer M_m. The relevant component of the output from the mixer is

$$i_e = \tfrac{1}{2} K' 2[S(t) E_{\mathrm{LO}}]^2 \sin 2(\Delta\omega\, t + \theta_e) \tag{12.25}$$

where K' is a constant associated with the electronic mixer. $[S(t)]^2$ can be rewritten as

$$[S(t)]^2 = S_{\text{dc}} + S_{\text{ac}}\ (t) \tag{12.26}$$

where S_{dc} is the average component and $S_{\text{ac}}\ (t)$ is the ac component of $[S(t)]^2$. The frequency spectrum of $S_{\text{ac}}\ (t)$ is associated with the modulation by the signal and is normally much higher (tens of gigahertz) than $\Delta\omega$, provided ω_L is close to ω_c. With a low-pass filter, $S_{\text{ac}}\ (t)$ can be removed. The output of the low-pass filter becomes

$$i_e = K_1' E_{\text{LO}}^2 S_{\text{dc}} \sin 2(\Delta\omega\ t + \theta_e) \tag{12.27}$$

The output i_e from the mixer is used as an error signal to control the frequency and phase of the local oscillator laser diode. To explain the operation, assume first that the frequency is already matched but the phase is not: $\Delta\omega = \omega_c - \omega_L = 0$ and $\theta_e = \theta_c - \theta_L \neq 0$. Figure 12.10 shows a plot of the error signal i_e as a function of the phase mismatch θ_e.

Consider only the region of $-\pi/4 < \theta_e < \pi/4$. Let us say the operating point is at P in Fig. 12.10. A positive value of the error signal i_e means $\theta_e > 0$ or $\theta_c > \theta_L$. A positive error signal i_e raises the local oscillator frequency higher so that θ_L is increased and reaches the value of θ_c, thereby bringing the operating point P to the origin.

Next, suppose that ω_c makes an abrupt increase $\Delta\omega$ at $t = 0$. In this case, $\theta_e = \Delta\omega\, t$ starts to increase, and hence i_e starts to increase, raising the local oscillator frequency ω_L in order to catch up with the increase in ω_c and bring the operating point back to the origin. It is, however, crucial that the turnaround time of this feedback process be fast enough to react before θ_e goes beyond the range of $\pm\pi/4$ radians.

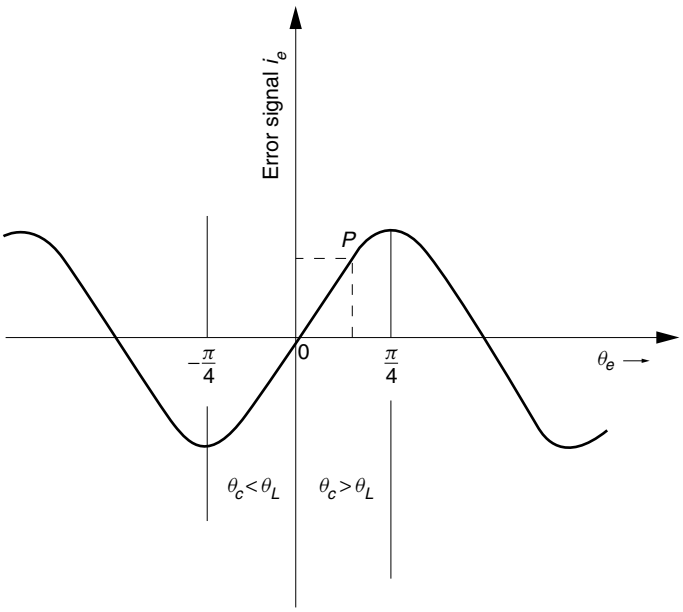


Figure 12.10 Error signal $\langle i_I i_Q \rangle_{\text{error}}$ as a function of phase error $\theta_e = \theta_c - \theta_L$.

In the case of ASK ("on–off") modulation, either the in-phase or the quadrature component alone of Eq. (12.24) could have been used directly as an error signal to the feedback loop. For phase modulation, where $S(t)$ takes on positive and negative values (or 0° and 180° phases), the sign of the error signal changes as $S(t)$ changes sign, and both in-phase and quadrature components had to be used as the error signal.

In this way, the phase as well as the frequency of the local oscillator is locked to the received carrier frequency. In other words, the local oscillator is completely coherent with the received signal, and in traditional communication theory, only homodyne is called coherent detection. In the field of optics, both heterodyne and homodyne are generally called coherent detection.

12.6.3 Intradyne System

The intradyne system [10] is a hybrid of the heterodyne and homodyne systems. It enjoys the merits of both systems. The stringent requirements on the frequency bandwidth and phase stability are relaxed. The intradyne system normally incorporates the phase diversity technique in which the signal is first divided into more than one channel before detection. Signal processing circuits, such as squaring and differentiating circuits, are installed in each channel. After being processed, the signals from the divided channels are added to render a final signal that is free from phase jitter.

In the homodyne system shown in Fig. 12.9, if the local oscillator frequency is not completely matched, both the in-phase, i_I, and quadrature, i_Q, components become modulated at a frequency $(\omega_c - \omega_L)$ as shown in Figs. 12.11c and 12.11d. Such outputs can almost be used as the final signal, but modulation of the amplitude at frequency $\Delta\omega = (\omega_c - \omega_L)$ creates a problem.

The intradyne system shown in Fig. 12.12 uses both in-phase and quadrature components to remove the modulation of the amplitude. The sum of the squares of the in-phase component and the quadrature component, Eq. (12.24), gives the output

$$i^2 = 2[KE_{\mathrm{LO}}S(t)]^2 \tag{12.28}$$

i^2 is free from $(\omega_L - \omega_c)$ and ϕ. The quadrature current i_Q is fed to the frequency discriminator, which keeps $(\omega_L - \omega_c)$ within a given range. Such an intradyne detector is sometimes called a phase diversity receiver because two phase components are used. It should be noted that the output being $2K^2E_{\mathrm{LO}}^2S^2(t)$, this detection system is not suitable for either phase or frequency modulated signals.

It is also possible to split the signal into N components with equal phase spacing between components. But, the larger N is, the higher the local oscillator power is needed. The loss in the circuit is also increased. An advantage of a larger N, however, is that uniformity among the channels is less critical.

As already mentioned, the frequency of the input light is converted into either an IF or baseband signal by mixing the incoming signal with the local oscillator light. The quality of the mixer significantly influences the sensitivity of the heterodyne system because the mixer is the first element that the signal at its lowest power level sees. Because of this importance, the function of the mixer will be explained in detail here.

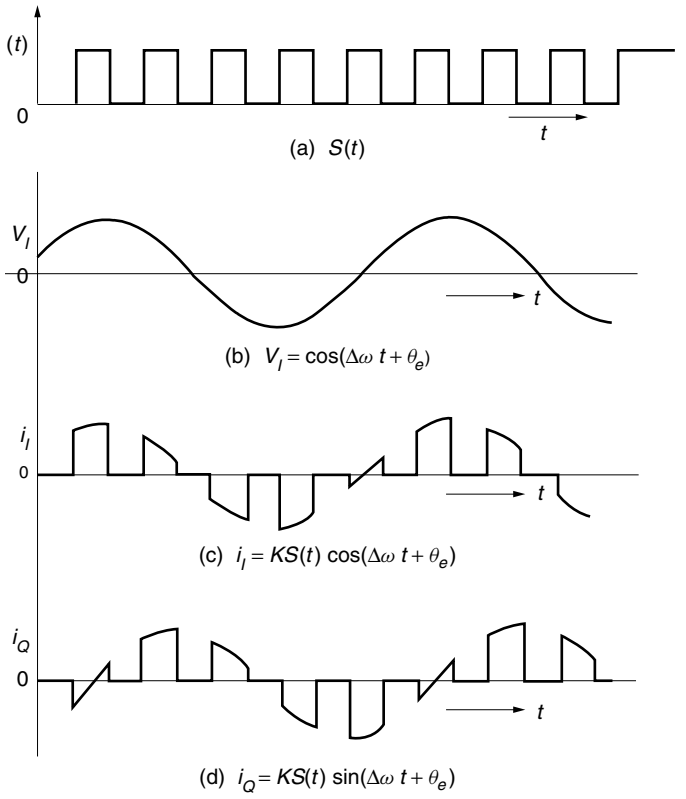


Figure 12.11 Intradyne signal.

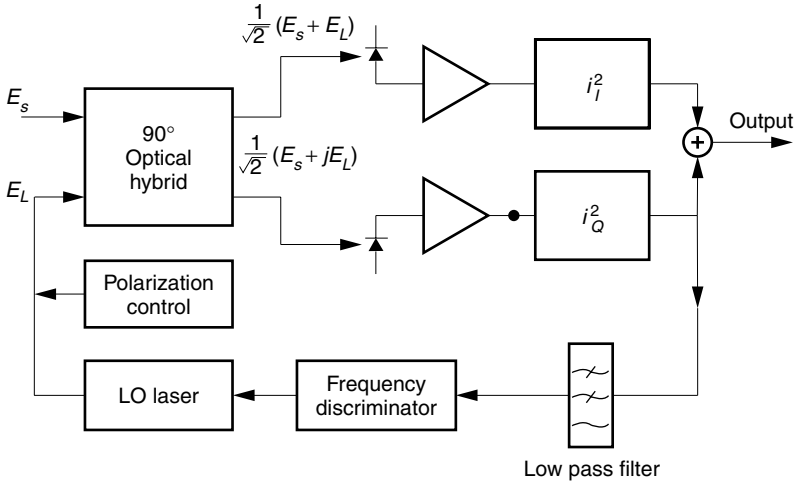


Figure 12.12 Intradyne detection.

12.7 BALANCED MIXER

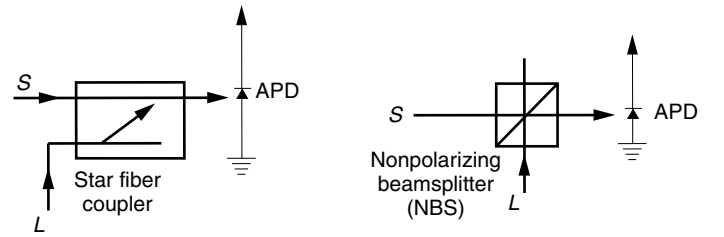
There are two types of mixers: the single crystal mixer as shown in Fig. 12.13a, and the double balanced mixer [13] as shown in Fig. 12.13b.

The most commonly used input elements to the mixer are optical fiber couplers or prism couplers. The substrate optical coupler shown in (1) of Fig. 12.13b will be treated. Assume that inside the substrate optical coupler, the field that crosses the guide from the upper to lower or from the lower to upper guide experiences a 90° phase shift compared to the field that goes straight through. The input fields E_1 to mixer M_1 and E_2 to mixer M_2 will be calculated first. The 90° phase shift will be accounted for by changing the cosine function into a sine function. The input fields to the mixers are expressed as

$$E_1 = \frac{1}{\sqrt{2}}[S(t)\cos\omega_c t + E_{\mathrm{LO}}\sin\omega_L t] \tag{12.29}$$

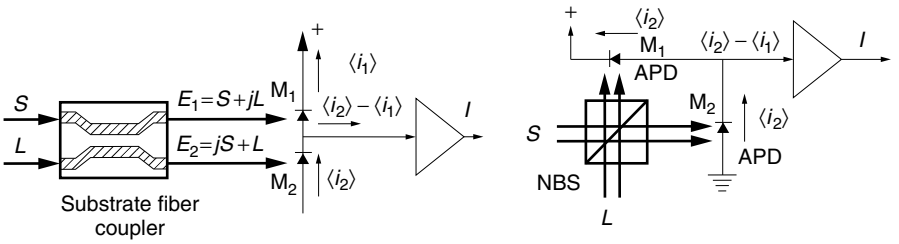
$$E_2 = \frac{1}{\sqrt{2}}[S(t)\sin\omega_c t + E_{\mathrm{LO}}\cos\omega_L t] \tag{12.30}$$

where $S(t)$ and E_{LO} are the signal and local oscillator fields, respectively. The factor $1/\sqrt{2}$ is necessary because the light power is split 50–50.



(1) Star fiber coupling (2) Nonpolarizing beamsplitter coupling

(a)



(1) Fiber coupling (2) Prism coupling

(b)

Figure 12.13 Comparison between (a) the single crystal mixer and (b) the double balanced mixer.

The output currents i_1 and i_2 from mixer M_1 and mixer M_2 are

$$\langle i_1 \rangle = K_1 \left[\tfrac{1}{2} S^2(t) - S(t) E_{\mathrm{LO}} \sin(\omega_c - \omega_L)t + \tfrac{1}{2} E_{\mathrm{LO}}^2 \right] \tag{12.31}$$

$$\langle i_2 \rangle = K_2 \left[\tfrac{1}{2} S^2(t) + S(t) E_{\mathrm{LO}} \sin(\omega_c - \omega_L)t + \tfrac{1}{2} E_{\mathrm{LO}}^2 \right] \tag{12.32}$$

Equation (12.31) or (12.32) alone is the expression for the output current from the single crystal detection. If the intensity of the local oscillator laser fluctuates, the mixer output fluctuates accordingly, and the fluctuations in E_{LO}^2 become a source of noise.

In a double balanced mixer, however, i_1 is subtracted from i_2. This means both first and third terms in Eqs. (12.31) and (12.32) cancel each other,* provided the quantum efficiencies of the two mixers are well balanced and

$$\eta_1 = \eta_2 = \eta \quad \text{and hence} \quad K_1 = K_2 = K \tag{12.33}$$

Assuming Eq. (12.33) is valid, the output from the double balanced mixer is

$$i = 2KS(t)E_{\mathrm{LO}} \sin \omega_{\mathrm{IF}} t \tag{12.34}$$

The signal current in Eq. (12.34) is indeed double that of the single crystal mixer given by Eq. (12.31) or (12.32).

Fluctuations in E_{LO} will cause i to fluctuate, but the amount is different for the two cases. Equation (12.31) for the single mixer contains a term proportional to E_{LO}^2, whereas Eq. (12.34) for the double mixer is proportional to $S(t)E_{\mathrm{LO}}$. Normally, $S(t)$ is significantly weaker than E_{LO}, so that the effect of local oscillator fluctuations is significantly improved over the case of the single crystal mixer.

Even though the double balanced mixer is more complicated in structure than the single crystal detector, it has such advantages as:

1. The output signal current is twice as much as that from the single crystal mixer.
2. The mixer output is practically free from the noise generated by the intensity fluctuations of the local oscillator laser.
3. Even when the output of the mixer is dc coupled to the preamplifier, saturation of the preamplifier can be avoided because of the absence of the dc current.

Next, a method of detection that resembles none of the above will be explained.

12.8 DETECTION BY STIMULATED EFFECTS

Another unique method of detection is the use of the stimulated effect in the optical fiber. It can be considered as a kind of homodyne detection whose local oscillator light is generated by stimulated Brillouin scattering (SBS) inside the optical fiber during transmission. Such a scheme does not require phase stabilization of the local oscillator because the signal induced by the stimulated Brillouin scattering is automatically in phase with the carrier of the signal. It also provides an additional gain of 15–25 dB during transmission. A drawback, however, is that stimulated Brillouin scattering cannot be effectively suppressed during the "off" period of the ASK modulation.

* The $S(t)E_{\mathrm{LO}} \sin(\omega_C + \omega_L)t$ terms were not included in Eqs. (12.31) and (12.32) because they are outside the spectral range of interest. Even if they had been included, these terms would cancel each other when i_1 is subtracted from i_2.

12.8.1 Stimulated Effects

Detection associated with the nonlinear effect of the glass will be explained [14–18]. When a light beam with a diameter of a few millimeters is focused into the core of an optical fiber with a diameter of a few micrometers, the light intensity is increased by 10^6 times. When one compares the length of an optical fiber with the lengths of other optical components, the fiber is typically 10^4 times longer. The power density-interaction length is used as a barometer of the enhancement of the nonlinear effect. An optical fiber is 10^{10} times more susceptible to nonlinear effects than a bulk optical component.

Close observation reveals that even when monochromatic incident light is used, the spectrum of the light reflected back from inside the fiber is no longer monochromatic. This phenomenon is quite different from the electrooptic effect that relates to the change in the index of refraction but does not involve a change in frequency spectrum.

As shown in Fig. 12.14, on both sides of the spectrum of the incident light, two new lines become observable as the intensity of the incident light is raised. Spectra appearing on the lower frequency side are called lines of Stokes radiation, and those appearing on the higher frequency side are called lines of anti-Stokes radiation. The ratio of the intensity I_S of Stokes radiation to the intensity I_a of anti-Stokes radiation obeys the Boltzmann distribution and

$$I_a/I_S = e^{-2h\Delta f/kT} \tag{12.35}$$

where Δf is the frequency shift of the Stokes wave. Thus, the intensities of the Stokes waves are larger than those of the anti-Stokes waves.

Both Stokes and anti-Stokes waves consist of two sets of spectral lines. One is due to Brillouin scattering and the other is due to Raman scattering. The mechanism

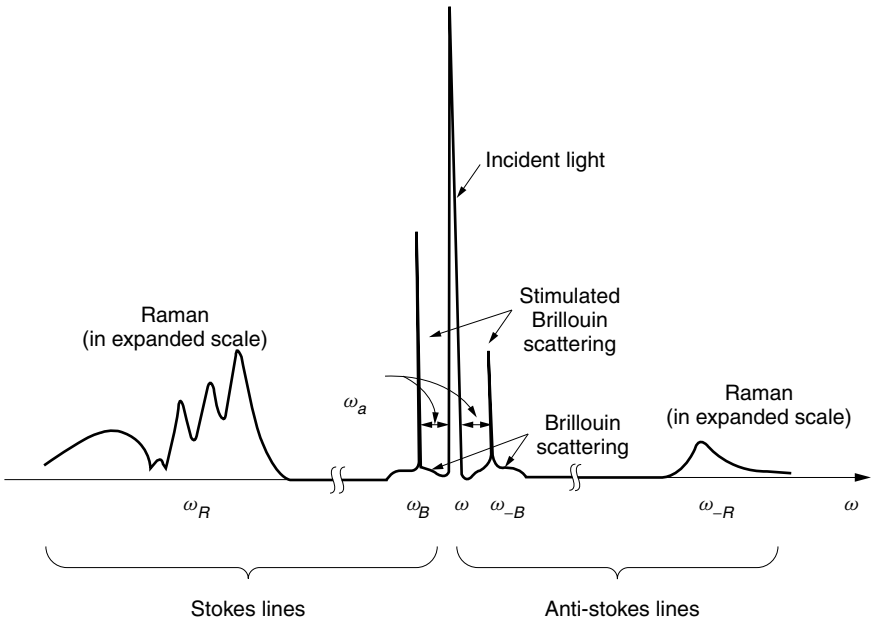


Figure 12.14 Typical Stokes and anti-Stokes lines.

of the frequency shift for Brillouin and Raman scatterings are different. While Raman scattering is associated with vibrations of dipole moments formed by two ions in a unit cell of the crystal lattice, Brillouin scattering is the reflection of light from acoustic waves (phonons) generated by the electrostrictive effect. The characteristics of the Brillouin scattering are different from that of the Raman scattering.

12.8.2 Homodyne Detection by Stimulated Brillouin Scattering

Figure 12.15 shows a block diagram of a homodyne receiver that uses the wave generated by stimulated Brillouin scattering (SBS). The pump laser light with a power of several milliwatts at frequency ω_L is injected from the receiver end into the optical fiber in the reverse direction from that of the signal flow. The signal light is much weaker in intensity compared to the pump. A beat wave is generated between the pump laser light at ω_L and the signal wave whose carrier frequency ω_s is set to be identical with the frequency ω_B of the Brillouin scattering. Two waves traveling in opposite directions but with slightly different frequencies generate a "moving" standing wave in the fiber in the direction of propagation of the higher carrier frequency light. This "moving" standing wave generates an acoustic wave due to the electrostrictive effect. The acoustic wave spatially modulates the index of refraction of the core. Consequently, a moving optical grating is formed in the core. If the Bragg condition is satisfied, the moving optical grating efficiently scatters the pump beam into the same direction as the signal, creating a second beam heading toward the receiver. The frequency of the second wave is shifted lower with respect to that of the pump beam due to the Doppler effect of the moving scatterer (SBS). This second beam has the same frequency as the signal light. The net result is an increase in the signal power.

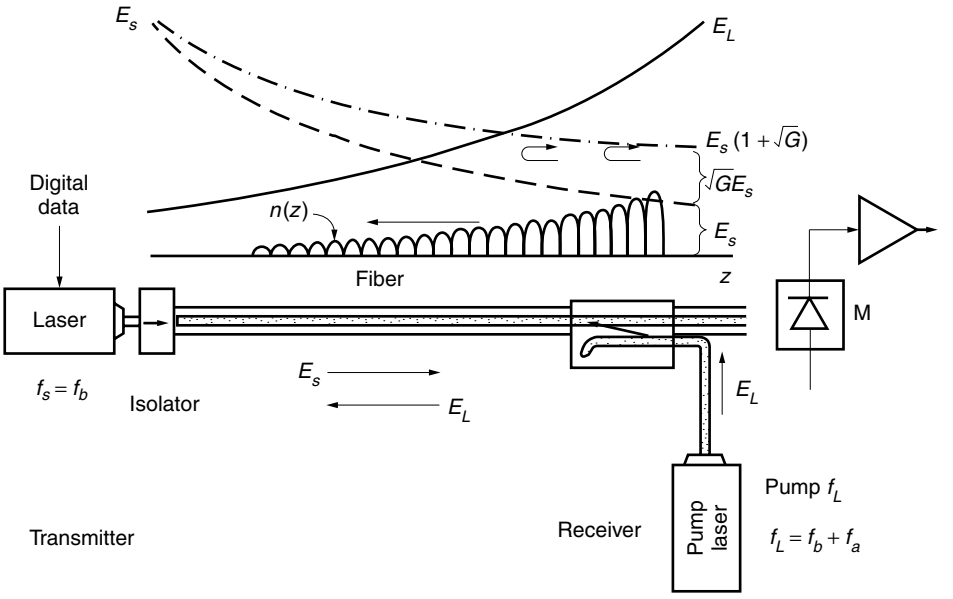


Figure 12.15 Homodyne receiver based on stimulated Brillouin scattering.

This amplification is immune to the phase jitter of the signal. If the phase of the signal is delayed by $\Delta\phi$ radians, the peak positions of the grating shift by $d = \frac{1}{2}\Delta\phi/\beta$ toward the source. The local oscillator signal reflected from the grating has to travel an extra distance $2d$, causing a phase delay of $2d\beta = \Delta\phi$. Thus, the reflected local oscillator signal is delayed by the same amount. Thus, the secondary beam has both the same frequency and the same phase. Two special features of this amplification are the automatic synchronization of the phase of the pump beam with that of the signal beam and the narrowness of the frequency bandwidth, typically 20 MHz. Thus, this phenomenon can play the role of the synchronous homodyne receiver.

This scheme not only detects the signal but also has a power gain of G that is proportional to the pump power and is typically 15–25 dB. The drawback is the difficulty in applying ASK modulation (on–off modulation) because the stimulated Brillouin scattering tends to continue even during the light off period of the on–off modulation.

Let us find the frequency relationships. The frequency f_B of Brillouin scattering that will beat with a frequency f_L of the pump light and generate the acoustic frequency f_a is

$$f_L - f_B = f_a \tag{12.36}$$

The grating has to satisfy the Bragg condition if the reflected wave is enhanced by Brillouin scattering. If the acoustic wave propagates at velocity V_a in the crystal, the

When two atoms in a unit crystal cell consist of positive and negative ions, these two atoms form an electric dipole. Such a dipole oscillates when an oscillating external field **E** is applied. The oscillating dipole in turn reradiates an electromagnetic wave. Such a dipole has a resonant frequency ω_v of oscillation that is determined, among other physical parameters, by the atomic weight. When the frequency ω_v of the external field coincides with this resonant frequency, the intensity of the reradiation is greatly enhanced. If, as a start, a small-intensity signal at ω_R in the thermal noise spectrum is mixed with the frequency ω_L of the external field in the nonlinear medium, beat waves are generated at the sum frequency $\omega_L + \omega_R$ and the difference frequency $\omega_L - \omega_R$. Only the frequency ω_R in the thermal noise spectrum that satisfies

$$\omega_L - \omega_R = \omega_v$$

can effectively excite the dipole. The reradiated field at ω_v from the dipole again mixes with the incident wave at ω_L and contributes to the signal field at frequency ω_R. The cyclic reaction repeats and builds up the signal at ω_R. To summarize, the incident energy at ω_L is transferred into the Raman spectrum at ω_R using the ionic dipole and the nonlinear properties of the medium.

The amplification based on the stimulated Raman scattering (SRS) goes through the same procedure but the difference is that the spectrum ω_R is injected rather than that originating from the thermal noise spectrum. The intensity of the output at ω_R is larger than that of the input power due to the enhancement of the Raman scattering. Such an enhancement can be used for a Raman amplifier or a Raman oscillator [19,20].

The intensity of Stokes radiation due to Raman scattering in a SiO_2 fiber is about two orders of magnitude smaller than that of Brillouin scattering. The frequency shift of the spectrum due to Brillouin scattering is 11.4 GHz while that of Raman scattering is about a few hundred terahertz. The linewidth of Raman scattering is of the order of terahertz, while that of stimulated Brillouin scattering is of the order of tens of megahertz.

wavelength λ_a of the acoustic wave is $\lambda_a = V_a/f_a$. For the pump wave to satisfy the Bragg condition, λ_a has to be one-half of the wavelength of the pump light:

$$\frac{1}{2}\frac{c}{n}\cdot\frac{1}{f_L} = \frac{V_a}{f_a} \tag{12.37}$$

Now, the value of f_B that satisfies both Eqs. (12.36) and (12.37) is

$$f_B = f_L\left(1 - \frac{2n}{c}V_a\right) \tag{12.38}$$

Thus, the frequency of the Brillouin scattering is shifted by $(2n/c)V_a f_L$, which is much smaller than f_L because V_a/c is quite small. For the case of a fused silica fiber, $V_a = 5960$ ms^{-1}, $\lambda_L = 1.5$ μm, and $n = 1.44$. The frequency difference between the signal and the local oscillator is 11.4 GHz.

12.9 JITTER IN COHERENT COMMUNICATION SYSTEMS

The two kinds of jitter (unpredictable change) that impact coherent communication systems are polarization jitter and phase jitter. The presence of these jitters will cause the output signal to wax and wane. If the jitter is severe enough, there will be times when the signal fades completely, thus interrupting communication. Countermeasures have to be incorporated.

12.9.1 Polarization Jitter Controls

When linearly polarized light is sent through a fiber that does not preserve the state of polarization of the input light, the received signal is generally elliptically polarized. For optimum mixing, the state of polarization of both the received signal and the local oscillator signal should match. A means of adjusting the state of polarization of one or the other is required.

Polarization-maintaining fibers could be used in place of polarization controllers; however, this is not always an option when using existing cables that have been installed with non-polarization-maintaining fiber.

The fiber-loop polarization controller described in Chapter 6 can be used to convert the polarization of the received light into a state that is linearly polarized parallel to the local oscillator light. However, the fiber-loop polarizer [21,22] cannot cope with a fluctuating state of polarization. In ordinary optical fibers, the polarization is apt to change over time because of changes in stress caused by temperature changes or by mechanical vibration of the transmission fiber [23]. An automatic polarization control is necessary to dynamically correct for changes in polarization.

Example 12.1 Both the signal and the local oscillator signals are linearly polarized, but they are not parallel to each other on the input surface of the mixer. What is the IF output current when they are angled spatially at ϕ as shown in Fig. 12.16?

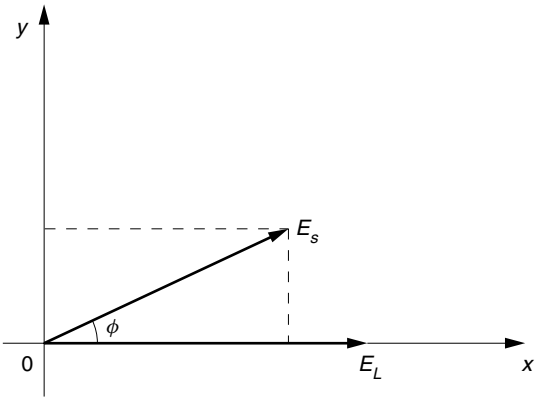


Figure 12.16 IF output current when E_S and E_L have a separation angle of ϕ.

Solution Let us say the input signal and local oscillator signal are

$$\begin{aligned} E_s &= S(t)\cos(\omega_c t + \theta) \\ E_L &= E_{\text{LO}}\cos\omega_L t \end{aligned} \tag{12.39}$$

The component of the signal parallel to the local oscillator is $E_s \cos\phi$ and that of the perpendicular component is $E_s \sin\phi$.

The output from the mixer is proportional to the power or square of the resultant **E** field, which is the sum of the squares of resultant fields parallel and perpendicular to the local oscillator field.

$$\begin{aligned} i &= 2K\langle[E_{\text{LO}}\cos\omega_L t + S(t)\cos\phi\cos(\omega_c t + \theta)]^2 \\ &\quad + [S(t)\sin\phi\cos(\omega_c t + \theta)]^2\rangle \\ &= 2K\left[\frac{1}{2}E_{\text{LO}}^2 + \frac{1}{2}S^2(t) + E_{\text{LO}}S(t)\cos\phi\cos(\omega_{\text{IF}}t + \theta)\right] \end{aligned} \tag{12.40}$$

Thus, the IF current component is the third term in Eq. (12.40).

$$i_{\text{IF}} = 2KE_{\text{LO}}S(t)\cos\phi\cos(\omega_{\text{IF}}t + \theta) \tag{12.41}$$

In conclusion, only the field component that is parallel to the local oscillator field contributes to the IF current. □

There are two approaches for automatically correcting the state of polarization. These are the computer-controlled method and the polarization diversity method. As evident from the name, the computer-controlled approach corrects the polarization of the received signal through computer control. The polarization diversity method uses an analog circuit that can process the signal independent of the state of polarization.

12.9.1.1 Computer-Controlled Method of Jitter Control

A designated computer either searches for the maximum output by interrogating the reaction to a small perturbation of the control signal, or actually monitors the state of

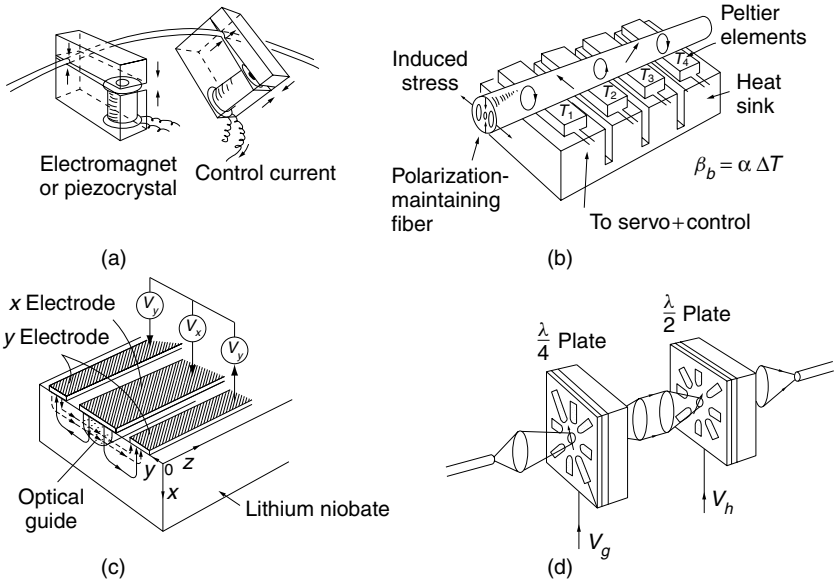


Figure 12.17 Various types of automatic polarization controller elements. (a) Based on squeezing the fiber. (After A. S. Siddiqui and Y. Aihua [25].) (b) Based on heat. (After N. G. Walker and G. R. Walker [26].) (c) Based on the electrooptic effect. (After N. G. Walker and G. R. Walker [26].) (d) Electrically rotatable liquid crystal waveplates. (After T. Chiba, Y. Ohtera, and S. Kawakami [29].)

polarization at an intermediate stage and issues commands to the appropriate elements. In either way, elements are needed whose birefringence can be computer controlled [24]. Figure 12.17 shows examples of such elements.

The fiber squeezer polarization controller [25] makes use of the stress-induced birefringence created by squeezing the fiber, as mentioned in Chapter 6. The fiber is squeezed either by an electromagnet or piezoelectro mechanical transducers, as shown in Fig. 12.17a.

The heat-induced polarization controller in Fig. 12.17b uses the temperature dependence of a short length of the birefringence of a polarization-maintaining fiber [26]. The polarization-maintaining fiber, as discussed in Chapter 11, is made of two types of glass having different amounts of shrinkage at the time the glass fiber is set. The birefringence is zero at the temperature T_s when the glass is set. As the temperature T is decreased, the birefringence linearly increases [26] and

$$\beta_b = \alpha \Delta T \tag{12.42}$$

where

$$\Delta T = T_s - T$$

Thus, by changing the temperature, the birefringence can be controlled.

Another type of element is based on the electrooptic effect of an electrooptic [27,28] crystal. The birefringence is varied by an externally applied electric field. An example of an optical guide on a lithium niobate substrate is shown in Fig. 12.17c. The potential V_x between the center and both side electrodes controls the E_x field (vertical field),

and the potential V_y between the two side electrodes controls the E_y field (horizontal field). The proper combination of V_x and V_y provides the required birefringence.

In Fig. 12.17d, a couple of electrically rotatable waveplates are used to convert elliptically polarized input light into linearly polarized light [29]. Light of an arbitrary elliptical polarization can be converted by aligning the fast axis of the quarter-waveplate with either the major or minor axis of the polarization ellipse. The half-waveplate rotates the linearly polarized light to the desired polarization direction (See Section 6.4).

The electrically rotatable waveplates make use of the birefringence of a nematic liquid crystal. The director of the liquid crystal, which behaves like the crystal axis, can be rotated by rotating the external electric field (See Section 5.10.4.2).

12.9.1.2 Polarization Diversity Method

Figure 12.18 shows a block diagram of an example of the polarization diversity method [30,31]. Incident light of an arbitrary polarization state is first fed through a non polarizing beam splitter (NPBS) and then is decomposed into horizontally and vertically polarized waves by means of a polarizing beamsplitter (PBS). The local oscillator light is coupled through the NPBS and is decomposed into horizontally and vertically polarized waves by being fed into the PBS with its direction of polarization at an angle of 45° to the horizontal direction. The horizontally polarized waves of both the signal and the local oscillator are fed to mixer M_H while the vertically polarized waves are fed into mixer M_V. The output current from the mixer is the IF signal current. Each IF current is then squared by a square operator. The outputs from the square operators are summed to obtain the sum of the squares of the horizontal and vertical components. This sum is independent of the state of polarization of the input signal, as proved below.

Any state of polarization can be designated by two parameters. One is the phase difference α between the vertical and horizontal waves, and the other is the power fraction β between the two waves. Both α and β are subject to fluctuations during transmission. The polarization diversity method eliminates both α and β from the final

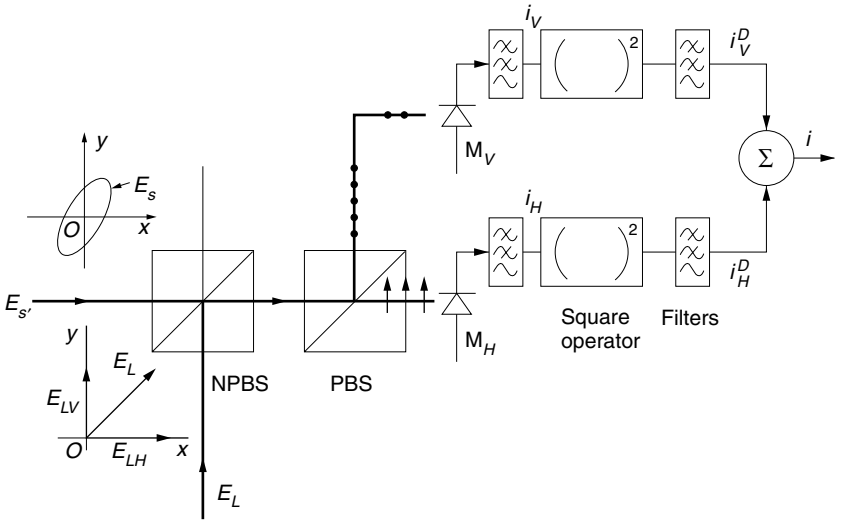


Figure 12.18 Block diagram of the polarization diversity method.

expression. The horizontally and vertically polarized signal and local oscillator fields are represented by

$$\left.\begin{aligned} E_H &= \sqrt{\beta}S(t)\cos\omega_c t \\ E_V &= \sqrt{(1-\beta)}S(t)\cos(\omega_c t+\alpha) \end{aligned}\right\} \tag{12.43}$$

$$E_{LH} = E_{LV} = \frac{1}{\sqrt{2}}E_{\mathrm{LO}}\cos\omega_L t \tag{12.44}$$

The outputs from mixers M_H and M_V are

$$i_H = 2K\left\langle\left(\sqrt{\beta}S(t)\cos\omega_c t + \frac{E_{\mathrm{LO}}}{\sqrt{2}}\cos\omega_L t\right)^2\right\rangle \tag{12.45}$$

$$i_V = 2K\left\langle\left(\sqrt{1-\beta}S(t)\cos(\omega_c t+\alpha) + \frac{E_{\mathrm{LO}}}{\sqrt{2}}\cos\omega_L t\right)^2\right\rangle \tag{12.46}$$

After time averaging and filtering the signal currents from M_H and M_V, the IF frequency components become

$$i_H = \sqrt{2}K\sqrt{\beta}S(t)E_{\mathrm{LO}}\cos\omega_{\mathrm{IF}}t \tag{12.47}$$

$$i_V = \sqrt{2}K\sqrt{1-\beta}S(t)E_{\mathrm{LO}}\cos(\omega_{\mathrm{IF}}t+\alpha) \tag{12.48}$$

The outputs from the square operator after low-pass filtering are

$$i_H^D = [KE_{\mathrm{LO}}S(t)]^2\beta \tag{12.49}$$

$$i_V^D = [KE_{\mathrm{LO}}S(t)]^2(1-\beta) \tag{12.50}$$

It is good to remember that the square operator followed by low-pass filtering always removes the effect of carrier phase fluctuation. Finally, the sum of i_H^D and i_V^D gives

$$i = [KE_{\mathrm{LO}}S(t)]^2 \tag{12.51}$$

which is completely independent of fluctuations in both α and β.

12.9.2 Phase Jitter

Besides polarization jitter, the other jitter to be concerned about is phase jitter, which is generated not only in the laser diode but also in the fiber due to temperature fluctuations and mechanical vibrations. For the amplitude or intensity modulated signal, the modulation information is extracted from only the amplitude of the signal, and the phase jitter does not show up in the output.

Phase jitter in the phase modulated system, however, is more difficult to remove, simply because the system has to identify the origin of the phase change and find out if it is due to the intended phase change associated with the phase modulation, or due to unwanted phase jitter of the system. The two-frequency method is one of the ways to remove the phase jitter and still retain the modulated phase information. The transmitter

sends two light signals whose carrier frequencies are slightly different [31]. Only one lightwave is phase modulated. The frequency of the other unmodulated lightwave is shifted by a small amount $\Delta\omega$. The unmodulated light is a tone signal used as reference. The two light signals are transmitted through the same fiber, at the same time, under the same conditions, and at the same state of polarization, and it is reasonable to expect that both suffer the same phase jitter during transmission. If the two phases are subtracted from each other at the receiver, only the intended phase modulation will be detected.

First, it is assumed that the polarization jitter is absent. Figure 12.19 shows the circuit of a differential phase shift keying (DPSK) system. The transmitter laser lightwave is divided into two lightwaves. One is phase modulated as $\Phi(t)$, and the other is the frequency-shifted tone $\omega_c + \Delta\omega$. In this example, the frequency shift $\Delta\omega$ is created by an acoustooptic modulator. Prior to entering the fiber-optic cable, both waves are combined such that the two waves are polarized in the same direction. At the receiver end, both waves are mixed with the local oscillator and undergo detection and signal processing. Both waves emerging from the fiber will have undergone similar phase and polarization jitter. A fiber-loop polarization controller [22] (FPC) is used to convert the waves back to a linear polarization. Then, the waves are mixed with the local oscillator through a nonpolarizing beamsplitter (NPBS). Let α denote the phase jitter from the fiber during transmission. The output from the mixer M is

$$i = 2K\langle\{\cos[\omega_c t + \Phi(t) + \theta + \theta_p + \alpha] + \cos[(\omega_c + \Delta\omega)t + \theta + \alpha] + E_{\mathrm{LO}}\cos(\omega_L t + \theta_L)\}^2\rangle \tag{12.52}$$

where θ, θ_p, and θ_L are the phase jitters in the transmitter laser, the phase modulator, and the local oscillator laser, respectively. For simplicity, the wave amplitude $S(t)$ has been taken as unity.

The output is band-pass filtered so that the $\Delta\omega$ component is rejected. Only the ω_{IF} and $(\omega_{\mathrm{IF}} + \Delta\omega)$ components are retained, where $2\omega_c$, $2\omega_L$, and $(\omega_c + \omega_L)$ terms are ignored.

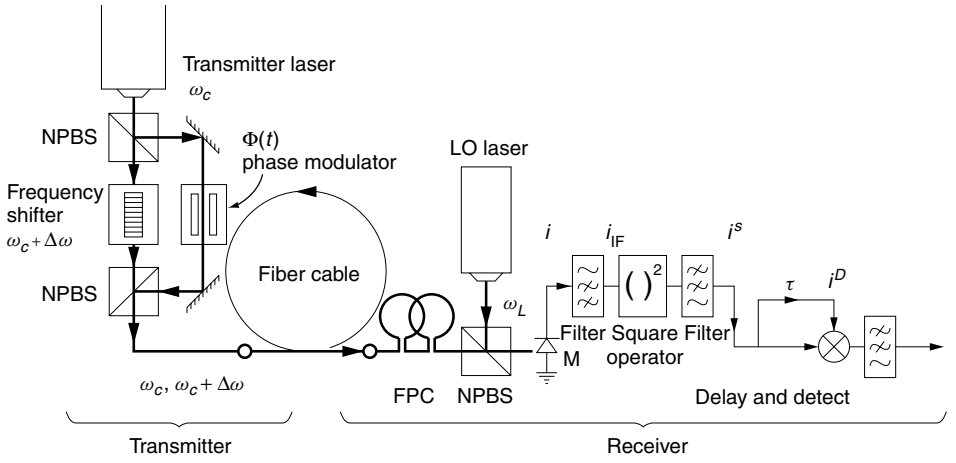


Figure 12.19 Phase jitter removal from a DPSK system by a frequency shift delay- and-detect method.

The output current from the filter is

$$\begin{aligned} i_{\mathrm{IF}} = 2KE_{\mathrm{LO}}\{&\cos(\omega_{\mathrm{IF}}t + \Phi(t) + \theta - \theta_L + \theta_p + \alpha] \\ &+ \cos[(\omega_{\mathrm{IF}} + \Delta\omega)t + \theta - \theta_L + \alpha]\} \end{aligned} \tag{12.53}$$

After the square operator, the term with $\Delta\omega$ is selected by filtering. The ac output is

$$i^s = 4(KE_{\mathrm{LO}})^2 \cos[\Delta\omega t - \Phi(t) - \theta_p] \tag{12.54}$$

Equation (12.54) contains the desired phase modulation information $\Phi(t)$ and is free from fluctuations in θ, θ_L, and α. However, the phase jitter term from the phase modulator itself is still present.

DPSK is the coding system that uses the change in the phase of the carrier. A "0" is represented by the absence of change; a "1" is represented by the presence of change. The means by which the delay-and-detect circuit interrogates the DPSK coding will be explained. Besides the function of interrogation, the delay-and-detect circuit also removes the θ_p noise. The signal path of the delay-and-detect circuit is divided into two paths, one having a shorter length than the other. The output from the longer path is time delayed from that of the shorter path by exactly one bit period. The sum of these two signals are squared, and all the terms except the product of these two signals are filtered out. The result is

$$i^D = K_1 \cos[\Delta\omega(t - \tau) - \Phi(t - \tau) - \theta_p] \cos[\Delta\omega t - \Phi(t) - \theta_p] \tag{12.55}$$

where K_1 absorbs all the physical parameters, and τ is one bit period. The output is fed through a low-pass filter and

$$i^D = \tfrac{1}{2}K_1 \cos[\Delta\omega\tau + \Phi(t - \tau) - \Phi(t)] \tag{12.56}$$

Imposing the condition that

$$\Delta\omega\tau = 2N\pi, \quad \text{where } N \text{ is an integer} \tag{12.57}$$

the "0" bit and the "1" bit of the phase modulation satisfy

$$\begin{aligned} \Phi(t - \tau) &= \Phi(t) && \text{"0" bit} \\ \Phi(t - \tau) &= \Phi(t) + \pi && \text{"1" bit} \end{aligned} \tag{12.58}$$

Thus, the output from the detector becomes

$$i^D = \begin{cases} K/2 & \text{for "0" bit} \\ -K/2 & \text{for "1" bit} \end{cases} \tag{12.59}$$

Therefore, all the phase jitter is removed from the detected signal. The signs of $K/2$ in Eq. (12.59) are reversed if $2N$ in Eq. (12.57) is replaced by $2N + 1$.

It may be added that even though the delay-and-detect circuit has the merit of simplicity, there is a restriction on the choice of the bit rate, as specified by Eq. (12.57). Equation (12.57) can be rewritten in terms of the bit rate $B = 1/\tau$ as

$$B = \frac{\Delta f}{N} \tag{12.60}$$

where $\Delta f = \Delta\omega/2\pi$ and $N = 1, 2, 3\ldots$. The delay-and-detect circuit cannot function (See Problem 12.3) when

$$B = \frac{4}{(2N-1)}\Delta f \tag{12.61}$$

where $N = 0, 1, 2, 3\ldots$.

12.10 COHERENT DETECTION IMMUNE TO BOTH POLARIZATION AND PHASE JITTER

Two systems that are free from both polarization and phase jitter will be presented here. Both systems combine a circuit designed to eliminate phase jitter with a circuit designed to get rid of polarization jitter, but they combine them in different ways. The first example uses the two diversities separately while the second example uses the two diversities in a more comprehensive manner.

Figure 12.20 shows a block diagram of the first example [32]. Phase diversity is applied to the horizontal and vertical waves separately and polarization diversity

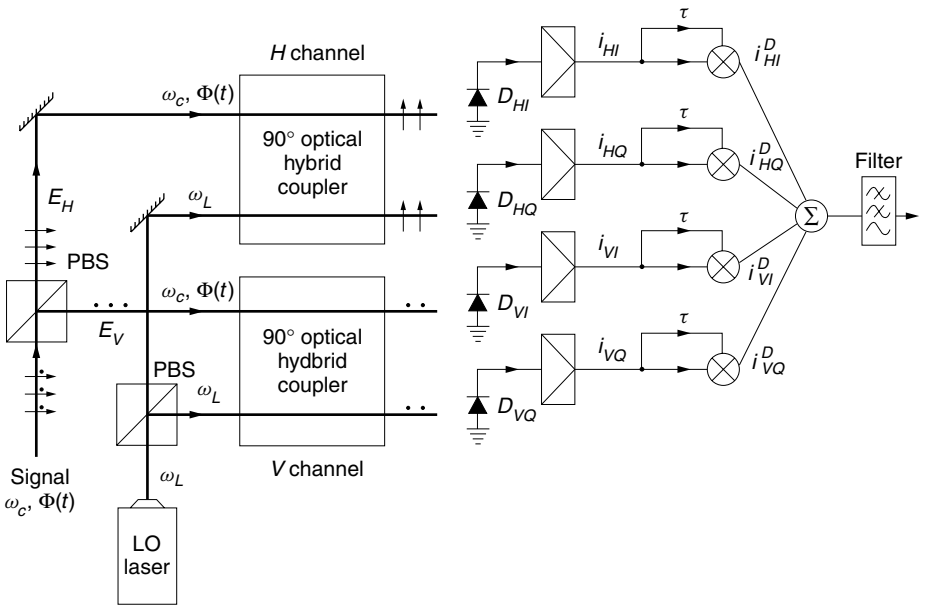


Figure 12.20 Block diagram of DPSK that is free from both polarization and phase jitter. (After Y. H. Cheng, T. Okoshi, and O. Ishida [32].)

is applied to these two phase-jitter-free waves. This example again uses DPSK modulation. The received signal is first fed to the polarization beamsplitter. The horizontal and vertical wave outputs from the beamsplitter are

$$E_H = \sqrt{\beta} S(t) \cos[\omega_c t + \Phi(t) + \theta_H] \tag{12.62}$$

$$E_V = \sqrt{1-\beta} S(t) \cos[\omega_c t + \Phi(t) + \theta_V] \tag{12.63}$$

where the fluctuating quantities are β, θ_H, and θ_V. They are fed to separate 90° optical hybrid couplers together with the local oscillator light. The output from the 90° optical hybrid is similar to the one in Fig. 12.9 described in connection with the optical Costas loop in Section 12.6.2. The in-phase and quadrature phase IF frequency outputs for the horizontal wave components are

$$i_{HI} = \sqrt{2} K \sqrt{\beta} E_{\text{LO}} S(t) \cos[\omega_{\text{IF}} t + \Phi(t) + \theta_H] \tag{12.64}$$

$$i_{HQ} = \sqrt{2} K \sqrt{\beta} E_{\text{LO}} S(t) \sin[\omega_{\text{IF}} + \Phi(t) + \theta_H] \tag{12.65}$$

where the local oscillator power is equally divided into each 90° hybrid coupler and its amplitude is $(1/\sqrt{2})E_{\text{LO}}$.

Similarly, the outputs from the 90° hybrid coupler for the vertical wave components are

$$i_{VI} = \sqrt{2} K \sqrt{(1-\beta)} E_{\text{LO}} S(t) \cos[\omega_{\text{IF}} t + \Phi(t) + \theta_V] \tag{12.66}$$

$$i_{VQ} = \sqrt{2} K \sqrt{(1-\beta)} E_{\text{LO}} S(t) \sin[\omega_{\text{IF}} t + \Phi(t) + \theta_V] \tag{12.67}$$

These IF signals are separately amplified and then fed to sets of delay-and-detect circuits. The delay-and-detect circuit performs the operation of the product between the present and previous bits, as described in Section 12.9.2.

Now let us obtain the horizontal wave output from the top delay-and-detect circuits. Note that the delay-and-detect circuit in Fig. 12.20 does not employ a low-pass filter as compared to that in Fig. 12.19. The output for the in-phase component is the product of i_{HI} at $t = t$ and $t = t - \tau$ with $\Phi(t - \tau) = \Phi(\tau)$ for "0" and $\Phi(t - \tau) = \Phi(t) + \pi$ for "1".

$$i_{HI}^D = \pm \beta K_m [K E_{\text{LO}} S(t)]^2 \cos^2[\omega_{\text{IF}} t + \Phi(t) + \theta_H] \tag{12.68}$$

where the + sign corresponds to the "0" bit and the − sign to the "1" bit, and where K_m is a constant associated with the electronic mixer and $\omega_{\text{IF}} \tau = 2N\pi$. Similarly, the output for the quadrature phase component is

$$i_{HQ}^D = \pm \beta K_m [K E_{\text{LO}} S(t)]^2 \sin^2[\omega_{\text{IF}} t + \Phi(t) + \theta_H] \tag{12.69}$$

The sum of Eqs. (12.68) and (12.69) is independent of ω_{IF}, $\Phi(t)$, and θ_H.

$$i_H = \pm \beta K_m [K E_{\text{LO}} S(t)]^2 \tag{12.70}$$

The components for the vertically polarized waves are

$$i_{VI}^{D} = \pm(1-\beta)K_m[KE_{\mathrm{LO}}S(t)]^2\cos^2[\omega_{\mathrm{IF}}t + \Phi(t) + \theta_V] \tag{12.71}$$

$$i_{VQ}^{D} = \pm(1-\beta)K_m[KE_{\mathrm{LO}}S(t)]^2\sin^2[\omega_{\mathrm{IF}}t + \Phi(t) + \theta_V] \tag{12.72}$$

Addition of Eqs. (12.71) and (12.72) gives

$$i_V = \pm(1-\beta)K_m[KE_{\mathrm{LO}}S(t)]^2 \tag{12.73}$$

Finally, the sum of Eqs. (12.70) and (12.73) removes the fluctuation of β in polarization, and gives a completely jitter-free signal:

$$i = \pm K_m[KE_{\mathrm{LO}}S(t)]^2 \tag{12.74}$$

The system with the square operation that also achieves jitter-free operation for DPSK modulation is shown in Fig. 12.21. This system combines the two diversity methods more comprehensively [33]. This system uses the frequency shift method for phase jitter cancellation, and the square and addition operation for polarization diversity. This system is simpler than the previous system in that both the unshifted signal and the frequency-shifted signal share the same circuit, but it requires a frequency-shifted signal light. The transmitting light consists of two waves. One is at frequency ω_c and is phase modulated by $\Phi(t)$ of the DPSK data. The other is frequency shifted to $\omega_c + \Delta\omega$ but is not modulated. Both of these waves are linearly polarized and the directions of polarization of the two waves are collinear.

Just the same as the system in Fig. 12.19, during the traversal of the fiber, both waves experience the same phase jitter and change of state of polarization because the

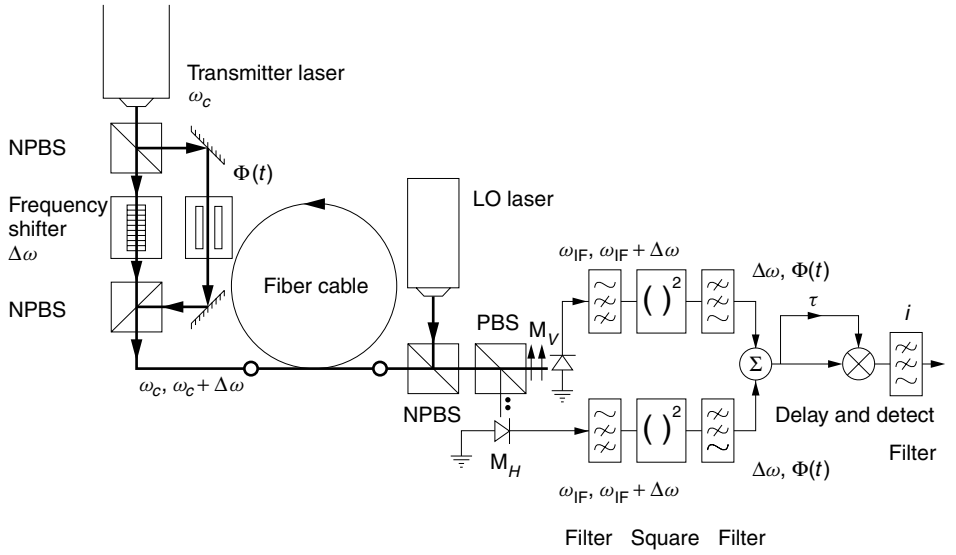


Figure 12.21 System with the square operation for the removal of both polarization and phase jitter from DPSK.

only difference between the two waves is that one is frequency shifted by $\Delta\omega$. The received signals become

$$E_H = \frac{1}{\sqrt{2}}\sqrt{\beta}S(t)\{\cos[\omega_c t + \Phi(t) + \theta + \theta_p + \theta_H] + \cos[(\omega_c + \Delta\omega)t + \theta + \theta_H]\} \tag{12.75}$$

$$E_V = \frac{1}{\sqrt{2}}\sqrt{1-\beta}S(t)\{\cos[\omega_c t + \Phi(t) + \theta_p + \theta_V] + \cos[(\omega_c + \Delta\omega)t + \theta + \theta_V]\} \tag{12.76}$$

where θ_p is the jitter in the phase modulator and θ, θ_H, and θ_V are the same as before and are the phase jitters of the source, the horizontal wave, and the vertical wave. The received wave is separated into horizontal and vertical wave channels. The local oscillator signal is

$$E_L = E_{\text{LO}} \cos(\omega_L t + \theta_L) \tag{12.77}$$

which is polarized at 45° to the horizontal direction, providing an equal-amplitude ($E_{\text{LO}}/\sqrt{2}$) local oscillator field to each channel, where E_{LO}, ω_L, and θ_L are the local oscillator amplitude, frequency, and phase jitter. The analysis of the operation is quite similar to that of the phase diversity without polarization diversity described in Section 12.9.1.2 and will be repeated only briefly. In each channel, the signal is converted down to the IF frequency by the mixers M_H and M_V. In each channel, a band-pass filter follows the mixer and filters out the $\Delta\omega$ component, keeping only ω_{IF} and $\omega_{\text{IF}} + \Delta\omega$ components. These two components experience the same phase jitter, and the next processing step consisting of the square operation followed by summation serves to cancel the phase jitter. Thus, signals in both the H and V channels become free from phase jitter. Using results similar to Eq. (12.53) as the input to the square operator, the squared results of the H and V channels are

$$i_H = 2[KE_{\text{LO}}S(t)]^2\beta\cos[\Delta\omega t - \Phi(t) - \theta_p] \tag{12.78}$$

$$i_V = 2[KE_{\text{LO}}S(t)]^2(1-\beta)\cos[\Delta\omega t - \Phi(t) - \theta_p] \tag{12.79}$$

Addition of the two outputs becomes independent of fluctuating β and

$$i = 2[KE_{\text{LO}}S(t)]^2\cos[\Delta\omega t - \Phi(t) - \theta_p] \tag{12.80}$$

Finally, the signal from the delay-and-detect of Eq. (12.80), followed by a low-pass filter, becomes

$$i = \frac{K_1}{2}\cos[\Delta\omega\tau + \Phi(t-\tau) - \Phi(t)] \tag{12.81}$$

where K_1 absorbs all the physical parameters. The output is completely free from both phase and polarization jitter. An advantage of this system over the previous system is that the lightwaves of ω_c and $\omega_c + \Delta\omega$ share the same paths, and the receiving system is simpler in this regard. The disadvantage is the requirement of a more elaborate transmitting system.

The existing PBS in Fig. 12.21 cannot be used to mix the local oscillator beam with the signal beam, thereby sparing a NPBS, because the direction of the polarizations of the emergent signal and the local oscillator beams become perpendicular to each other.

12.11 CONCLUDING REMARKS

At the receiving end, the transmitted light signal is converted back into the original encoded electrical signal. Receiving systems can broadly be categorized into the following five systems:

1. Direct detection (or video detection).
2. Heterodyne detection.
3. Homodyne detection.
4. Intradyne detection.
5. Detection by stimulated effects.

The advantages and disadvantages of these systems are summarized below.

1. The direct detection scheme is simple, but the trade-off for simplicity is mediocre sensitivity and the lack of a built-in mechanism for channel selectivity. With direct detection, channel selectivity is achieved with additional components such as wavelength filters. Even when such filters are used, channel selectivity is generally poorer than what can be achieved with heterodyne and homodyne detection.
2. The heterodyne system is equipped with a local oscillator laser and is therefore more complicated than direct detection, but it enjoys more than a 20-dB improvement in receiver sensitivity as well as more than 10^3 times the wavelength selectivity.
3. The homodyne system is similar to the heterodyne system but with zero intermediate frequency (IF). With zero IF, the frequency bandwidth per channel is smaller, and the receiver gain is 3 dB higher, compared to a heterodyne system. The price to be paid is that the homodyne system requires a local oscillator laser that can track the phase of the signal wave, and this requirement is very severe.
4. Intradyne detection is a mixture of heterodyne and homodyne detection systems. The stability requirement of the local oscillator is not as severe as that of the homodyne detection.
5. The stimulated effect uses stimulated Brillouin scattering and the system can be interpreted as detection with built-in phase tracking. The narrowness of the frequency bandwidth and difficulty in applying ASK modulation are disadvantages of this system.

PROBLEMS

12.1 (a) What is the theoretical limit ($\eta = 1$) for the output current from a PIN photodiode when the power of the incident light is -43 dBm and the wavelength is $\lambda = 0.84$ μm?

(b) What value was assigned to the quantum efficiency to obtain the rule of thumb equation $i = 0.5$ A/W for expressing the output current i from a PIN photodiode?

12.2 With the polarization diversified circuit shown in Fig. 12.18, prove that as long as the ASK format is used, the phase jitter incurred in both the transmitting and local oscillator lasers, as well as in the fiber during transmission, can be removed.

12.3 Prove that the delay-and-detect circuit cannot function when the bit rate B is

$$B = \frac{4}{2N-1}\Delta f$$

where Δf is the frequency shift between the two frequencies of the double-frequency scheme mentioned in Section 12.9.2.

REFERENCES

1. Y. Tsuchiya, "Streak cameras for ultrafast time resolved photon detection," *Proc. SPIE* **1599**, 244–270 (1991).
2. C. Kittel, *Introduction to Solid State Physics*, 7th ed., Wiley, New York, 1996.
3. B. Sapoval and C. Herman, *Physics of Semiconductors*, Springer-Verlag, Berlin, 1995.
4. H. Yonezu, *Engineering of Components of Fiber Optic Communication*, 2nd ed. (in Japanese), Kogakutosho, Tokyo, 1986.
5. F. Zappa, A. Lacaita, S. Cova, and P. Lovati, "Counting, timing, and tracking with a single-photon germanium detector," *Opt. Lett.* **21**(1), 59–61 (1996).
6. S. Ryu, *Coherent Lightwave Communication Systems*, Artech House, Boston, 1995.
7. T. Okoshi, "Recent advances in coherent optical fiber communication systems," *J. Lightwave Technol.* **5**(1), 44–52 (1987).
8. L. G. Kazovsky, S. Benedetto, and A. Willner, *Optical Fiber Communication Systems*, Artech House, Boston, 1996.
9. M. A. Grant, W. C. Michie, and M. J. Fletcher, "The performance of optical phase-locked loops in the presence of nonnegligible loop propagation delay," *J. Lightwave Technol.*, **5**(4), 592–597 (1987).
10. A. W. Davis, M. J. Pettitt, J. P. King, and S. Wright, "Phase diversity techniques for coherent optical receivers," *J. Lightwave Technol.* **5**(4), 561–572 (1987).
11. G. Nicholson and T. D. Stephens, "Performance analysis of coherent optical phase diversity receivers with DPSK modulation," *J. Lightwave Technol.* **7**(2), 393–399 (1989).
12. I. B. Djordjević, M. C. Stefanović, S. S. Ilić, and G. T. Djordjević, "An example of a hybrid system: Coherent optical system with Costas loop in receiver-system for transmission in baseband," *J. Lightwave Technol.* **16**(2), 177–183 (1998).
13. G. L. Abbas, V. W. S. Chan, and T. K. Yee, "A dual-detector optical heterodyne receiver for local oscillator noise suppression," *J. Lightwave Technol.* **3**(5), 1110–1122 (1985).
14. D. Cotter, "Stimulated Brillouin scattering in monomode optical fiber," *J. Opt. Commun.* **4**, 10–19 (1983).
15. Y. -H. Lee, H. -W. Tsao, M. -S. Kao, and J. Wu, "Application of fiber Brillouin amplifiers on coherent optical wavelength division and subcarrier-multiplexing (WDM-SCM) system," *J. Opt. Commun.* **13**(3), 99–103 (1992).

16. S. Fischer, P. Meyrueis, and W. Schroder, "Brillouin backscattering in fiber optic gyroscope," *Proc. SPIE* **2510**, 49–58 (1995).
17. G. P. Agrawal, *Fiber-Optic Communication Systems*, 2nd ed., Wiley, New York, 1997.
18. U. L. Österberg, "Nonlinear optics: Theory and applications," *Trends in Optical Fibre Metrology and Standards*, edited by O. D. D. Soares, *Proceedings of the NATO Advanced Study Institute*, **285**, Viana do Castelo, Portugal, pp. 711–725, 1994.
19. E. M. Dianov, M. V. Grekov, L. A. Bufetov, S. A. Vasiliev, O. I. Medvedkov, G. A. Ivanov, A. V. Belov, V. G. Plotnichenko, V. V. Koltashev, and A. M. Prokhorov, "Phosphosilicate fiber: Simple high-power CW 1.24 and 1.48 μm Raman lasers," *CLEO'98*, San Francisco, p. 225, May 3–8, 1998.
20. J. -L. Archambault and S. G. Grubb, "Fiber gratings in lasers and amplifiers," *J. Lightwave Technol.* **15**(8), 1378–1390 (1997).
21. R. Ulrich, S. C. Rashleigh, and W. Eickhoff, "Bending-induced birefringence in single-mode fibers," *Opt. Lett.* **5**(6), 273–275 (1980).
22. H. C. Lefevre, "Single-mode fiber fractional wave devices and polarization controllers," *Electron. Lett.* **16**(20), 778–780 (1980).
23. R. A. Harmon, "Polarization stability in long lengths of monomode fibre," *Electron. Lett.* **18**(24), 1058–1059 (1982).
24. I. Andonovic and D. Uttamchandani, *Principles of Modern Optical Systems*, Artech House, Norwood, MA, 1989.
25. A. S. Siddiqui and Y. Aihua, "A new method to analyze the endless characteristics of polarization controllers," *J. Opt. Commun.* **10**(4), 141–143 (1989).
26. N. G. Walker and G. R. Walker, "Survey and current status of polarization control techniques for coherent optical communications," *Proc. SPIE* **1175**, 67–79 (1990).
27. N. G. Walker and G. R. Walker, "Polarization control for coherent communications," *J. Lightwave Technol.* **8**(3), 438–458 (1990).
28. P. T. Fredericksen, M. Flüge, E. Dietrich, and H. G. Weber, "A fast polarization controller for coherent phase diversity receivers," *J. Opt. Commun.* **10**(4), 149–153 (1989).
29. T. Chiba, Y. Ohtera, and S. Kawakami, "Polarization stabilizer using liquid crystal rotatable waveplates," *J. Lightwave Technol.* **17**(5), 885–890 (1999).
30. B. Glance, "Polarization independent coherent optical receiver," *J. Lightwave Technol.* **5**(2), 274–276 (1987).
31. R. T. B. James, K. Iizuka, and Y. Imai, "Phase noise cancelled polarization diversity scheme for DPSK optical communication," *Electron. Lett.* **25**(20), 1394–1396 (Sept. 28, 1989).
32. Y. H. Cheng, T. Okoshi, and O. Ishida, "Performance analysis and experiment of a homodyne receiver insensitive to both polarization and phase fluctuations," *J. Lightwave Technol.* **7**(2), 368–374 (1989).
33. Y. Imai, K. Iizuka, and R. T. B. James, "Phase-noise-free coherent optical communication system utilizing differential polarization shift keying (DpolSK)," *J. Lightwave Technol.* **8**(5), 691–697 (1990).

13

OPTICAL AMPLIFIERS

The word laser is an acronym for lightwave amplification by stimulated emission of radiation. As this acronym includes the word amplification, the optical amplifier and the laser are closely related. They are so closely related that the order of presentation does not really matter. Historically, commercialization of lasers preceded that of optical amplifiers, but for heuristic reasons, the chapter on amplifiers is placed ahead of the chapter on lasers. Optical amplifiers suppress the reflection of light to prevent self-oscillation while lasers enhance the reflection to create oscillation. All other basic principles are common to both devices.

13.1 INTRODUCTION

Optical amplifiers are used as power amplifiers, line amplifiers, and preamplifiers. At the transmitter end, if the light signal is externally modulated or divided into multiple channels, the output from the transmitter is reduced. This reduced light power is amplified by a power amplifier before it is transmitted into the fiber.

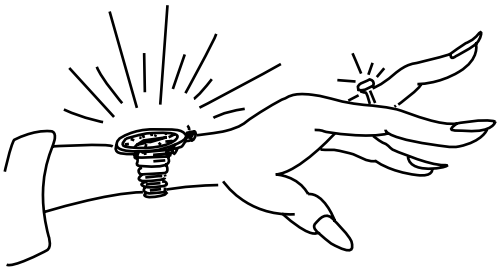
One of the most important applications of the optical amplifier is in repeater stations. Long-haul transmission through a fiber-optic communication line requires repeaters; otherwise, the signal level becomes too low to detect. Repeater design is greatly simplified by using an optical amplifier. Without the optical amplifier, the input light to the repeater is first detected and converted into an electrical signal. The electrical signal is reshaped into a well-defined pulse shape, and the timing of the pulse is readjusted. This new electrical pulse is used to regenerate the light power to send to the next repeater station. In the repeater station, reshaping, retiming, and regeneration, the so-called 3Rs, have to be performed. This rather involved process can be replaced by an optical amplifier. Simplicity and reliability of the repeater compartment are especially important when the lightwave cable is used as an undersea transmission cable.

Another application of optical amplifiers is in the receiver as a preamplifier. When the received light signal is preamplified by an optical amplifier before direct detection, the sensitivity is enhanced.

13.2 BASICS OF OPTICAL AMPLIFIERS

Laser materials are fluorescent, and fluorescence is the key to understanding both lasers and optical amplifiers. As a matter of fact, when searching for a new laser material, the fluorescent properties of the material are the first thing to be checked. A fluorescent substance is one that glows when illuminated, and additionally, the color of the glow is different from the color of the illuminating light. The illuminating light is called the pump light. Even though fluorescence begins as soon as the pump light is turned on, when the pump light is turned off, the fluorescence is sustained. It decays with a characteristic lifetime. Another feature of fluorescence is that the color of the glow is very specific, whereas the color of the pump light does not have to be as specific.

From these observations, scientists have found an explanation using the concept of energy levels and transitions between energy levels. The model shown in Fig. 13.1 will be used as a starting point. This model is known as the three-energy-level model. Pumping takes place between energy levels E_1 and E_3, where $h\nu_{31} = E_3 - E_1$ and ν_{31} is the frequency of the pump light. The fluorescent glow takes place between E_2 and E_1, where $h\nu_{21} = E_2 - E_1$ and ν_{21} is the frequency of the glow light.



Fluorescent crystals are gems for lasers and amplifiers.

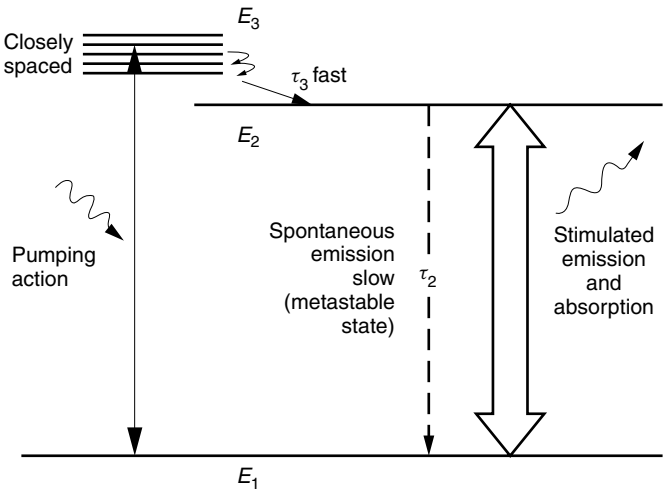


Figure 13.1 Principle of a three-level amplifier.

Energy level E_3 is actually a bundle of many closely spaced energy levels, rather than one discrete level. When the pump light is turned on, upward transitions from E_1 to the E_3 band take place, provided the energies contained in the pump light match the E_1 to E_3 band transition. Almost immediately, downward transitions are initiated between the closely spaced E_3 levels, as well as from E_3 to the nearby level E_2. Because of the narrow spacing, the transitions between the E_3 levels are primarily associated with phonons and nonradiative transitions and occur very quickly over lifetimes of the order of femtoseconds to nanoseconds. The released energy is converted into crystal lattice vibrations or phonons with energy $h\nu_{\text{phonon}} = E_{n+1} - E_n$.

According to Fig. 13.1, there is a large energy gap between E_2 and E_1, which means photons are involved rather than phonons. The downward transition from E_2 to E_1 is responsible for the fluorescent glow. We are particularly interested in materials where E_2 is a metastable state, which means the downward transition from E_2 to E_1 occurs over a lifetime τ_2 of milliseconds to hours. Thus, $\tau_2 > \tau_3$ and the glow lasts a long time after the pump has been turned off.

E_1 and E_2 are discrete levels, whereas E_3 is a band of levels. Hence, the wavelength of the fluorescent glow is very specific (E_2 to E_1 transition), but bands of wavelengths will work for the pump light (E_1 to E_3. band transition).

When the transition from E_2 to E_1 occurs spontaneously, a photon of energy $h\nu_{21}$ is released. This process is called *spontaneous emission*. Spontaneously emitted photons not only travel in different directions but also have different phases. These photons are said to be incoherent.

There is another important mechanism for the emission of light that is called *stimulated emission*. Stimulated emission will be explained by referring once again to the three-level model in Fig. 13.1. The pump light causes the population of atoms in level E_3 to increase. This population increase in E_3, however, is quickly transferred to that of E_2 because of the fast decay from E_3 to E_2, finally resulting in the population buildup in E_2 because of long τ_2. If the material is illuminated by light of frequency ν_{21} while this buildup of population in E_2 exists, a significant increase in the light intensity is observed at frequency ν_{21}. The amount of the increase is proportional to the ν_{21} illumination.

Let's take a closer look at what is going on. Before the buildup in level E_2 has a chance to decay spontaneously, the input photons of energy $h\nu_{21}$ come along. Because these photons happen to have the right frequency, they induce downward transitions from E_2 to E_1. For each induced downward transition, a photon of energy $h\nu_{21}$ is released. Furthermore, both the released photon and the photon that induced the release are not only identical in frequency but also in phase and direction. Hence, the ν_{21} stimulated emission is coherent with the ν_{21} illuminating light.

Stimulated emission is the basis for lasers and optical amplifiers. In a laser, photons capable of causing stimulated emission make multiple passes through a laser cavity, inducing the release of photons of identical energy with each pass. A coherent beam of collimated light with a narrow frequency bandwidth is created. In an optical amplifier, the photons capable of causing stimulated emission (supplied by the signal light) make a single pass through the material. The power of the signal light is amplified by the stimulated emission while traveling through the amplifier.

Emission and the reverse process of absorption can be interpreted as a kind of resonance phenomenon taking place between level E_1 and E_2. A photon is emitted whenever there is a downward transition from E_2 to E_1, and a photon is absorbed

whenever there is an upward transition from E_1 to E_2. Moreover, their transition probabilities are identical.

Why then does stimulated emission grow? Would not absorption cancel out the emission? The answer has to do with the populations in levels E_1 and E_2. Under normal circumstances, E_1 is more heavily populated, and absorption dominates. When E_2 is more populated than E_1, this situation is called a population inversion, and in this case, emission dominates. In the three-level model, a population inversion exists between E_1 and E_2 because the pump light keeps E_2 well supplied via the fast decay from E_3 and slow decay to E_1.

If there is no signal light at frequency ν_{21}, then the emission is mainly spontaneous emission. When both a population inversion and input signal light of the right frequency ν_{21} are present, then the output light due to the stimulated emission grows in proportion to the strength of the input signal light, with the population inversion as a proportionality constant.

In the amplifier, there exists not only stimulated emission that is used for amplification, but also spontaneous emission that harms the quality of the amplifier. Spontaneous emission is independent of the input signal light and causes noise problems. Indeed, this noise source is a major concern when using an optical amplifier. Choosing an amplifying medium with a long lifetime is important because the probability of spontaneous emission is $1/\tau_{21}$.

An example of an amplifier that corresponds to the three-level model is the erbium-doped fiber amplifier. This amplifier is used at $\lambda = 1.55$ μm. Not all amplifiers are based on the three-level model. An example of an amplifier based on a four-level model is the neodymium (Nd)-doped fiber amplifier, which is useful at $\lambda = 1.06$ and 1.32 μm.

The four-level model has an extra level E_0 *below* E_1. E_0 is the ground level. The pumping takes place between E_0 and E_3 (see Fig. P13.1b). The population inversion is between E_2 and E_1. As before, the lifetime τ_{32} is short and the lifetime τ_{21} is long. The pumping action populates E_2 via the fast decay from E_3. Compared to the three-level model, the four-level model has a larger population inversion between E_2 and E_1 because E_1 is very quickly emptied out into E_0, which is constantly being evacuated by the pump light. The population inversion between E_2 and E_1 is not only larger but is also less sensitive to the condition of the pump light.

With the three-level material, however, the population in E_1 depends on the strength of the pumping action and its ability to evacuate level E_1. Consequently, the population inversion between E_2 and E_1 is more dependent on the pump power and likewise, the gain of the amplifier is more dependent on the pump power.

13.3 TYPES OF OPTICAL AMPLIFIERS

As shown in Fig. 13.2, there are two major types of optical amplifiers: the fiber amplifier and the semiconductor laser amplifier.

In the fiber amplifier, the doped core of an optical fiber serves as the amplifying medium as shown in Fig. 13.2a. As already mentioned, two examples of fiber amplifiers are the neodymium-doped fiber amplifier [1] that operates at 1.06 and 1.32 μm, and the erbium-doped fiber amplifier (EDFA) that operates in the window of lowest fiber loss at 1.55 μm. A third example of a fiber amplifier is the praseodymium-doped fiber

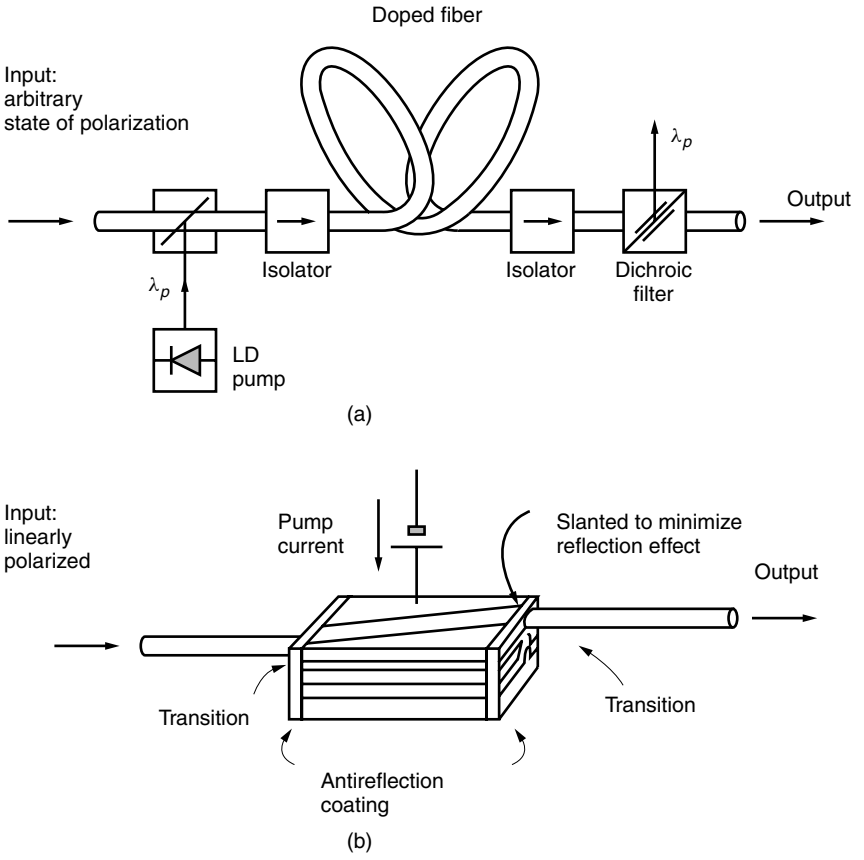


Figure 13.2 Two types of optical amplifiers. (a) Fiber amplifier. (b) Semiconductor laser amplifier.

amplifier (PDFA) that operates in the window of least fiber dispersion at 1.3 μm [2]. Of these three examples, the discussions to follow will feature the EDFA as a typical fiber amplifier.

In order to fabricate an erbium-doped fiber by the two stage process (see Section 11.10.2.1), the core soot of the preform is immersed in alcohol containing erbium chloride ($ErCl_3$). After the alcohol evaporates, the $ErCl_3$ remains in the core soot. The cladding soot is then deposited over the core soot, and the fiber is drawn.

The other type of optical amplifier is the semiconductor laser amplifier (SLA) as shown in Fig. 13.2b. The structure of the SLA is essentially the same as a laser diode, except that end reflections are suppressed in the SLA. The population inversion is achieved by injecting an electric current.

There are advantages and disadvantages to each type of amplifier. The EDFA is ideal for fiber-based systems. Since the EDFA is itself an optical fiber, connections with other optical fibers are easy to make. On the other hand, in systems that employ monolithic integrated optical wafers, the SLA is the amplifier of choice because it can be incorporated directly into the wafer.

The pumping schemes are different for the two types of amplifiers. A light pump is necessary to pump the EDFA. Usually the light pump is the light output from a laser

Table 13.1 Comparison between EDFA and SLA

Characteristic	EDFA	SLA
Structure	Not suitable for monolithic wafer	Suitable for monolithic wafer
Input coupling	To the fiber	To the optical guide
State of polarization	Independent	Dependent
Wavelength	Limited choice in wavelength	Various wavelengths
Method of pumping	Laser diode pump	Electric current
Amplifier gain	Higher	Lower
Saturated output power	Higher	Lower
ASE noise	Present	Present

diode. A coupler is needed to get the pump light into the fiber. An optical filter is required at the end of the fiber to prevent the pump light from emanating out of the amplifier.

By comparison, pumping in the case of the SLA is much simpler. In the SLA, pumping is achieved quite easily with the injection of an electric current.

In the EDFA, the operating wavelength is 1.53–1.55 μm, which matches the wavelength of the lowest fiber transmission loss. For a given fiber, the wavelength is not selectable. A greater range of operating wavelengths is available with the SLA.

The two types of amplifiers behave differently with respect to light polarization. The EDFA is independent of the polarization of the signal light. The same is not true of the SLA. Even when the active region of the SLA has a rectangular shape, the TE mode, whose **E** field is parallel to the surface of the waveguide, is amplified with higher gain than the TM mode. The performance of the SLA is polarization dependent.

Both the EDFA and the SLA generate inherent amplified spontaneous emission (ASE) noise. The gain of the EDFA, typically 30 dB, is higher than the 20-dB gain of the SLA. The level of power saturation is also higher for the EDFA, typically 20 dBm as compared to approximately 15 dBm for the SLA. Distortion of the signal is minimal in the case of the EDFA. Even in the region of gain compression, distortion is practically nonexistent.

In any amplifier, the presence of self-oscillation will cause the operation to become unstable. Self-oscillation is brought on by optical feedback caused by reflections. In continuous fiber systems, reflections inside the fiber are practically nonexistent, and self-oscillation is rarely a problem within the amplifier itself. In the SLA, however, the problem of end reflections can be severe, and elaborate antireflection measures are necessary in the design of the SLA. Both types of amplifiers are vulnerable to reflections that arise from how the various system components are connected together. An isolator is used to prevent oscillations due to the amplified light reflected back from the external components.

Table 13.1 summarizes the comparisons between the EDFA and the SLA.

13.4 GAIN OF OPTICAL FIBER AMPLIFIERS

The expression for the gain of an optical amplifier depends solely on the transitions between the two levels involved in the population inversion. This is true for both

three-level and four-level models. In the preliminary discussion of fluorescence, the levels involved in the population inversion were distinctly well-defined levels. However, the description can easily be extended to a population inversion between two *bands* of energy levels, as is often the case with semiconductor materials. Whether the transitions involve molecules, atoms, ions, or electrons in the energy bands depends on the material. For the sake of generality, the term "carrier" will be used to represent an atom, molecule, ion, or electron in a particular level.

Consider the two energy levels E_2 and E_1 shown in Fig. 13.3, disregarding pump level E_3. Energy is released in a downward transition of a carrier, and energy is absorbed in an upward transition. Stimulated transitions, that is, transitions initiated by the presence of an external photon, take place in both directions. In stimulated absorption, the external photon is absorbed as the upward transition takes place, and in stimulated emission, the external photon causes the release of a photon of identical energy as the downward transition takes place. Spontaneous emission is the release of a photon in a downward transition that happens of its own accord, without any external influences. There is no such thing as spontaneous absorption, or a spontaneous upward transition. Since there is only one type of absorption, the qualifier "stimulated' is often dropped, and the process is simply called absorption.

13.4.1 Spectral Lineshape

Every time a carrier makes a transition from E_2 to E_1, a burst of light with frequency $\nu_0 = (E_2 - E_1)/h$ is emitted. Each emission lasts only for a finite time as shown in Fig. 13.4. This finiteness of the duration of emission broadens the power spectrum of the light.

Keep in mind that if the pulse occurrence is completely random, as mentioned in Section 1.4.8, the intensity of the Fourier transform of a train of the pulses is the same as that of a single pulse multiplied by the number of the pulses. Let us find the power spectrum of each emission of light. The light intensity is assumed to decay exponentially with the intensity attenuation constant γ, or the amplitude attenuation constant $\gamma/2$. The amplitude of one of the bursts of light is expressed as

$$E(t) = E_0 e^{-(\gamma/2)t} \cos 2\pi\nu_0 t \tag{13.1}$$

with $t > 0$.

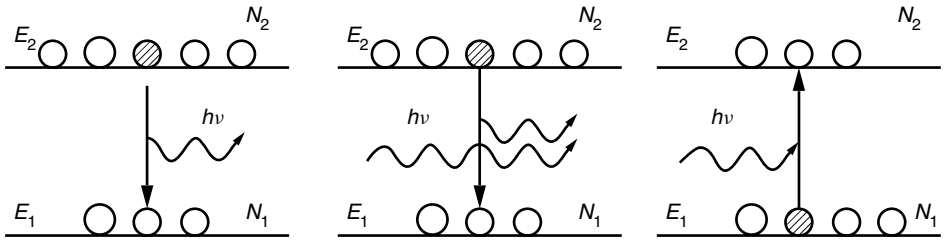


Figure 13.3 Transitions between two energy levels. (a) Spontaneous emission. (b) Stimulated emission. (c) Stimulated absoprtion.

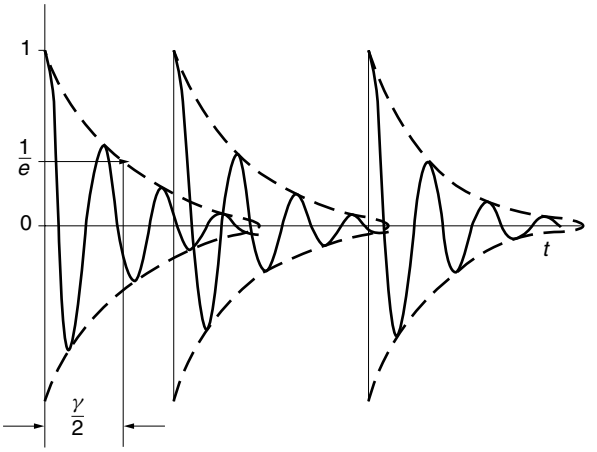


Figure 13.4 Burst of emissions at the times of transitions.

The spectral distribution is obtained from the Fourier transform of Eq. (13.1):

$$E(\nu) = \frac{E_0}{2}\left(\frac{1}{\gamma/2 + j2\pi(\nu - \nu_0)} + \frac{1}{\gamma/2 + j2\pi(\nu + \nu_0)}\right) \tag{13.2}$$

Only the spectrum near ν_0 is considered and the light intensity spectrum is

$$I(\nu) = E(\nu)E^*(\nu) = \frac{I_0}{(\gamma/2)^2 + [2\pi(\nu - \nu_0)]^2} \tag{13.3}$$

The normalized line function $g(\nu)$ will be found by finding I_0 such that

$$\int_{-\infty}^{\infty} \frac{I_0}{(\gamma/2)^2 + [2\pi(\nu - \nu_0)]^2}\, d\nu = 1$$

Using the finite integral formula

$$\int_{-\infty}^{\infty} \frac{a}{x^2 + a^2}\, dx = 2\pi$$

the lineshape function that is normalized as

$$\int_{-\infty}^{\infty} g(\nu)\, d\nu = 1 \tag{13.4}$$

is obtained as

$$g(\nu) = \frac{\gamma/2}{(\gamma/2)^2 + [2\pi(\nu - \nu_0)]^2} \tag{13.5}$$

The lineshape function in Eq. (13.5) is called the *Lorentz lineshape function*. This line broadening is due to the finiteness of the duration of each emission and it takes place even when the medium is homogeneous. There is another kind of broadening

that is due to the inhomogeneity of the medium. The former is called homogeneous broadening, and the latter, inhomogeneous broadening. The combination of these two broadening effects determines the lineshape. The overall line broadening will be designated by $g_t(\nu)$ here.

The function $g_t(\nu)$ is sometimes approximated as a constant value g_t and is normalized in a similar manner as Eq. (13.5):

$$g_t(\nu) \doteqdot g_t$$

with

$$g_t \Delta\nu_t = 1$$

and

$$g_t(\nu) \doteqdot \frac{1}{\Delta\nu_t} \tag{13.6}$$

The line shape functions are shown in Fig. 13.5.

Having completed a brief description of lineshape functions, the rate of transitions of the carriers will now be calculated. Let N_2 be the number of carriers per unit volume in energy level E_2, and N_1 be the same in energy level E_1. The number of carriers making the downward transition per second per unit volume due to spontaneous emission is

$$-\left.\frac{dN_2}{dt}\right|_{\text{spon}} = AN_2 \tag{13.7}$$

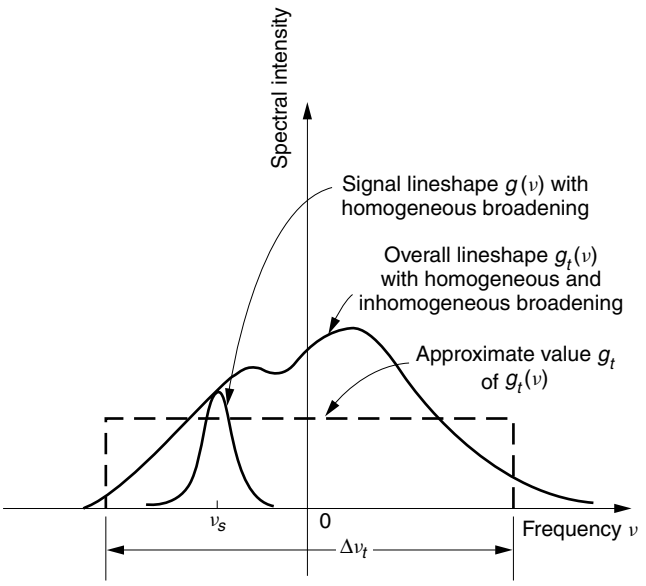


Figure 13.5 Relationships between $g(\nu)$, $g_t(\nu)$, and g_t as a function of frequency.

Let us prove

$$-\frac{dN}{dt} = \frac{N}{\tau}$$

An analogy is made to the calculation of the number of apples falling per day from a tree with a large number N of apples. The lifetime of the apples on the tree is 30 days. The age group of the apples is assumed to be equally distributed. There are $N/30$ apples in each age group and there are $N/30$ apples that are 30 days old and are ready to fall every day.

Thus, the number of apples that fall from the tree each day is

$$-\frac{dN}{dt} = \frac{N}{\tau}$$

where A is Einstein's A coefficient. The spread in frequency of the spontaneous emission is wide and considered constant with frequency. A is a measurable quantity equal to the inverse of the lifetime τ_{21} of the carriers in the E_2 level:

$$A = \frac{1}{\tau_{21}} = \frac{1}{\tau_{\text{spon}}} \tag{13.8}$$

The rate of change in carrier population N_2 per unit time only due to stimulated processes is the difference between the number of carriers that undergo absorption by making the upward transition from E_1 to E_2 and the number of carriers that undergo stimulated emission by taking the downward transition from E_2 to E_1. The induced transition probabilities are identical for both upward and downward transitions and are represented by one probability W_s. The net number of carriers making the downward transition per second per unit volume due to stimulated emission is

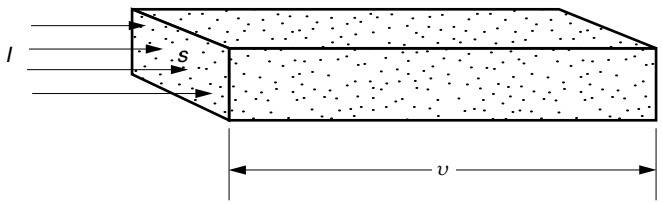
$$-\left.\frac{dN_2}{dt}\right|_{\text{stim}} = W_s(N_2 - N_1) \tag{13.9}$$

The value of W_s is proportional to the light energy density E_d of the stimulating light,

$$W_s = BE_d \tag{13.10}$$

Consider a volume sv defined by light passing through a cross section s m^2 at velocity v in 1 second. The light energy contained in this volume is Is, where I is the light intensity in W/m^2. The light energy density E_d is then

$$E_d = \frac{Is}{sv} = \frac{I}{v} \tag{13.11}$$



where the proportionality constant is known as Einstein's B coefficient. E_d is related to the light intensity I by $E_d = (I/v)$.

By including the spectral lineshape function $g_t(\nu)$, which is a combination of both homogeneous and inhomogeneous line broadenings, W_s can be rewritten as [3,4]

$$W_s(\nu) = B\frac{g_t(\nu)}{v}I \tag{13.12}$$

Next, B will be rewritten in terms of A.

Einstein's A and B coefficients are related to each other. Einstein derived the relationship between A and B from the equilibrium condition of blackbody radiation [3] as

$$\frac{A}{B} = \frac{8\pi n_1^3}{c^3}h\nu^3 \tag{13.13}$$

where n_1 is the index of refraction of a blackbody radiator and c is the speed of light.

Inserting Eqs. (13.8) and (13.13) into Eq. (13.12), W_s becomes

$$W_s(\nu) = \sigma_s\frac{I}{h\nu} \tag{13.14}$$

where

$$\sigma_s = \frac{\lambda^2 g_t(\nu)}{8\pi n_1^2 \tau_{\text{spon}}} \tag{13.15}$$

The quantity σ_s is called the stimulated emission cross section.

Let us further rewrite Eq. (13.13) noting the fact that there are $(8\pi n_1^3\nu^2/c^3)\Delta V \cdot \Delta\nu$ modes in the frequency range between $\nu + \Delta\nu$ and ν in the volume ΔV of a blackbody radiator (see Appendix B of Volume II). With this fact, Eq. (13.13) is described as

$$\frac{A}{B} = m(\nu)h\nu \tag{13.16}$$

where

$$m(\nu) = \frac{8\pi n_1^3}{c^3}\nu^2 = \begin{pmatrix}\text{Number of modes per}\\ \text{unit volume per}\\ \text{unit frequency}\end{pmatrix} \tag{13.17}$$

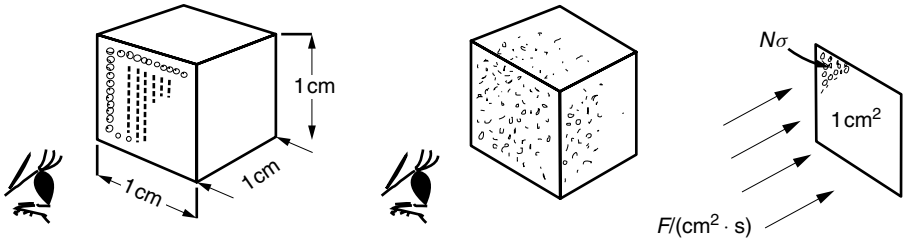
Inserting Eq. (13.17) into Eq. (13.15) or inserting Eq. (13.16) into Eq. (13.12) and comparing with Eq. (13.14) gives

$$\sigma_s = \frac{Ag_t(\nu)}{m(\nu)v} \tag{13.18}$$

Equation (13.14) can be interpreted as follows. The number of photons passing through unit area per second is the photon flux density F. Since $I/h\nu$ is the photon flux density F, Eq. (13.9) becomes (see boxed note)

$$\frac{dN_2}{dt} = \sigma_s F(N_2 - N_1)$$

Let us say there are $N = 1000$ carriers in a 1-cm^3 cube. If the carriers are packed in an orderly fashion and viewed straight on, then only the 100 carriers that are in the first face can be seen. But if the carriers are packed in a disorderly manner, then carriers throughout the cube can be seen, and at maximum, all 1000 carriers are seen.



When orderly 100 out of 1000 are seen

When disorderly, at maximum 1000 are seen

Number of photons intercepted is $F\dfrac{N\sigma}{1\,\text{cm}^2}$

In the maximum disorderly case, the total cross-sectional area that is seen is 1000 σ or $N\sigma$, where σ cm^2 is the cross section of each carrier. Now the number of incident stimulating photons passing through the 1-cm^2 cross-section per second is

$$F = \frac{I}{h\nu}$$

where I is in units of W/cm^2. The ratio of the number of photons that are intercepted by the carriers to that of the incident photons is the ratio of σN to 1 cm^2. The total number of carriers that are stimulated therefore is

$$\frac{dN}{dt} = \left(\frac{\sigma N}{1\ \text{cm}^2}\right) F = \sigma N \frac{I}{h\nu}$$

Including both emission and absorption carriers,

$$\frac{dN_2}{dt} = (N_2 - N_1)\sigma_s \frac{I}{h\nu}$$

The value of σ_s depends on the host material and the wavelength, as shown in Fig. 13.6. The peaks are in the vicinity of 1540 nm, and σ_s tapers off in 1540 ± 40 nm [5].

Now, let us consider the specific case of an erbium-doped fiber amplifier (EDFA). Imagine a fictitious open sided box, such as shown in Fig. 13.7, moving with the photons through a stationary sea of Er^{3+} ions without friction. The increase in the light energy per unit time inside the box is calculated. For each photon having energy $h\nu$, the total contributions from both stimulated and spontaneous emissions are, from Eqs. (13.7) and (13.9),

$$\frac{dE_d}{dt} = \left((N_2 - N_1)\sigma_s \frac{I}{h\nu} + AN_2\right) h\nu \tag{13.19}$$

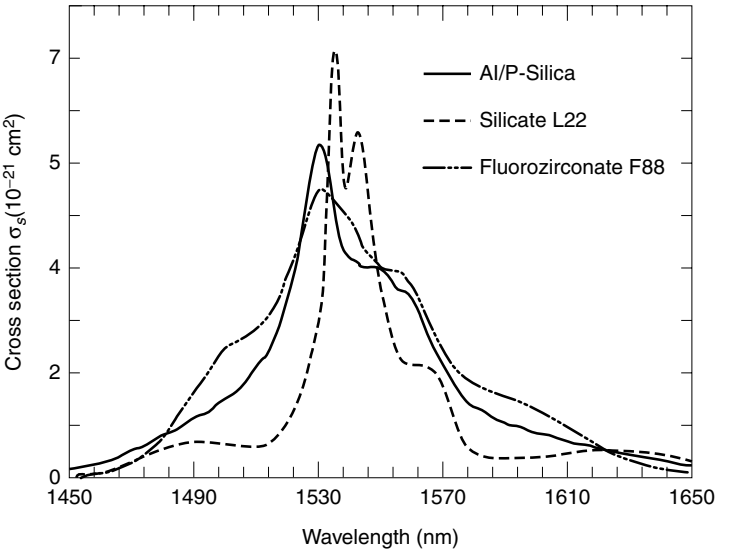


Figure 13.6 Signal cross section σ_s of Er^{3+} hosted in various kinds of glass, as a function of wavelength. (After W. J. Miniscalco [5].)

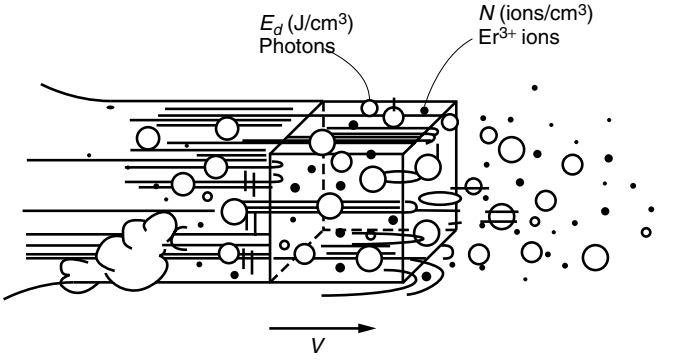


Figure 13.7 Fictitious moving pill box in a fiber amplifier.

From Eqs. (13.11),

$$\frac{dE_d}{dt} = \frac{dI}{v\,dt} = \frac{dI}{dz}$$

and Eq. (13.19) becomes

$$\frac{dI}{dz} = (N_2 - N_1)\sigma_s I + AN_2 h\nu \tag{13.20}$$

The solution of the differential equation is found by letting

$$\begin{aligned} g &= (N_2 - N_1)\sigma_s \\ h &= AN_2 h\nu \end{aligned} \tag{13.21}$$

Equation (13.20) becomes

$$\frac{dI}{dz} = gI + h \tag{13.22}$$

and

$$\int \frac{dI}{gI + h} = \int dz \tag{13.23}$$

Integrating Eq. (13.23) gives

$$\ln(gI + h) = gz + C$$

Solving for I yields

$$I = C'e^{gz} - \frac{h}{g} \tag{13.24}$$

With the boundary condition at $z = 0$,

$$I = I_s \tag{13.25}$$

the solution becomes

$$I = I_s e^{gz} + \frac{h}{g}(e^{gz} - 1) \tag{13.26}$$

Inserting Eq. (13.21) into (13.26) gives the output light intensity I at $z = L$:

$$I = GI_s + (G - 1)\frac{N_2}{N_2 - N_1}\frac{Ah\nu}{\sigma_s} \tag{13.27}$$

where

$$G = e^{gL} \tag{13.28}$$

The first step toward determining the value of σ_s for an erbium-doped fiber is to find the value of $m(\nu)$ specifically for a fiber. Let us consider the entire fiber as one big cavity with length L, a cross section s, and reflectors at each end. The longitudinal modes that are generated by two oppositely traveling waves in the cavity have to satisfy the condition of

$$\frac{\lambda_p}{2} p = L \tag{13.29}$$

In the fiber amplifier there is no reflection and no longitudinal modes, but in order to match the classical approach, we assume there are modes due to the reflected and forward waves. In the end, only one-half of the modes will be used because of the absence of the reflected wave in the fiber amplifier, where λ_p is the wavelength of the pth longitudinal mode. In terms of the frequency f_p, Eq. (13.29) is written as

$$f_p = \frac{v}{2L} p \tag{13.30}$$

where v is the phase velocity in the core. That of the $(p+1)$st mode is

$$f_{p+1} = \frac{v}{2L}(p+1) \tag{13.31}$$

The separation between adjacent modes is

$$f_{p+1} - f_p = \frac{v}{2L} \tag{13.32}$$

The number of modes per unit frequency in this cavity is

$$\frac{1}{f_{p+1} - f_p} = \frac{2L}{v} \tag{13.33}$$

Only the modes that carry the energy forward will be considered. Also, the number m_t of transverse modes has to be accounted for. For a circularly polarized wave, which has two orthogonal modes, m_t is 2. For a linearly polarized mode, m_t is 1. The number $m(\nu)$ per unit volume is obtained by further dividing by sL:

$$m(\nu) = m_t \frac{L}{v}\frac{1}{Ls} = m_t \frac{1}{vs} \tag{13.34}$$

Now, the transition cross section σ_s is found by combining Eqs. (13.18) and (13.34):

$$\sigma_s = \frac{Asg_t(\nu)}{m_t} \tag{13.35}$$

Inserting Eqs. (13.6) and (13.35) into Eq. (13.27) and integrating across the cross section of the fiber gives

$$P = GP_s + (G-1)n_{\text{spon}}\, m_t h\nu\, \Delta\nu_t \tag{13.36}$$

P and P_s are the output and input light powers.

$$n_{\text{spon}} = \frac{N_2}{N_2 - N_1} \tag{13.37}$$

The first term of Eq. (13.36) is the amplified signal power while the second term is the *amplified spontaneous emission* (ASE) noise. The factor n_{spon} is called the *population inversion factor*. The gain factor g of the amplifier can be raised by increasing the population difference $N_2 - N_1$. The ASE noise can be lowered by reducing n_{spon}, but n_{spon} cannot be made zero by making $N_2 = 0$ because the amplification also disappears. A more appropriate minimum value of n_{spon} is unity when $N_1 = 0$ and $N_2 = N$ with $N = N_1 + N_2$. This occurs when all the carriers are emptied out of the energy level E_1. The value of n_{spon} cannot be smaller than unity even for the lossless case. As seen from Eq. (13.36), this noise is amplified by $(G-1)$ with almost the same gain as that for the signal. It is essential for an optical amplifier to be equipped with an optical filter to suppress the ASE noise. The equivalent circuit derived from Eq. (13.36) is shown in Fig. 13.8.

A similar treatment of the gain of the SLA will be given together with lasers in the next chapter.

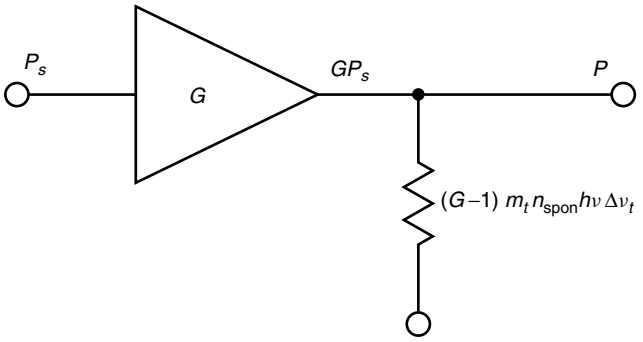


Figure 13.8 Equivalent circuit of an optical amplifier.

13.5 RATE EQUATIONS FOR THE THREE-LEVEL MODEL OF Er^{3+}

The previous section looked at the expression for the amplifier gain, which was derived from the rate equations for the two levels E_2 and E_1 involved in the population inversion. This section explores the solutions of the rate equations for all three levels of a three-level model [6]. The object of this analysis is to reveal the behavior of the threshold and saturation of the gain with respect to the pump light. The example chosen is the three-level Er^{3+} doped fiber amplifier, which is shown in Fig. 13.9. This figure indicates the following quantities: the number of carriers N_i per unit volume in the energy level E_i; the spontaneous transition lifetime τ_{ij}, which is the inverse of the rate of the transition from the ith energy level to the jth energy level (where $i > j$); the stimulated transition probability W_p between E_3 and E_1; and the stimulated transition probability W_s between E_2 and E_1. The quantity W_p is defined similar to Eq. (13.14) as

$$W_p(\nu_p) = \sigma_p \frac{I_p}{h\nu_p} \tag{13.38}$$

The rates of change in the populations in the three levels are

$$\frac{dN_3}{dt} = W_p(N_1 - N_3) - \frac{N_3}{\tau_3} \tag{13.39}$$

$$\frac{dN_2}{dt} = \frac{N_3}{\tau_{32}} - \frac{N_2}{\tau_{21}} - W_s(N_2 - N_1) \tag{13.40}$$

$$\frac{dN_1}{dt} = -W_p(N_1 - N_3) + \frac{N_2}{\tau_{21}} + W_s(N_2 - N_1) \tag{13.41}$$

where

$$\frac{1}{\tau_3} = \frac{1}{\tau_{32}} + \frac{1}{\tau_{31}} \tag{13.42}$$

The arguments (ν_s) and (ν_p) are suppressed.

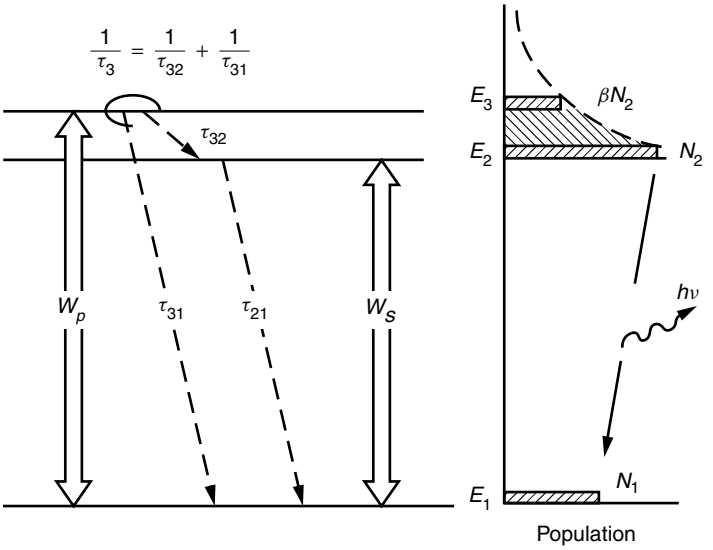


Figure 13.9 Three-level model for Er^{3+} doped fiber amplifier.

The simultaneous rate equations will be solved for the steady-state solution with

$$\frac{d}{dt} = 0 \tag{13.43}$$

The population ratio, the population difference, the gain, and the saturation gain of the amplifier will be found.

13.5.1 Normalized Steady-State Population Difference

The population difference between the E_1 and E_2 levels, which determines the gain of the amplifier, will be found. Since

$$\tau_{31} \gg \tau_{32} \tag{13.44}$$

and with Eqs. (13.42) to (13.44), Eq. (13.39) becomes

$$W_p(N_1 - N_3) = \frac{N_3}{\tau_{32}} \tag{13.45}$$

Putting Eq. (13.45) into (13.40) gives

$$0 = W_p(N_1 - N_3) - \frac{N_2}{\tau_{21}} - W_s(N_2 - N_1) \tag{13.46}$$

The E_2 and E_3 levels of the Er^{3+} doped fiber amplifier pumped by 1.48-μm pump light are very closely spaced, and because of the fast relaxation process, the population ratio between these two levels quickly reaches the Boltzmann population ratio [4]:

$$\beta = \frac{N_3}{N_2} = e^{-(\Delta E/kT)} \tag{13.47}$$

At room temperature, $\beta = 0.38$. Putting Eq. (13.47) into (13.46) gives

$$\frac{N_2}{N_1} = \frac{W_p + W_s}{W_p\beta + 1/\tau_{21} + W_s} \tag{13.48}$$

N is defined as

$$N = N_1 + N_2 \tag{13.49}$$

Let the numerator and denominator of Eq. (13.48) be A and B,* respectively. N becomes

$$N = N_1 + \frac{A}{B}N_1 \tag{13.50}$$

and hence

$$N_1 = \frac{B}{A+B}N \tag{13.51}$$

$$N_2 = \frac{A}{A+B}N \tag{13.52}$$

Equations (13.51) and (13.52) can be rewritten as

$$\frac{N_2 - N_1}{N} = \frac{A-B}{A+B} \tag{13.53}$$

Putting the numerator and denominator of Eq. (13.48) into this equation and letting $\tau_{21} = \tau$ gives

$$\frac{N_2 - N_1}{N} = \frac{(1-\beta)W_p\tau - 1}{(1+\beta)W_p\tau + 2W_s\tau + 1} \tag{13.54}$$

$W_p\tau$ is called the normalized pumping rate.

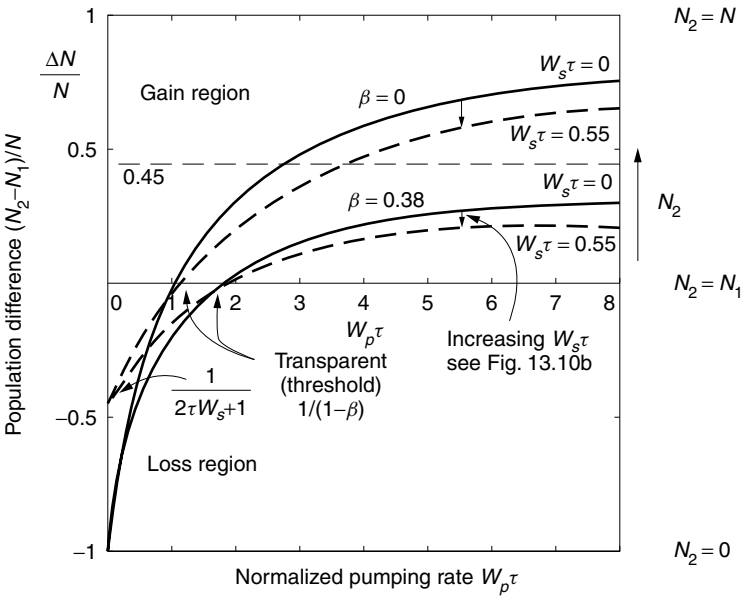
Recall from Eqs. (13.14) and (13.38) that $W_s\tau$ and $W_p\tau$ are quantities representing the signal and pump light intensities, respectively. The quantity $N_2 - N_1$ represents the gain factor from Eq. (13.21). Thus, Eq. (13.54) is an important equation that relates signal and pump powers to the gain of the amplifier. In Fig. 13.10a, $(N_2 - N_1)/N$ is plotted as a function of the pump power $W_p\tau$ with the signal power $W_s\tau$ and β as parameters. In Fig. 13.10b, the corresponding experimental results [7] are shown.

Example 13.1 Is the following logic true or false?

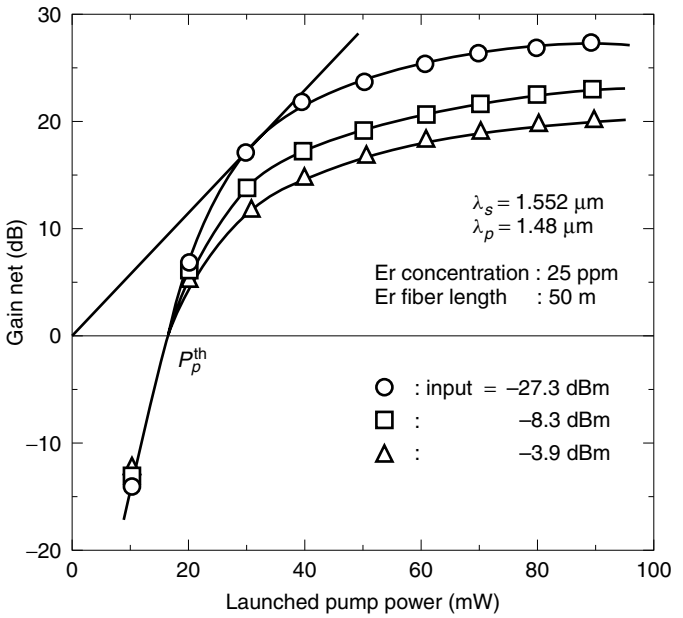
As shown in the energy level diagram in Fig. 13.3, there exists only the stimulated transition in the upward transition $1 \rightarrow 2$, whereas there exist both stimulated and spontaneous transitions in the downward transition $2 \rightarrow 1$. As a result, the overall probability of a downward transition $2 \rightarrow 1$ is always larger than an upward transition, so that a net downward transition of the carriers always exists.

Solution The conclusion that there is always a net downward transition of the carriers is false. But for Boltzmann's theorem, it could have been true. The population of the

*A and B are *not* Einstein's coefficients.



(a)



(b)

Figure 13.10 Population difference versus pump power with signal power as a parameter. (a) Population difference versus normalized pumping rate with signal power and β as parameters. (b) Signal gain characteristics versus pump power. (After K. Nakagawa et al. [7]).

carriers was overlooked. According to Eq. (13.47), the population N_2 of the upper state is smaller than N_1 of the lower state at equilibrium. □

13.5.2 Gain of the Amplifier

From the expression for the gain factor g of the optical amplifier in Eqs. (13.21) and (13.54), the criteria for whether or not the amplifier experiences gain are

$$\frac{N_2 - N_1}{N} \begin{cases} > 0 & \text{gain} \\ = 0 & \text{transparent} \\ < 0 & \text{loss} \end{cases} \tag{13.55}$$

When a medium starts to experience gain, the medium is said to become active. Equation (13.54) means that the medium does not become active until it reaches the threshold value of

$$W_p^{\text{th}}\tau = \frac{1}{1-\beta} \tag{13.56}$$

This threshold occurs at the same value regardless of the signal power level. The experimental curves in Fig. 13.10b clearly indicate this fact. Equation (13.56) with $\beta = 0$ reduces to

$$W_p^{\text{th}} = \frac{1}{\tau}$$

Thus, the threshold condition is interpreted as the case when the pump power is just enough to sustain the carriers for spontaneous emission.

From Eq. (13.56), we see that the longer the lifetime τ of the metastable state of erbium is, the lower the threshold pump power becomes. This is an additional merit of the long lifetime besides the earlier mentioned reduced spontaneous emission noise.

As seen from Eq. (13.54), the population difference is at its maximum when

$$2W_s\tau = 0 \tag{13.57}$$

Let us define the maximum population difference $\Delta N_{\max}$ from Eqs. (13.54) and (13.57):

$$\frac{\Delta N_{\max}}{N} = \frac{(1-\beta)W_p\tau - 1}{(1+\beta)W_p\tau + 1} \tag{13.58}$$

Inserting Eq. (13.58) into (13.54) gives

$$\frac{N_2 - N_1}{N} = \frac{\Delta N_{\max}}{N} \frac{(1+\beta)W_p\tau + 1}{(1+\beta)W_p\tau + 2W_s\tau + 1}$$

or

$$\frac{N_2 - N_1}{N} = \frac{\Delta N_{\max}}{N} \frac{1}{1 + \dfrac{2W_s\tau}{(1+\beta)W_p\tau + 1}} \tag{13.59}$$

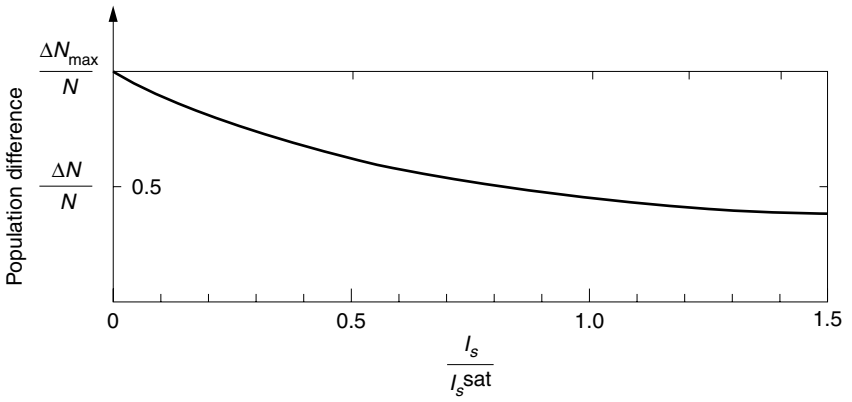


Figure 13.11 Population difference versus normalized signal power density.

The value of $W_s\tau$ (signal power) that reduces $N_2 - N_1$ to one-half of ΔN_{max} is designated as W_s^{sat}.

$$\frac{N_2 - N_1}{N} = \frac{\Delta N_{max}}{N} \frac{1}{1 + W_s/W_s^{sat}} \tag{13.60}$$

where

$$W_s^{sat} = \tfrac{1}{2}\left((1+\beta)W_p + \frac{1}{\tau}\right) \tag{13.61}$$

With Eq. (13.56), Eq. (13.61) becomes

$$W_s^{sat} = \frac{1}{2\tau}\left(1 + \frac{1+\beta}{1-\beta}\frac{W_p}{W_p^{th}}\right) \tag{13.62}$$

Recalling Eq. (13.14), Eq. (13.60) can be rewritten as

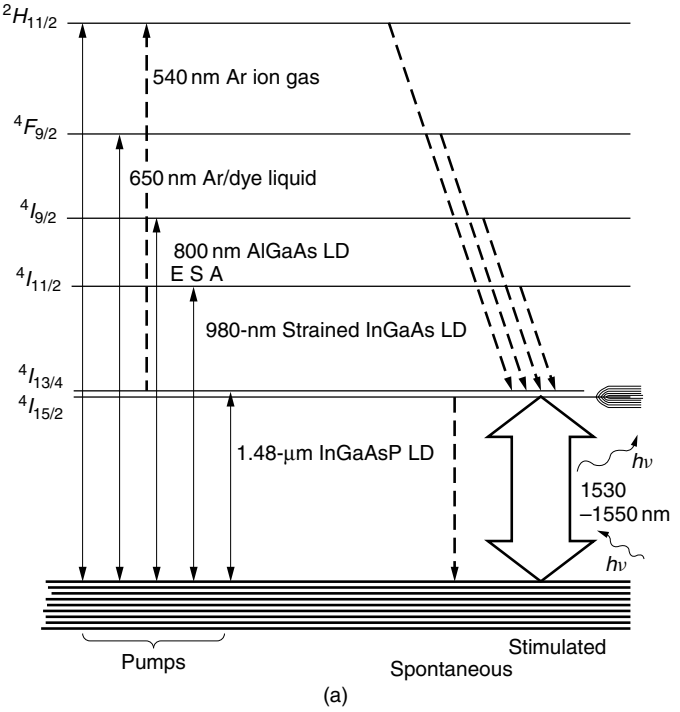
$$\frac{N_2 - N_1}{N} = \frac{\Delta N_{max}}{N} \frac{1}{1 + I_s/I_s^{sat}} \tag{13.63}$$

Equation (13.63) is plotted in Fig. 13.11.

The above simple analysis reveals several important conclusions. Recall that the gain factor of the amplifier is proportional to the population difference. An increase in the pump power results in an increase not only in the signal saturation intensity (Eq. (13.62)) but also in the gain of the amplifier (Eq. (13.59)). With an increase in the signal power, the gain of the amplifier gradually decreases and finally reaches zero (Eq. (13.63)).

13.6 PROS AND CONS OF 1.48-μm AND 0.98-μm PUMP LIGHT

Figure 13.12a shows the energy level diagram of Er^{3+}. The types of lasers that may be used as the pump light are indicated in the diagram. Figure 13.12b shows the absorption



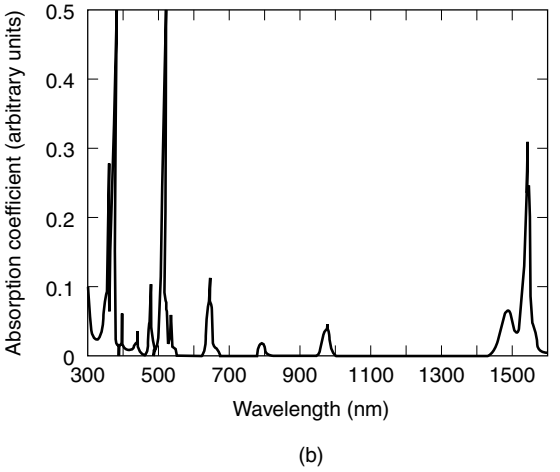


Figure 13.12 Energy level diagram and absorption lines of Er^{3+}. (a) Energy level diagram of Er^{3+}. (b) Absorption spectrum of Er^{3+}-doped silicate glass. The peaks of the strong bands at 380 and 520 nm are 1.5 and 1.0, respectively. (After W. J. Miniscalco [5]).

spectrum of Er^{3+} associated with the transitions shown in Fig. 13.12a. The higher the absorption coefficient is, the larger the absorption cross section σ_p is, and hence the easier to pump. Unfortunately, high absorption lines are in the region of wavelengths shorter than 540 nm where solid state pump lasers are not available. From the viewpoint of reliability, longevity, and compact size, only solid state devices are practical for

optical communication, which means the pump light candidates are 800 nm, 980 nm, and 1.48 μm. The 800-nm pump, however, is inefficient due to a process known as excited state absorption (ESA). Figure 13.13 explains the mechanism of ESA loss. If there is still a higher energy level E_4 such that the spacing between E_4 and E_2 is identical with that between E_3 and E_1, there is a possibility that some of the carriers in level E_2 would be brought up further to the E_4 level by the pump light because $E_3 - E_1 = E_4 - E_2$. The attractive candidates now become only 980-nm and 1.48-μm pump light. Comparisons between these two pump wavelengths will be made using the solutions of the rate equations obtained in Section 13.5.

First, the values of β for these two pump wavelengths are considered. The 1.48-μm pump is special in that the ratio N_3/N_2 of the population density is tightly clamped by the Boltzmann distribution because the energy level of the 1.54-μm signal is so close to that of the 1.48-μm pump. At room temperature, β is

$$\beta = \frac{N_3}{N_2} = e^{-\Delta E/kT} = 0.38 \tag{13.64}$$

In contrast, β is practically zero for 0.98-μm pump light. This creates differences between 1.48-μm and 0.98-μm pumping. As seen from Eq. (13.54), the maximum obtainable population difference for a large normalized pumping rate is

$$\left(\frac{N_2 - N_1}{N}\right)_{\max} = \frac{1-\beta}{1+\beta} \tag{13.65}$$

In the case of 1.48-μm pumping, substituting $\beta = 0.38$ in Eq. (13.65) gives

$$\left(\frac{N_2 - N_1}{N}\right)_{\max,\ 1.48-\mu\text{m pump}} = 0.45 \tag{13.66}$$

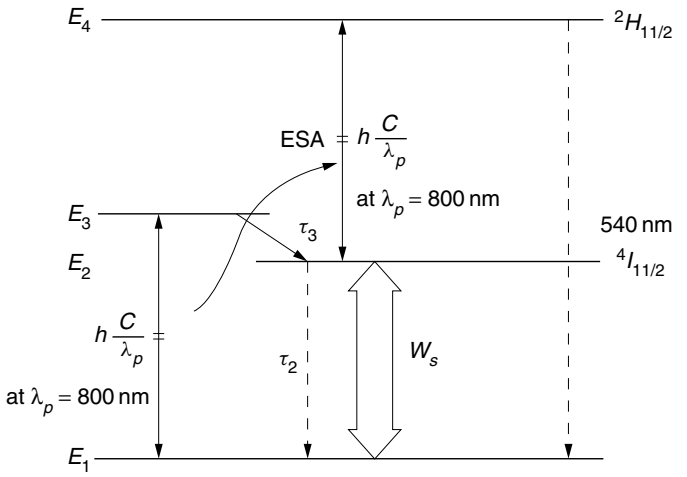


Figure 13.13 Explanation of excited state absorption (ESA) loss: stepwise two-photon absorption.

While for 0.98-μm pumping, Eq. (13.65) evaluates to

$$\left(\frac{N_2 - N_1}{N}\right)_{\text{max, 0.98}-\mu\text{m pump}} = 1 \tag{13.67}$$

Thus, the 0.98-μm pump amplifier has superior gain to its 1.48-μm pump counterpart.

The 0.98-μm pump is also superior to the 1.48-μm pump in terms of the spontaneous noise of the amplifier. As obtained earlier, the amplified spontaneous emission (ASE) noise is proportional to

$$n_{\text{spon}} = \frac{N_2}{N_2 - N_1} \tag{13.68}$$

From Eqs. (13.51) and (13.52), N_{spon} is rewritten as

$$\frac{N_2}{N_2 - N_1} = \frac{A}{A - B} \tag{13.69}$$

and from Eq. (13.48), n_{spon} becomes

$$n_{\text{spon}} = \frac{W_p + W_s}{W_p(1 - \beta) - 1/\tau} \tag{13.70}$$

Increasing the pump power reduces n_{spon} and n_{spon} approaches the limit

$$n_{\text{spon}} = \frac{1}{1 - \beta} \tag{13.71}$$

The 0.98-μm pump can ultimately reduce n_{spon} to 1, but the 1.48-μm pump can only reduce n_{spon} to 1.61.

When comparing the pumping efficiency, which is the gain in dB obtained per milliwatt of pump power, the 0.98-μm pump is also superior to the 1.48-μm pump. The threshold pump power is given by Eq. (13.56). For 1.48 μm, the threshold is defined as $W_p^{\text{th}}\tau = 1.61$, while the same quantity for the 0.98-μm pump is $W_p^{\text{th}}\tau = 1.0$. This difference is related to a measured difference in the energy extraction efficiency from the pump laser. The gain of the amplifier per unit pump power for a 0.98-μm pump laser is 10 dB/mW, while that for a 1.48-μm pump laser is 5 dB/mW.

As for the saturation signal power intensity, Eq. (13.61) contains a factor $(1 + \beta)$. For the same pump power W_p, the saturation signal power intensity W_s^{sat} is higher for the 1.48-μm pump. Thus, the 1.48-μm pump is better than the 0.98-μm pump in this respect.

One of the difficulties associated with the 0.98-μm pump is that the linewidth of the pump transition is narrower than the linewidth of the 1.48-μm transition. The 0.98-μm pump laser diode (LD) needs tight control of its wavelength of emission.

In summary, the 0.98-μm pump is superior with respect to the gain per unit pump power, the amount of ASE noise, and the threshold pump power, while the 1.48-μm pump is superior with respect to saturation signal power and the tolerance of the wavelength of the pump light.

These conclusions are summarized in Table 13.2.

Table 13.2 Comparison of EDFAs pumped by 1.48-μm and 0.98-μm pump light

Characteristic	1.48-μm Pump	0.98-μm Pump
Pump laser material	InGaAs/InP	Strained InGaAs
Typical output power from the pump laser	20–100 mW	10–20 mW
Amplifier differential gain per mW of pump light	5 dB/mW	10 dB/mW
Saturation signal power	20 dBm	5 dBm
Pump transition linewidth	20 nm	2.5 nm
Noise figure (NF)	5 dB	3 dB

13.7 APPROXIMATE SOLUTIONS OF THE TIME-DEPENDENT RATE EQUATIONS

Whether or not a short-duration pulse can be amplified depends on the behavior of the amplifier gain over short time spans. In other words, the time-dependent solutions for N_2 and N_1 have to be found. The time-dependent rate equations will be solved.

The lifetime τ_3 is less than 1 ns and N_3 is large, which means $-N_3/\tau_3$ is a large negative number. This means that N_3 reaches the steady state in less than 1 ns, and it is usually safe to assume after 1 ns that

$$\frac{dN_3}{dt} = 0 \tag{13.72}$$

This assumption means that no significant change in N_3 takes place after 1 ns. This assumption remarkably simplifies the solution, yet gives answers very close to the answers obtained without the assumption [8]. With Eq. (13.72), the time-dependent rate equations — Eqs. (13.40), (13.41), and (13.45) — become

$$\dot{N}_1 = -W_p(N_1 - \beta N_2) + \frac{N_2}{\tau} + W_s(N_2 - N_1) \tag{13.73}$$

$$\dot{N}_2 = W_p(N_1 - \beta N_2) - \frac{N_2}{\tau} - W_s(N_2 - N_1) \tag{13.74}$$

Rewriting Eqs. (13.73) and (13.74) gives

$$\dot{N}_1 = -aN_1 + bN_2 \tag{13.75}$$

$$\dot{N}_2 = aN_1 - bN_2 \tag{13.76}$$

where

$$a = W_p + W_s \quad \text{and} \quad b = \beta W_p + W_s + \frac{1}{\tau} \tag{13.77}$$

The dot means the derivative with respect to time t. The derivative of Eq. (13.76) and the relationship $\dot{N}_1 = -\dot{N}_2$ obtained from Eqs. (13.75) and (13.76) gives

$$\ddot{N}_2 + (a + b)\dot{N}_2 = 0 \tag{13.78}$$

The same form of differential equation is obtained for N_1.

Let's assume the solution of N_2 is

$$N_2 = C_1 e^{\gamma t} + C_2 \tag{13.79}$$

The value of γ is found by putting Eq. (13.79) back into Eq. (13.78) and differentiating:

$$\gamma^2 + (a+b)\gamma = 0 \tag{13.80}$$

The two possible values of γ are obtained as

$$\gamma(\gamma + a + b) = 0 \tag{13.81}$$

$$\gamma = 0 \quad \text{and} \quad \gamma = -(a+b) \tag{13.82}$$

The general solution for $N_2(t)$ is

$$N_2(t) = C_1 e^{-(a+b)t} + C_2 \tag{13.83}$$

With the initial values of $N_2(0)$ and $\dot{N}_2(0)$, C_1 and C_2 are found. The result is

$$N_2(t) = N_2(0) + \frac{\dot{N}_2(0)}{a+b}\left(1 - e^{-t/\tau_{\text{eff}}}\right) \tag{13.84}$$

where

$$\tau_{\text{eff}} = \frac{1}{a+b} \tag{13.85}$$

Invoking the conservation of the number of carriers $N = N_1(t) + N_2(t)$, the time-dependent population difference becomes

$$\frac{N_2(t) - N_1(t)}{N} = 2\frac{N_2(t)}{N} - 1 \tag{13.86}$$

which leads to a time-dependent expression of the gain of the amplifier from Eq. (13.21).

Inserting Eq. (13.77) into (13.85), the effective time constant is obtained as

$$\tau_{\text{eff}} = \frac{\tau}{(1+\beta)W_p\tau + 2W_s\tau + 1} \tag{13.87}$$

Using τ in Eqs. (13.56) and (13.62), Eq. (13.87) can be rewritten as

$$\tau_{\text{eff}} = \frac{\tau}{\dfrac{W_p}{W_p^{\text{th}}}\left(\dfrac{1+\beta}{1-\beta}\right) + \dfrac{W_s}{W_s^{\text{sat}}}\left(1 + \dfrac{1+\beta}{1-\beta}\dfrac{W_p}{W_p^{\text{th}}}\right) + 1} \tag{13.88}$$

The gain factor of the amplifier is proportional to the population difference. Equations (13.83) through (13.88) tell us how quickly the gain of the amplifier reacts to a change of conditions, as, for example, a sudden decrease in $N_2(t)$ due to a sharp high-intensity input pulse.

The lifetime τ of Er^{3+} is quite long and $\tau = 10$ ms. The derived expression of τ_{eff} in Eq. (13.88) shows that τ_{eff} is a slow function of W_s and W_p. Even with the extreme case of $W_s/W_s^{\text{sat}} = W_p/W_p^{\text{th}} = 5$ (= 7 dB), the effective lifetime τ_{eff}

The results without the assumption of $dN_3/dt = 0$ can be found in the literature [8], and the solution is of the form of $N_2(t) = N_0(e^{-t/\tau_{\text{eff}}} + \epsilon e^{-t/\tau_3}) + C$, where $\epsilon = \tau_3/\tau \approx 10^{-7}$ and N_0 and C are constants. The contribution of the term with τ_3 is small and insignificant.

is 14 μs to 0.28 ms, depending on β, and is reasonably slow. This means that as long as the rise time of the input pulse is faster than τ_{eff}, no change in the gain of the amplifier, which is proportional to Eq. (13.86) takes place, and the pulse can be amplified without distortion. Amplification without distortion is an important feature of the erbium-doped fiber amplifier. Provided that the signal rates are higher than $1/(0.28 \times 10^{-3}) = 35.7$ kb/s for RZ coding (see Section 16.2.9), no distortion is generated. This is true even when the amplitude of the signal is beyond saturation because the gain of the amplifier cannot change faster than 0.28 ms. Any signal with a bit rate lower than 35.7 kb/s, however, experiences a change in the gain and, hence, distortion of the signal. Distortionless amplification at saturation power levels is utilized to achieve a stable operating point of the amplifier. Another advantage of setting the operating point at several decibels beyond the point of the saturation power level is that even when the input signal power decreases abruptly for a short time due to unknown causes in the system, the change in the amplifier's gain is minimum. This is also true when the power level increases unexpectedly.

Normally, in the case of a multichannel system, the operating power level is set below gain saturation in order to avoid crosstalk between the channels. Consider two independent channels with slightly different wavelengths that are combined and amplified by the same optical amplifier. Suppose that channel A is the signal channel, and channel B is just providing the operating point. If channel B does not change with time, then channel A is amplified at constant gain and no distortion results. However, if channel B is also changing with time, the operating point of channel A changes with time, and the gain of channel A changes as the operating point fluctuates. The output of channel A is influenced by the channel B signal, and crosstalk between channels A and B occurs. In the case of the EDFA, if the movement of the bias point is fast enough, the gain does not change, and even from the viewpoint of crosstalk between channels, the EDFA has advantages over the SLA.

Example 13.2 From the following measured results, determine the signal stimulated emission cross section σ_s of Er^{3+}.

$\lambda = 1.55$ μm
$\Delta\lambda_t = 50$ nm
$n_1 = 1.46$
$\tau = 10$ ms

The unfiltered ASE noise spectrum was used to measure $\Delta\lambda_t$, which is the full-width half-power point of the ASE spectrum.

Solution From Eq. (13.15)

$$\sigma_s = \frac{\lambda^2 g_t(\nu)}{8\pi n_1^2 \tau_{\text{spon}}}$$

First, $g_t(\nu)$ Will be found.

$$\Delta\nu = 3 \times 10^8 \left(\frac{1}{1.525} - \frac{1}{1.575}\right) \times 10^6$$
$$= 6.2 \times 10^{12} \text{ Hz}$$
$$g_t(\nu) \simeq \frac{1}{\Delta\nu} = 1.6 \times 10^{-13} \text{ Hz}^{-1}$$
$$\sigma_s = \frac{(1.55 \times 10^{-6})^2 (1.6 \times 10^{-13})}{8\pi (1.46)^2 (10^{-2})}$$
$$= 7.2 \times 10^{-25} \text{ m}^2 = 7.2 \times 10^{-21} \text{ cm}^2 \qquad \square$$

Example 13.3 An EDFA is constructed with a 30-m length of erbium-doped fiber. What is the population difference $N_2 - N_1$ that is required to achieve a 30-dB gain in this amplifier? Use the value of the stimulated emission cross section obtained in Example 13.2.

Solution The gain from Eq. (13.28) is

$$G = e^{gL}$$

where, from Eq. (13.21),

$$g = (N_2 - N_1)\sigma_s$$

The gain (in dB) is

$$G(\text{dB}) = 10\ gL \log e$$
$$g = \frac{G(\text{dB})}{10L \log e}$$
$$= \frac{(30)}{10(30)(0.434)}$$
$$= 0.230 \text{ m}^{-1}$$
$$(N_2 - N_1) = \frac{g}{\sigma_s}$$
$$= \frac{0.230}{7.2 \times 10^{-25}}$$
$$(N_2 - N_1) = 3.2 \times 10^{17} \text{ cm}^{-3} \qquad \square$$

Example 13.4 Calculate the threshold light intensity for a 1.48-μm pumped erbium-doped fiber amplifier with the following parameters:

$\sigma_p = 0.42 \times 10^{-21}$ cm^2
$\tau = 10$ ms

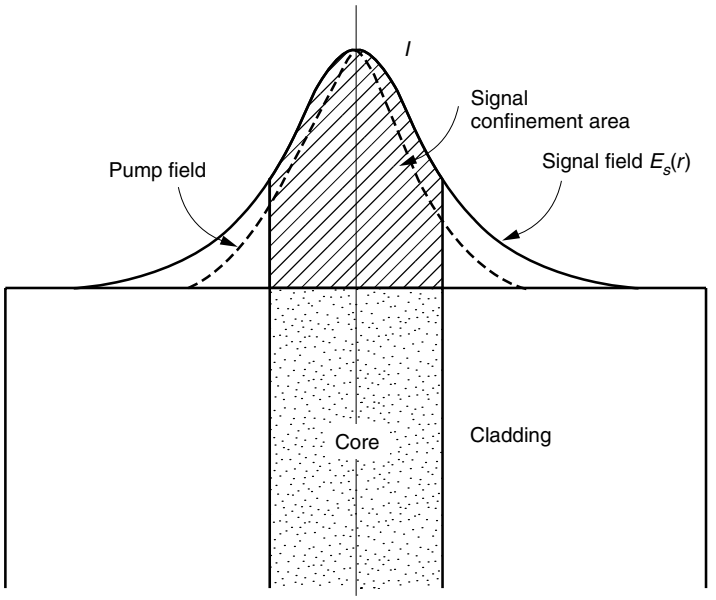


Figure 13.14 Confinement area of light and core.

$h = 6.63 \times 10^{-34}$ J · s
$\beta = 0.38$
$\Gamma = 0.4$
$s = 12.6\ \mu m^2$

Γ is the confinement factor of the pump light distribution with the shape of the fiber core. It is used to quantify the coupling of the pump light as shown in Fig. 13.14.

Compare the calculated results with the experimental curve in Fig. 13.10b.

Solution The first parameter to be calculated is the frequency of the pump:

$$\nu_p = \frac{c}{\lambda} = 2.03 \times 10^{14}\ \text{Hz}$$

As shown in Fig. 13.14, the field $E_s(r)$ is not entirely confined inside the active region (core of the fiber). Only the center portion of the light is inside the active layer and participates in the amplification. The actual gain should be reduced by the *confinement factor*, defined as the ratio of the power inside the core to the entire field:

$$\Gamma_s = \frac{\int_0^a |E_s(r)|^2 r\, dr}{\int_0^\infty |E_s(r)|^2 r\, dr} \tag{13.89}$$

Because of the difference in wavelengths, the confinement factor Γ_s of the signal is not necessarily the same as Γ_p of the pump.

The relationship between the transition probability W_p and the light intensity I_p is, from Eq. (13.38),

$$\begin{aligned} W_p &= \sigma_p \frac{I_p}{h\nu_p} \\ &= (0.42 \times 10^{-25}) \frac{I_p}{(6.63 \times 10^{-34})(2.03 \times 10^{14})} \\ &= 0.312 \times 10^{-6} I_p \end{aligned}$$

At the threshold,

$$W_p^{\text{th}} = 0.312 \times 10^{-6} I_p^{\text{th}}$$

On the other hand, from Eq. (13.56),

$$\begin{aligned} W_p^{\text{th}} &= \frac{1}{(1-\beta)\tau} = \frac{1}{(1-0.38)(10^{-2})} = 161 \text{ s}^{-1} \\ I_p^{\text{th}} &= \frac{161}{0.312 \times 10^{-6}} = 5.16 \times 10^8 \text{ W/m}^2 \\ &= 5.16 \times 10^8 \frac{10^3}{10^{12}} \text{ mW/}\mu\text{m}^2 \\ &= 0.516 \text{ mW/}\mu\text{m}^2 \end{aligned}$$

The confinement factor is $\Gamma = 0.4$, and the cross-sectional area of the fiber is 12.6 μm^2. The threshold pump power is $P_p^{\text{th}} = 0.516 \times (1/0.4) \times 12.6 = 16.3$ mW. This agrees well with experimentally measured values shown in Fig. 13.10b. □

Example 13.5 What is the maximum gain G of a 1.48-μm pumped erbium-doped fiber amplifier when the pump power $P_p = 80$ mW? The maximum gain is the gain near zero signal power. What is the gain when the output light power P is 1 mW? The relevant parameters are:

$\lambda_p = 1.48$ μm
$\lambda_s = 1.55$ μm
$\sigma_p = 0.42 \times 10^{-21}$ cm^2
$\sigma_s = 3.6 \times 10^{-21}$ cm^2
$\tau = 10$ ms
$\Gamma_p = \Gamma_a = 0.4$
$s = 12.6$ μm^2
$L = 35$m
$\beta = 0.38$
$N = 2.0 \times 10^{18}$ cm^{-3}

Ignore the transmission loss.

Solution The gain is

$$G = e^{gL}$$

where

$$g = \sigma_s(N_2 - N_1)$$

The maximum population difference is, from Eq. (13.58),

$$\left(\frac{N_2 - N_1}{N}\right)_{\max} = \frac{(1-\beta)W_p\tau - 1}{(1+\beta)W_p\tau + 1}$$

First, W_p is found from Eq. (13.38):

$$\begin{aligned} W_p &= \sigma_p \frac{1}{h\nu_p}\frac{P_p}{s}\Gamma_p \\ &= 0.42 \times 10^{-25}\frac{8 \times 10^{-2}}{(6.63 \times 10^{-34})(2.03 \times 10^{14})(12.6 \times 10^{-12})}(0.4) \\ &= 793 \text{ s}^{-1} \end{aligned}$$

Next, $(N_2 - N_1)_{\max}$ is calculated.

$$\begin{aligned} \frac{\Delta N_{\max}}{N} &= \frac{(1-0.38)(793)(10^{-2}) - 1}{(1+0.38)(793)(10^{-2}) + 1} = 0.328 \\ (N_2 - N_1)_{\max} &= 2 \times 10^{24} \times 0.328 = 6.56 \times 10^{23} \text{ m}^{-3} \end{aligned}$$

Finally, the maximum gain factor is

$$\begin{aligned} g_{\max} &= \sigma_s(N_2 - N_1) \\ &= (3.6 \times 10^{-25})(6.56 \times 10^{23}) \\ &= 0.24 \text{ m}^{-1} \end{aligned}$$

The maximum gain (in dB) is

$$\begin{aligned} G(\text{dB}) &= 10 \log e^{0.24 \times 35} \\ &= 36.5 \text{ dB} \end{aligned}$$

Next, the gain of the fiber in the region of the nonzero signal light output is calculated. The signal power is not uniform along the fiber but the output signal power $P = 1$ mW is used for the calculation.

$$\frac{N_2 - N_1}{N} = \frac{\Delta N_{\max}}{N}\frac{(1+\beta)W_p\tau + 1}{(1+\beta)W_p\tau + 2W_s\tau + 1}$$

$W_s\tau$ is to be calculated.

$$\begin{aligned} W_s &= \sigma_s \frac{1}{h\nu_s}\frac{P}{s}\Gamma_s \\ &= (3.6\times 10^{-25})\frac{(10^{-3})}{(6.63\times 10^{-34})(1.94\times 10^{14})(12.6\times 10^{-12})}(0.4) \\ &= 88.9\ \text{s}^{-1} \end{aligned}$$

Note that W_s is one order of magnitude smaller than W_p.

$$\begin{aligned} \frac{N_2-N_1}{N} &= \frac{\Delta N_{\max}}{N}\frac{(1+0.38)(793)(10^{-2})+1}{(1+0.38)(793)(10^{-2})+2(88.9)(10^{-2})+1} \\ &= \frac{\Delta N_{\max}}{N}0.87 \end{aligned}$$

$N_2 - N_1$ drops by 13% from the maximum value, and hence,

$$\begin{aligned} g &= 0.24\times 0.87 = 0.21\ \text{m}^{-1} \\ G &= e^{gL} = e^{0.21\times 35} = 1556 \\ G(\text{dB}) &= 10\log e^{0.21\times 35} \\ &= 31.9\ \text{dB} \end{aligned}$$

□

13.8 PUMPING CONFIGURATION

The pump light can be injected either from the front or back end of the erbium-doped fiber. Double-clad fibers are used for high-power amplifiers.

13.8.1 Forward Pumping Versus Backward Pumping

The pump light can be injected so that it propagates in the same direction as the signal light, which is called forward pumping, or it can be injected in the opposite direction, which is called backward pumping. There is also bidirectional pumping, which is a combination of forward and backward pumping. All of these configurations are shown in Fig. 13.15. The reason for the different configurations has to do with the distribution of pump light in the fiber. The pump light is attenuated as it propagates.

In the case of the forward pumping, as shown in Fig. 13.15a, $(N_2 - N_1)/N$ is largest at the input end of the fiber. This means the population inversion factor $n_{\text{spon}} = N_2/(N_2 - N_1)$ is smaller at the input, and hence, the ASE noise is smaller at the input end than at the output end. When cascading two amplifiers with different noise figures, the overall noise performance is better when the lower noise amplifier is placed in front of the noisier one. Thus, forward pumping displays better noise performance than backward pumping.

In the case of backward pumping, as shown in Fig. 13.15b, the saturation signal power is higher than forward pumping. Toward the end of the amplifier fiber, the signal intensity has become high and closer to saturation. The saturation signal intensity, however, is increased with pump power by Eq. (13.62). With backward pumping, the pump power is largest toward the end of the fiber, where the high intensity is needed to

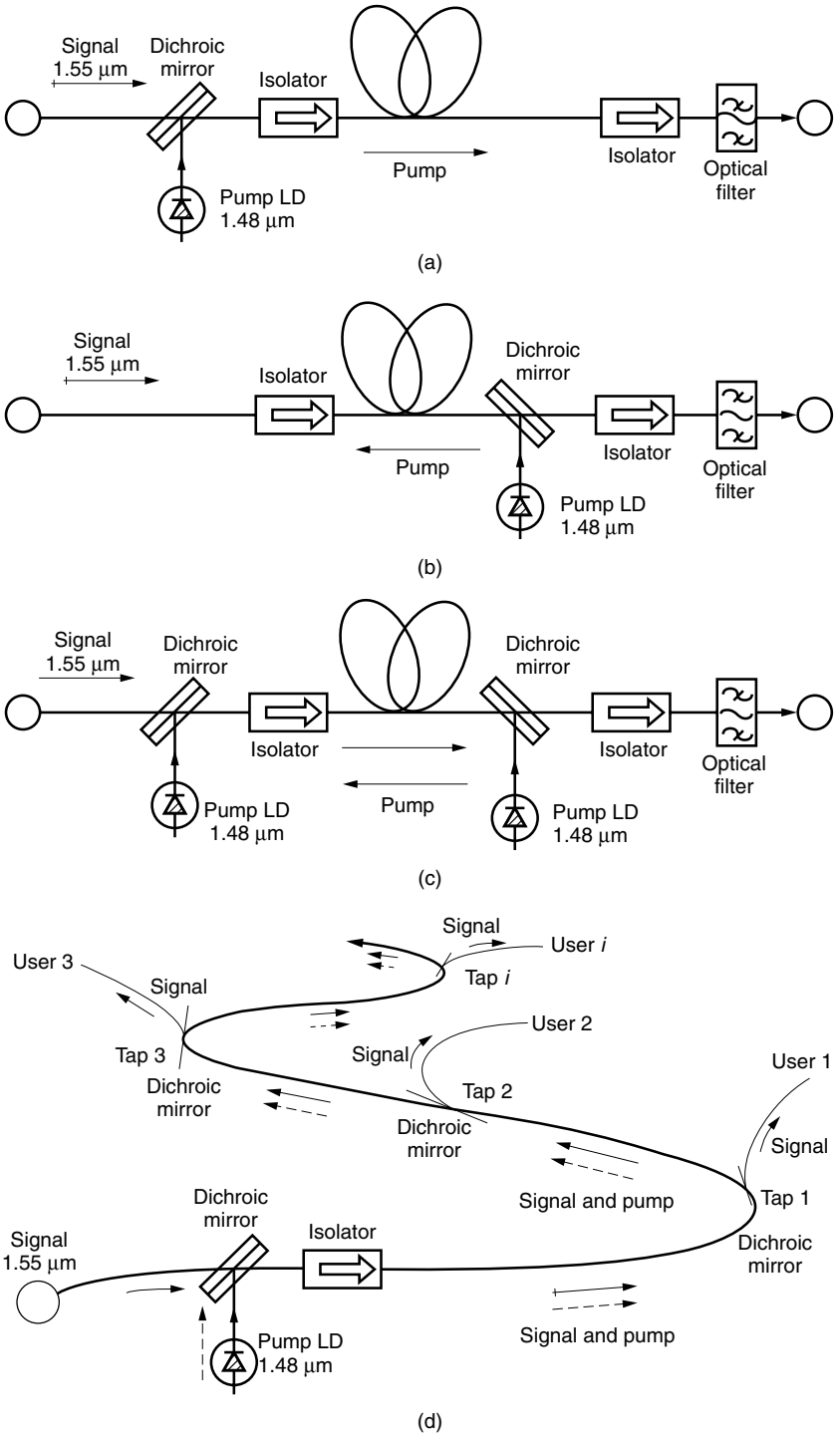


Figure 13.15 EDFA pumping configurations. (a) Forward pumping. (b) Backward pumping. (c) Bidirectional pumping. (d) Remotely pumped concatenated EDFA network.

raise the saturation signal power. Not only is the saturation signal power increased, but the total gain of the amplifier is also increased. If not for the increased pump power, the gain would have been saturated and the overall power would have been reduced. Thus, backward pumping has the advantages of higher saturation power and higher gain.

Finally, bidirectional pumping, as shown in Fig. 13.15c, enjoys the advantages of both at the cost of the complexity.

As shown in Fig. 13.15, the pump light from the 1.48-μm laser diode is coupled to the 1.55-μm signal light path by means of a dichroic mirror. The dichroic mirror is made up of layers of dielectric film whose thicknesses are arranged so as to pass the 1.55-μm light but reflect the 1.48-μm light. Optical isolators, which are not sensitive to the light polarization (see Section 5.5.2) but are sensitive to the direction of propagation, are installed on both ends of the fiber to prevent the reflected wave from going back into the fiber.

An optical filter is installed at the end of the fiber amplifier. This filter is necessary to prevent the pump light from going out of the amplifier and also to remove the effect of the ASE noise in the detector. From the viewpoint of removing the ASE noise, a narrowband filter (<1 nm) is desirable.

As shown in Fig. 13.15d, there is a scheme in which a single pump source supplies pump to concatenated EDFAs. It is a kind of forward pumping. The role of gain in such a network is to overcome propagation losses as well as to compensate for the signal power tapped off by users for detection. Such a scheme is worthy of consideration for managing the densely populated metropolitan area network (MAN) [9].

13.8.2 Double-Clad Fiber Pumping

Confinement of both the signal and pump light in a 5-μm core, such as shown in Fig. 13.14, becomes impractical with a high-power fiber amplifier because of the power density handling capability (nonlinear effects in the core). The double-clad fiber [10,11] subdivides the cladding into inner and outer cladding regions. The signal light propagates as a single mode inside the core, while the pump light propagates as multimodes inside the inner cladding.

Figure 13.16 shows various double-clad fiber geometries.

In Fig. 13.16a, the single-mode core and multimode inner cladding are concentric. Among the higher order modes, there are more skew rays than meridional rays. As shown in Fig. 11.1b, the skew ray goes around the guiding layer and the light energy is concentrated everywhere but the center, where the core is located. The confinement factor is low.

In order to increase the overlap of the skew rays of the pump with the signal in the core, the location of the core is offset in Fig. 13.16b. Figure 13.16c is similar to Fig. 13.16b, but with several offset cores. However, for both the geometries shown in Figs. 13.16b and 13.16c, the offset creates difficulties in light coupling or splicing the fibers.

The geometry in Fig. 13.16d is slightly different. The inner cladding is rectangular. The maximum intensity of the fundamental mode is located at the center as in the case of a rectangular microwave waveguide, and a high confinement factor is achieved.

A D-shaped inner cladding, such as shown in Fig. 13.16e improves the coupling of the skew ray with the signal light in the core.

The geometry of the multiple D fiber is shown in Fig. 13.16f. This geometry not only increases the amount of overlap but also restores the symmetry of the fiber cross section for easier light coupling or splicing.

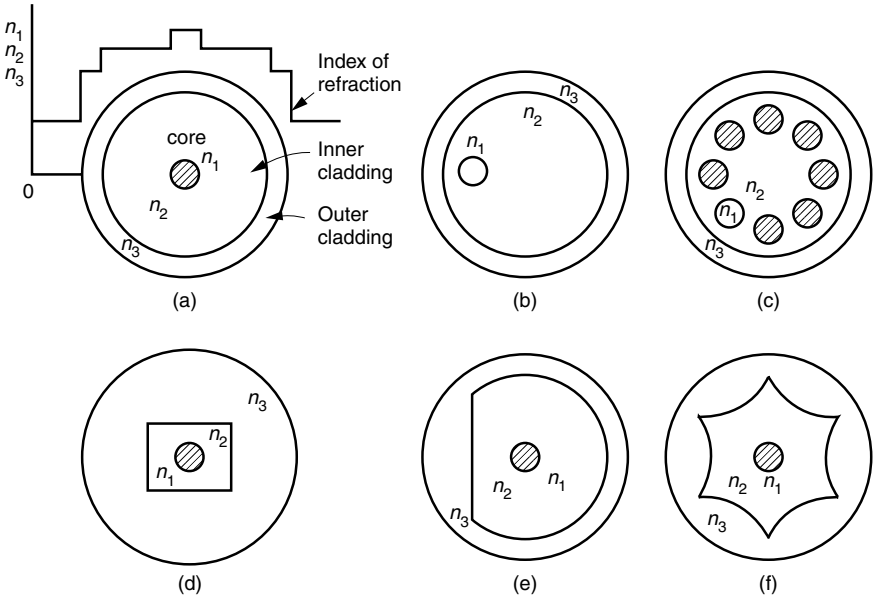


Figure 13.16 Various geometries of a double-clad fiber for high-power fiber amplifiers.

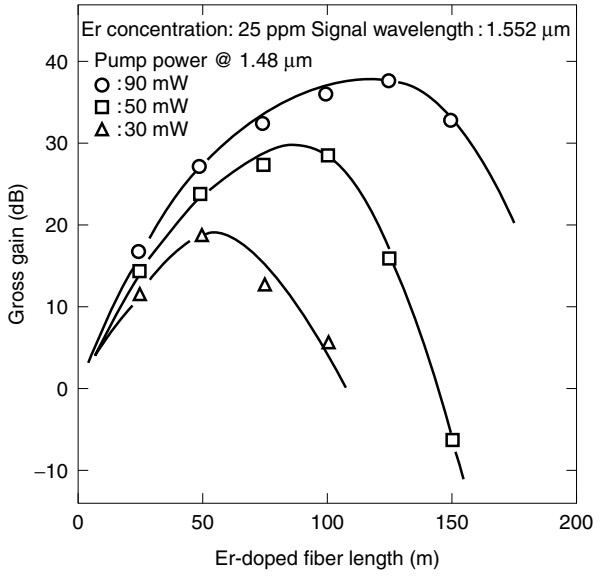


Figure 13.17 Signal gain characteristics versus EDFA length. (After K. Nakagawa et al. [7].)

13.9 OPTIMUM LENGTH OF THE FIBER

Figure 13.17 [7] shows experimental results for the gain of an erbium-doped fiber amplifier as a function of fiber length. The optimum length for maximum gain depends on the pump light power. With an increase in the pump power, the optimum length as well as the maximum gain increases.

If the length of the fiber is too long, there is a region of the fiber where the pump power is too small, the signal has reached saturation intensity, and the gain is decreased.

The ASE noise should also be a consideration when choosing the fiber length. As the fiber length increases, the ASE noise keeps on increasing while the signal is saturated until the ASE noise becomes so large that the ASE noise is the same power level as the signal. From these considerations, the ideal fiber length is in the range of 20–150 meters.

The major applications of the EDFA to fiber-optic communication systems are as a preamplifier to the photodiode detector and as relay amplifiers to stretch the distance of the communication systems. The next two sections will be devoted to noise considerations when the EDFA is used for these applications.

13.10 ELECTRIC NOISE POWER WHEN THE EDFA IS USED AS A PREAMPLIFIER

The electronic noise power when the optical amplifier is used as a preamplifier [4,12,13] to the photodiode detector will be calculated. The layout of the system is shown in Fig. 13.18.

When the incident light to the photodiode detector has already been contaminated by ASE noise, the output current from the detector suffers from excess noise over and above the shot noise and thermal noise of the detector, and the sensitivity of the receiver suffers.

Figure 13.19 shows the measured ASE noise, and Fig. 13.20 compares the size of the frequency bandwidth of the ASE noise $\Delta\nu_t$, the optical band-pass filter passband width $\Delta\nu_f$, and the electronic signal bandwidth $2B$.

In order to reduce the contribution of the ASE noise, a band-pass optical filter is installed at the end of the erbium-doped fiber, as shown in Fig. 13.18. If the ASE noise is uniformly spread over $\Delta\nu_t$, the light output P from the EDFA through the optical filter with passband $\Delta\nu_f$ is, from Eq. (13.36),

$$P = GP_s + GP_{\text{ASE}} \tag{13.90}$$

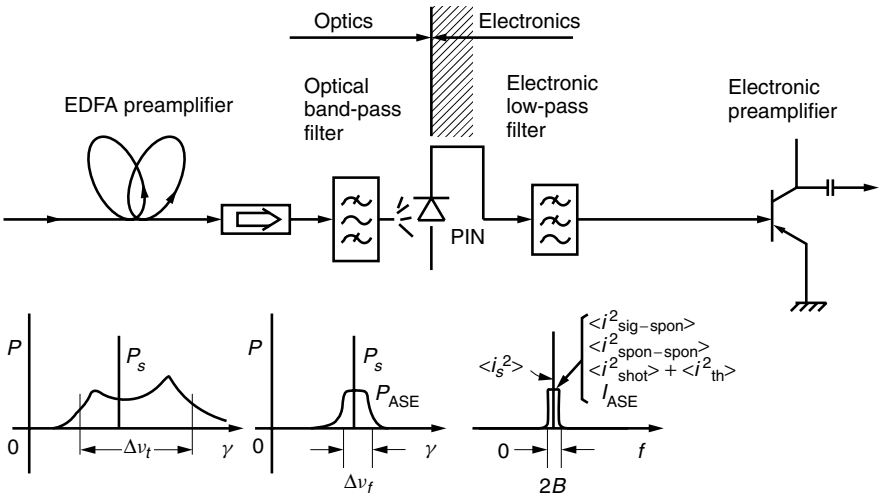


Figure 13.18 Receiver using EDFA as a preamplifier.

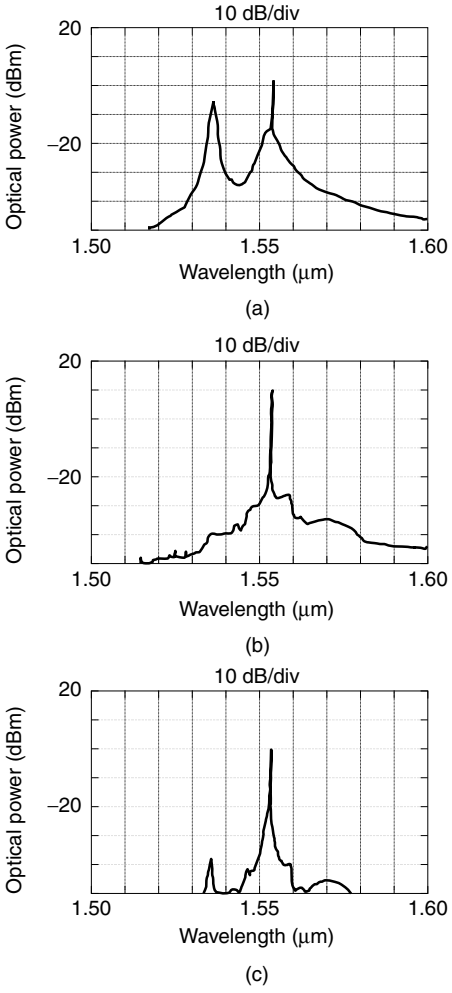


Figure 13.19 Noise spectrum of EDFA output. (a) Unsaturated region (input signal, −27.3 dBm). (b) Saturated region (input signal, −4.1 dBm). (c) Using 1-nm filter (input signal, −27.3 dBm). (After K. Nakagawa et al. [7].)

where

$$P_{\text{ASE}} = m_t n_{\text{spon}}\, h\nu \Delta\nu_f \tag{13.91}$$

$$G - 1 \doteqdot G \tag{13.92}$$

The output from the photodiode when the output P of Eq. (13.90) is fed to the photodiode will be calculated. The amplitude E_s and power GP_s of the light are related by

$$GP_s = \frac{1}{2}\frac{E_s^2}{\eta_0} s \tag{13.93}$$

where η_0 is the intrinsic impedance of air but not that of the photodiode medium (the definition of the quantum efficiency has already accounted for the reflection at the air

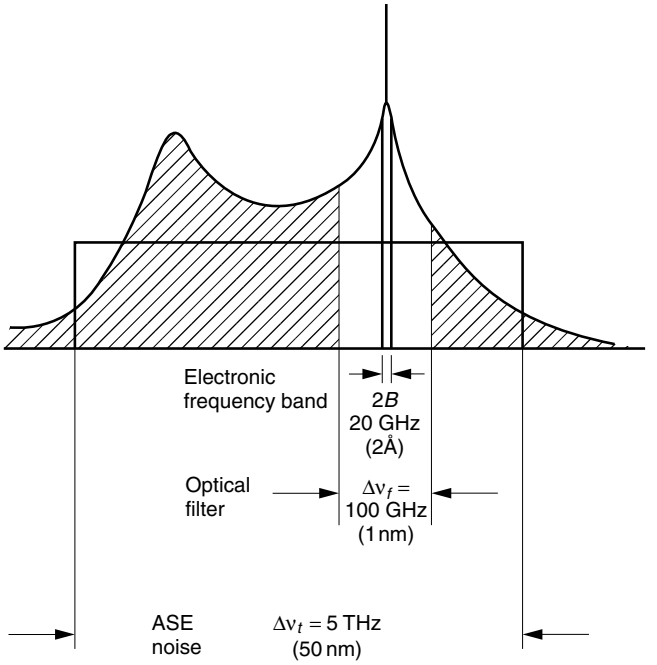


Figure 13.20 Relative frequency ranges compared: ASE noise, optical filter, and electronic frequency band.

and photodiode boundary in Section 12.4.2). The surface area of the photodiode is s. The amplitude of the input light signal into the photodiode is

$$E_s = \sqrt{\frac{2\eta_0}{s} G P_s} \tag{13.94}$$

If the assumption is made that the ASE noise spectrum can be represented as N discrete line spectra with equal powers of GP_{ASE}/N, as shown in Fig. 13.21, then the amplitude of each noise spectrum is

$$E_n = \sqrt{\frac{2\eta_0}{s} G P_{\text{ASE}}/N} \tag{13.95}$$

The photodiode plays the role of the mixer between the signal and the ASE noise. The calculation method is the same as that of heterodyne detection (see Section 12.6) but with the ASE noise as the local oscillator. Using Eqs. (12.17) and (12.18), the sinusoidal wave expression for the instantaneous output electrical current $i(t)$ from a PIN photodiode is

$$i(t) = \frac{\eta e}{h\nu} \frac{s}{\eta_0} \left(E_s \cos(2\pi \nu_s t + \phi_s) + \sum_{n=1}^{N} E_n \cos(2\pi \nu_n t + \phi_n) \right)^2 \tag{13.96}$$

The first term is the signal and the second term is the sum of the N discrete ASE noise spectrum lines. The square operation creates a number of different beat frequencies that belong to the following three categories:

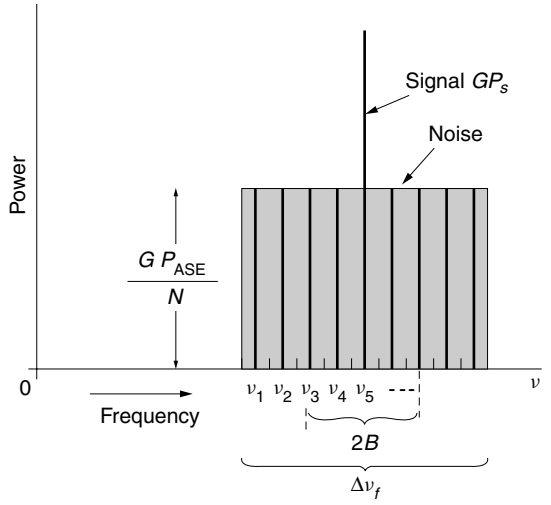


Figure 13.21 Division of continuous ASE noise spectrum into discrete ASE noise line spectra.

1. Signal current.
2. Noise current due to the beats between the signal and the ASE noise whose beat frequencies fall within the electronic frequency bandwidth of the detector. This noise is called signal–spontaneous beat noise.
3. Noise current due to the beats among the line spectra of the ASE noise whose beat frequencies fall within the electronic frequency bandwidth of the detector. This noise is called spontaneous–spontaneous beat noise.

Let us perform some calculations.

(1) The signal current $i(t)$ is

$$i(t) = \frac{\eta e}{h\nu}\frac{s}{\eta_0}E_s^2\cos^2(2\pi\nu_s t + \phi_s) \tag{13.97}$$

$$= \frac{\eta e}{h\nu}\frac{s}{\eta_0}E_s^2\tfrac{1}{2}\left[1 + \cos 2(2\pi\nu_s t + \phi_s)\right] \tag{13.98}$$

The photodiode is not sensitive to the $2\nu_s$ frequency component and the second term of Eq. (13.98) can be ignored. From Eq. (13.94), Eq. (13.98) becomes

$$i(t) = \frac{\eta e}{h\nu}GP_s \tag{13.99}$$

The electrical signal power s thus becomes

$$S = \left(\frac{\eta e}{h\nu}GP_s\right)^2 R_L \tag{13.100}$$

where R_L is the load resistance of the photodiode.

(2) Next, the signal–spontaneous beat noise $i_{\text{sig}-\text{spon}}$ is considered. By treating each noise spectrum as a local oscillator, the output current due to the beat between the signal and only the nth noise spectrum is derived from Eq. (13.96).

$$
\begin{aligned}
i_{\text{sig-spon}}(t) &= 2\frac{\eta e}{h\nu}\frac{s}{\eta_0}E_sE_n\sum_{n=1}^{N}\cos(2\pi\nu_s t+\phi_s)\cos(2\pi\nu_n t+\phi_n)\\
&= \frac{\eta_e}{h\nu}\frac{s}{\eta_0}E_sE_n\left(\sum_{n=1}^{N}\cos\left[2\pi(\nu_n-\nu_s)t+\phi_n-\phi_s\right]\right.\\
&\quad\left.+\sum_{n=1}^{N}\cos\left[2\pi(\nu_n+\nu_s)t+\phi_n+\phi_s\right]\right)
\end{aligned}
\tag{13.101}
$$

The second term in the large parentheses is outside the sensitivity range of the photodiode and the output current components from the photodiode are

$$
i_{\text{sig-spon}}(t) = \frac{\eta e}{h\nu}\frac{s}{\eta_0}E_sE_n\sum_{n=1}^{N}\cos\left[2\pi(\nu_n-\nu_s)t+\phi_n-\phi_s\right] \tag{13.102}
$$

$i_{\text{sig-spon}}(t)$ has N discrete spectra that are the beats made by ν_n and ν_s, and they are the components that can pass through the electronic filter. From Eqs. (13.94) and (13.95), the kth frequency current is

$$
\begin{aligned}
i_{\text{sig-spon},k}(t) &= 2\frac{\eta e}{h\nu}\sqrt{GP_s}\sqrt{GP_{\text{ASE}}/N}\\
&\quad\times\cos\left[2\pi(\nu_k-\nu_s)t+\phi_k-\phi_s\right]
\end{aligned}
\tag{13.103}
$$

The number of frequency components of $i_{\text{sig-spon}}(t)$ in Eq. (13.102) that are within $-B$ to B and get through the electronic preamplifier is

$$
N' = N\frac{2B}{\Delta\nu_f} \tag{13.104}
$$

Each spectrum has the same magnitude, and the total of the time averages of the beat current square is N' times and

$$
\langle i^2_{\text{sig-spon}}\rangle = 4\left(\frac{\eta e}{h\nu}\right)^2 GP_sGP_{\text{ASE}}\frac{B}{\Delta\nu_f} \tag{13.105}
$$

The current $\langle i^2_{\text{sig-spon}}\rangle$ is proportional to the product of the output signal power and the ASE noise.

(3) Next, the spontaneous–spontaneous beat noise $i_{\text{spon-spon}}(t)$ is calculated. The noise current due to the jth and kth ASE noise line spectra is obtained from the square of the second term in Eq. (13.96), suppressing ϕ_n:

$$
\begin{aligned}
i_{\text{spon-spon}}(t) &= \frac{\eta e}{h\nu}\frac{s}{\eta_0}\left(E_1\cos 2\pi\nu_1 t+E_2\cos 2\pi\nu_2 t+E_3\cos 2\pi\nu_3 t+\cdots\right)\\
&\quad\times\left(E_1\cos 2\pi\nu_1 t+E_2\cos 2\pi\nu_2 t+E_3\cos 2\pi\nu_3 t+\cdots\right)
\end{aligned}
\tag{13.106}
$$

The result of the multiplication in Eq. (13.106) will be grouped by the frequency components.

$$i_{\text{spon}-\text{spon}}(t) = \frac{\eta e}{h\nu}\frac{s}{\eta_0}\left[\begin{array}{c}\text{Zeroth group}\\ \left(\begin{array}{c} E_1^2\cos^2 2\pi\nu_1 t \\ + \\ E_2^2\cos^2 2\pi\nu_2 t \\ + \\ E_3^2\cos^2 2\pi\nu_3 t \\ + \\ \vdots \\ E_N^2\cos^2 2\pi\nu_N t \end{array}\right)\end{array} + 2\begin{array}{c}\text{First group}\\ \left(\begin{array}{c} E_1E_2\cos 2\pi\nu_1 t\cos 2\pi\nu_2 t \\ + \\ E_2E_3\cos 2\pi\nu_2 t\cos 2\pi\nu_3 t \\ + \\ E_3E_4\cos 2\pi\nu_3 t\cos 2\pi\nu_4 t \\ + \\ \vdots \\ E_{N-1}E_N\cos 2\pi\nu_{N-1} t\cos 2\pi\nu_N t \end{array}\right)\end{array}\right.$$

$$\left. + 2\begin{array}{c}\text{Second group}\\ \left(\begin{array}{c} E_1E_3\cos 2\pi\nu_1 t\cos 2\pi\nu_3 t \\ + \\ E_2E_4\cos 2\pi\nu_2 t\cos 2\pi\nu_4 t \\ + \\ E_3E_5\cos 2\pi\nu_3 t\cos 2\pi\nu_5 t \\ + \\ \vdots \\ E_{N-2}E_N\cos 2\pi\nu_{N-2} t\cos 2\pi\nu_N t \end{array}\right)\end{array} + \cdots\right] \quad (13.107)$$

Let the kth term of the zeroth group be

$${}_0i_{\text{spon}-\text{spon},k}(t) = \frac{\eta e}{h\nu}\frac{s}{\eta_0}E_k^2\tfrac{1}{2}(1+\cos 4\pi\nu_k t) \quad (13.108)$$

Terms of this group create dc and $2\nu_k$ frequency components. The $2\nu_k$ frequency components are outside the sensitivity of the photodiode and can be discarded. Thus, there are N terms in this group, all of which have the same magnitude, and using Eqs. (13.95) and (13.96), the total dc current is

$${}_0i_{\text{spon}-\text{spon}}(t) = \frac{\eta e}{h\nu}(GP_{\text{ASE}}/N)N \quad (13.109)$$

or

$${}_0i_{\text{spon}-\text{spon}}(t) = GI_{\text{ASE}} \quad (13.110)$$

where

$$I_{\text{ASE}} = \frac{\eta e}{h\nu}P_{\text{ASE}} \quad (13.111)$$

Note that $i_{\text{spon}-\text{spon}}(t)$ is independent of time and the contribution of I_{ASE} as noise is indirect in that an increase in the current flow through the photodiode increases the amount of the shot noise of the photodiode (Eq. (16.26)).

Next, the terms in the first group in Eq. (13.107) are considered. Frequencies of these components are all at one quantized frequency, $\Delta\nu_f/N$. For instance,

the first term is

$$\begin{aligned} {}_1i_{\text{spon}-\text{spon},1}(t) &= \frac{\eta e}{h\nu}\frac{s}{\eta_0}2E_1E_2\cos 2\pi\nu_1 t\cos 2\pi\nu_2 t \\ &= \frac{\eta e}{h\nu}\frac{s}{\eta_0}2E_1E_2\ \frac{1}{2}\left(\cos 2\pi\frac{\Delta\nu_f}{N}t+\cos 2\pi(\nu_1+\nu_2)t\right) \end{aligned} \tag{13.112}$$

Again, using the same reason of the insensitivity of the photodiode, the term with $\nu_1+\nu_2$ is ignored.

$${}_1i_{\text{spon}-\text{spon},1}(t) = \frac{\eta e}{h\nu}\frac{s}{\eta_0}E_1E_2\cos 2\pi\frac{\Delta\nu_f}{N}t \tag{13.113}$$

Using Eq. (13.95) gives

$${}_1i_{\text{spon}-\text{spon},1}(t) = 2\frac{\eta e}{h\nu}\left(GP_{\text{ASE}}/N\right)\cos 2\pi\frac{\Delta\nu_f}{N}t \tag{13.114}$$

and its current square component is

$$\langle {}_1i^2_{\text{spon}-\text{spon}}\rangle = \left(2\frac{\eta e}{h\nu}GP_{\text{ASE}}/N\right)^2\left(\frac{1}{2}\right) \tag{13.115}$$

There are $N-1$ such terms in the first group and the total current square of the first quantized frequency is

$$\langle {}_1i^2_{\text{spon}-\text{spon}}\rangle = 2\left(\frac{\eta e}{h\nu}GP_{\text{ASE}}/N\right)^2(N-1) \tag{13.116}$$

Next, the terms whose frequency is the second quantized frequency, $2\Delta\nu_f/N$, are in the second group in Eq. (13.107). The first term in the second group is

$${}_2i_{\text{spon}-\text{spon},1}(t) = \frac{\eta e}{h\nu}\frac{s}{\eta_0}2E_1E_3\cos 2\pi\nu_1 t\cos 2\pi\nu_3 t \tag{13.117}$$

If the term with $\nu_1+\nu_3$ is ignored because of the absence of sensitivity of the photodiode, then

$${}_2i_{\text{spon}-\text{spon},1}(t) = 2\frac{\eta e}{h\nu}\left(GP_{\text{ASE}}/N\right)\cos 4\pi\frac{\Delta\nu_f}{N}t \tag{13.118}$$

and its current square component is

$$\langle {}_2i^2_{\text{spon}-\text{spon},1}\rangle = 2\left(\frac{\eta e}{h\nu}GP_{\text{ASE}}/N\right)^2 \tag{13.119}$$

There are $N-2$ such terms in this group, and the total current square of the second quantized frequency, $2\Delta\nu_f/N$ is

$$\langle {}_2i^2_{\text{spon}-\text{spon}}\rangle = 2\left(\frac{\eta e}{h\nu}GP_{\text{ASE}}/N\right)^2(N-2) \tag{13.120}$$

Similarly, the total current square of the third quantized frequency, $3\Delta\nu_f/N$, is

$$\langle {}_3i^2_{\text{spon}-\text{spon}}\rangle = 2\left(\frac{\eta e}{h\nu}GP_{\text{ASE}}/N\right)^2 (N-3) \tag{13.121}$$

Comparing Eqs. (13.116), (13.120), and (13.121), the spectral density distribution of the spontaneous–spontaneous beat noise in the output current from the detector linearly decreases with the quantized frequency as long as the quantized frequency is less than B. It will, however, become zero as soon as the quantized frequency exceeds the electronic filter cutoff frequency.

Since each quantized frequency step is $\Delta\nu_f/N$, the number of frequency steps to reach B is

$$\frac{B}{\Delta\nu_f/N} = N\frac{B}{\Delta\nu_f} \tag{13.122}$$

Thus, the grand total of the spontaneous-to-spontaneous beat current square is

$$\begin{aligned}\langle i^2_{\text{spon}-\text{spon}}\rangle &= 2\left(\frac{\eta e}{h\nu}P_{\text{ASE}}/N\right)^2 \\ &\quad\times\left(N+(N-1)+(N-2)+(N-3)+\cdots+\left(N-N\frac{B}{\Delta\nu_f}\right)\right)\end{aligned} \tag{13.123}$$

Using the expression for the sum of an arithmetic series,

$$\langle i^2_{\text{spon}-\text{spon}}\rangle = 2\left(\frac{\eta e}{h\nu}GP_{\text{ASE}}/N\right)^2 \frac{\left(N+N-N\frac{B}{\Delta\nu_f}\right)\left(N\frac{B}{\Delta\nu_f}\right)}{2} \tag{13.124}$$

From Eq. (13.91), the final expression becomes

$$\langle i^2_{\text{spon}-\text{spon}}\rangle = \left(\frac{\eta e}{h\nu}GP_{\text{ASE}}\right)^2\left(2-\frac{B}{\Delta\nu_f}\right)\frac{B}{\Delta\nu_f} \tag{13.125}$$

Summarizing all beat noises gives

$$\begin{aligned}\langle i^2_{\text{sig}-\text{spon}}\rangle &= 4GI_s\cdot GI_{\text{ASE}}\cdot\frac{B}{\Delta\nu_f} \\ \langle i^2_{\text{spon}-\text{spon}}\rangle &= G^2I^2_{\text{ASE}}\left(2-\frac{B}{\Delta\nu_f}\right)\frac{B}{\Delta\nu_f} \\ {}_0i_{\text{spon}-\text{spon}}(t) &= GI_{\text{ASE}}\end{aligned} \tag{13.126}$$

where

$$I_s = \frac{\eta e}{h\nu}P_s \quad\text{and}\quad I_{\text{ASE}} = \frac{\eta e}{h\nu}P_{\text{ASE}}$$

It should be noted that I_s is the photocurrent from the photodiode with an optical amplifier gain of $G = 1$, and similarly I_{ASE} is the ASE current with $G = 1$.

Based on the results of the beat noises summarized in Eq. (13.126), the signal to noise ratio (S/N) of a receiver with an optical amplifier used as a preamplifier, as shown in Fig. 13.18, will be found. Besides the beat noise given by Eq. (13.126), the shot noise and thermal noise from the photodiode have to be included. The shot noise of the PIN diode is caused by the irregularity of the electron flow in the diode, and the shot noise power N_{shot} is given from Eq. (16.26) by

$$N_{\text{shot}} = 2eI_tBR_L \tag{13.127}$$

where e is the electron charge; I_t is the total average current including the signal current GI_s, the ASE noise current GI_{ASE}, and the dark current I_d of the PIN diode; B is the bandwidth of the noise power subjected to the electronic preamplifier; and R_L is the load resistance of the PIN diode (see Chapter 16). The thermal noise is emitted from any medium that has nonzero Kelvin temperature. The thermal noise is given from Eq. (16.27) by

$$N_{\text{th}} = \frac{4kTB}{R_L}R_L \tag{13.128}$$

where k is the Boltzmann constant, and T is the absolute temperature in kelvin.

The electrical S/N of the output from the PIN photodiode shown in Fig. 13.18 is, from Eqs. (13.126), (13.127), and (13.128),

$$\frac{\text{S}}{\text{N}} = \frac{(GI_s)^2}{4G^2I_sI_{\text{ASE}}\dfrac{B}{\Delta\nu_f} + G^2I_{\text{ASE}}^2\left(2 - \dfrac{B}{\Delta\nu_f}\right)\dfrac{B}{\Delta\nu_f} + 2e[G(I_s + I_{\text{ASE}}) + I_d]B + \dfrac{4kTB}{R_L}} \tag{13.129}$$

Dividing both denominator and numerator by G^2, the S/N becomes

$$\frac{\text{S}}{\text{N}} = \frac{I_s^2}{4I_sI_{\text{ASE}}\dfrac{B}{\Delta\nu_f} + I_{\text{ASE}}^2\left(2 - \dfrac{B}{\Delta\nu_f}\right)\dfrac{B}{\Delta\nu_f} + 2e\left(\dfrac{1}{G}(I_s + I_{\text{ASE}}) + \dfrac{I_d}{G^2}\right)B + \dfrac{4kTB}{G^2R_L}} \tag{13.130}$$

As seen from Eq. (13.130), as G is increased, the contribution of the thermal noise is diminished the most, and then that of the shot noise, while the contributions of the signal–spontaneous noise and the spontaneous–spontaneous noise are left unchanged. For large G, the S/N becomes

$$\frac{\text{S}}{\text{N}} = \frac{I_s}{\left[4 + \dfrac{I_{\text{ASE}}}{I_s}\left(2 - \dfrac{B}{\Delta\nu_f}\right)\right]I_{\text{ASE}}\dfrac{B}{\Delta\nu_f}} \tag{13.131}$$

If the input photocurrent I_s is maintained much larger than I_{ASE}, the ultimate S/N value of the receiver when the optical amplifier is used as an optical preamplifier is

$$\frac{\text{S}}{\text{N}} = \frac{I_s}{4I_{\text{ASE}}\dfrac{B}{\Delta\nu_f}} \tag{13.132}$$

where use was made of

$$G \gg 1 \tag{13.133}$$

$$I_s \gg I_{\text{ASE}} \tag{13.134}$$

In terms of the light power using Eqs. (13.91) and (13.126), Eq. (13.132) can be rewritten as

$$\frac{\text{S}}{\text{N}} = \frac{P_s}{4m_t n_{\text{spon}}\ h\nu B} \tag{13.135}$$

and Eq. (13.134) is rewritten as

$$P_s \gg m_t n_{\text{spon}}\ h\nu\ \Delta\nu_f \tag{13.136}$$

Before closing, Eq. (13.131) will be rewritten in terms of light powers, using Eqs. (13.91) and (13.126), as

$$\frac{\text{S}}{\text{N}} = \frac{P_s}{\left[4 + \dfrac{m_t n_{\text{spon}}\ h\nu\ \Delta\nu_f}{P_s}\left(2 - \dfrac{B}{\Delta\nu_f}\right)\right] m_t n_{\text{spon}}\ h\nu B} \tag{13.137}$$

In conclusion, in order to optimize the performance of the optical amplifier as a preamplifier, a high input light power has to be available. In other words, when this detection system is used for fiber-optic communication, the distance between the transmitter and the receiver should be limited such that the input power P_s to the optical amplifier is high enough not only to satisfy Eq. (13.136) but also to achieve the required S/N given by Eq. (13.135).

The value of P_s required for Eq. (13.136) to be satisfied can be reduced by reducing a few quantities. These are the number of modes m_t, $\Delta\nu_f$ of the optical filter, and the population ratio n_{spon}. Since $m_t = 2$ for a circularly polarized wave but $m_t = 1$ for a linearly polarized wave, m_t can be reduced by introducing a polarizer to permit only a linearly polarized wave. The quantity n_{spon} can be reduced by raising the pumping light level.

Example 13.6 Calculate the ASE noise power from the EDFA of Example 13.5 when the wavelength bandwidth of the optical filter is 0.1 nm and $m_t = 2$.

Solution From Eq. (13.36), the ASE noise power output with $m_t = 2$ is

$$GP_{\text{ASE}} = 2n_{\text{spon}}(G - 1)h\nu\ \Delta\nu_f$$

The expression for n_{spon} is found in Eq. (13.70):

$$n_{\text{spon}} = \frac{W_p + W_s}{W_p(1 - \beta) - 1/\tau}$$

From the results of Example 13.5, W_p and W_s are

$$W_p = 793\ \text{s}^{-1} \quad \text{and} \quad W_s = 88.9\ \text{s}^{-1}$$

The frequency interval is

$$\Delta\nu = \frac{3\times 10^8}{10^{-6}}\left(\frac{1}{1.55}-\frac{1}{1.5501}\right) = 1.25\times 10^{10}\ \text{Hz}$$
$$= 12.5\ \text{GHz}$$
$$n_{\text{spon}} = \frac{793+88.9}{(793)(1-0.38)-100} = 2.25$$
$$GP_{\text{ASE}} = 2(1556-1)(2.25)(6.63\times 10^{-34})(1.94\times 10^{14})(1.25\times 10^{10})$$
$$= 11.3\times 10^{-6}\ \text{W} = 11.3\ \mu\text{W}$$

which is a significant amount compared with the signal output of $P_s = 1$ mW. □

Example 13.7 As shown in Fig. 13.22, an EDFA is used as an optical preamplifier followed by a PIN photodiode to demonstrate the improvement of its S/N over the case when a PIN photodiode alone is used as a detector.

(a) Compare the values of the various components of the noise current squares.
(b) Find the S/N with and without the EDFA preamplifier.

The following parameters apply:

Input light power, $P_s = 3.2\ \mu$W
Wavelength, $\lambda = 1.55\ \mu$m
Bandwidth of optical filter, $\Delta\nu_f = 12.4$ GHz
Gain of EDFA, $G = 1097$
Number of orthogonal modes, $m_t = 2$
Population inversion factor, $n_{\text{spon}} = 2.25$
Responsivity of the PIN diode, $\eta e/h\nu = 0.5$
Dark current of the PIN diode, $I_d \doteqdot 0$
Load impedance to the PIN diode, $R_L = 50\ \Omega$
Room temperature, $T = 300$ K
Bandwidth of electrical signal, $B = 6.2$ GHz

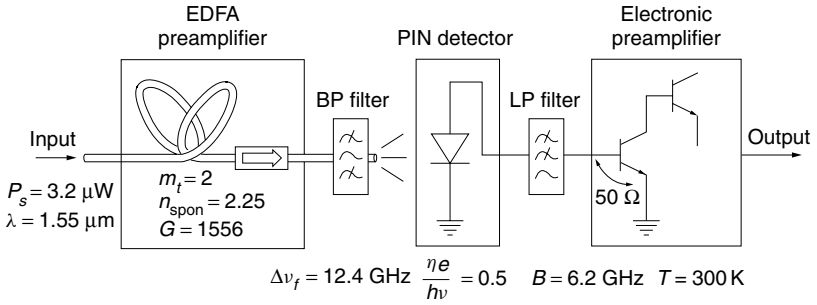


Figure 13.22 Physical layout when an EDFA is used as an optical preamplifier.

Solution The signal to noise ratio (S/N) of the receiver with the EDFA as an optical preamplifier is given by Eq. (13.130). The quantities needed to calculate the S/N are computed below.

$$
\begin{aligned}
I_s = \frac{\eta e}{h\nu} P_s &= (0.5)(3.2 \times 10^{-6}) = 1.6 \times 10^{-6} A \\
P_{\text{ASE}} &= m_t n_{\text{spon}}\, h\nu\, \Delta\nu_f \\
&= (2)(2.25)(6.63 \times 10^{-34})(1.94 \times 10^{14})(1.24 \times 10^{10}) \\
&= 7.18 \times 10^{-9}\ \text{W} \\
I_{\text{ASE}} = \frac{\eta e}{h\nu} P_{\text{ASE}} &= (0.5)(7.18 \times 10^{-9}) \\
&= 3.59 \times 10^{-9}\text{A} \\
\frac{B}{\Delta\nu_f} &= \frac{6.2 \times 10^9}{12.4 \times 10^9} = \tfrac{1}{2} \\
4\left(I_s I_{\text{ASE}} \frac{B}{\Delta\nu_f}\right) &= 4(1.6 \times 10^{-6})(3.59 \times 10^{-9})\left(\tfrac{1}{2}\right) \\
&= 1.15 \times 10^{-14}\ \text{A}^2 \\
I_{\text{ASE}}^2 \left(2 - \frac{B}{\Delta\nu_f}\right) \frac{B}{\Delta\nu_f} &= (3.59 \times 10^{-9})^2 \left(2 - \tfrac{1}{2}\right)\left(\tfrac{1}{2}\right) = 9.67 \times 10^{-18}\ \text{A}^2 \\
\frac{2e}{G}(I_s + I_{\text{ASE}})B &= \frac{2 \times 1.6 \times 10^{-19}}{1097}(1.6 \times 10^{-6} + 3.59 \times 10^{-9})(6.2 \times 10^9) \\
&= 2.9 \times 10^{-18}\ \text{A}^2 \\
\frac{4kTB}{R_L} &= \frac{4(1.38 \times 10^{-23})(300)(6.2 \times 10^9)}{50} \\
&= 2.06 \times 10^{-12}\ \text{A}^2 \\
\frac{4kTB}{G^2 R_L} &= \frac{2.06 \times 10^{-12}}{(1097)^2} = 1.7 \times 10^{-18}\ \text{A}^2
\end{aligned}
$$

Substituting the computed values into Eq. (13.130) gives

$$
\begin{aligned}
\frac{\text{S}}{\text{N}} &= \frac{I_s^2}{4I_s I_{\text{ASE}} \frac{B}{\Delta\nu_f} + I_{\text{ASE}}^2 \left(2 - \frac{B}{\Delta\nu_f}\right) \frac{B}{\Delta\nu_f} + \frac{2e}{G}\left(I_s + I_{\text{ASE}} + \frac{I_d}{G}\right) B + \frac{4kTB}{G^2 R_L}} \\
&= \frac{(1.6 \times 10^{-6})^2}{1.15 \times 10^{-14} + 9.67 \times 10^{-18} + 2.9 \times 10^{-18} + 1.7 \times 10^{-18}} = 223
\end{aligned}
$$

Thus, the magnitude of the signal–spontaneous beat noise is significantly larger than the spontaneous–spontaneous beat noise, the shot noise, and the thermal noise.

Next, the S/N without the optical preamplifier is calculated:

$$\begin{aligned}\frac{\mathrm{S}}{\mathrm{N}} &= \frac{I_s^2}{2eI_sB + \dfrac{4kTB}{R_L}} \\ &= \frac{(1.6\times10^{-6})^2}{(2)(1.6\times10^{-19})(1.6\times10^{-6})(6.2\times10^{9}) + 2.06\times10^{-12}} \\ &= \frac{2.56\times10^{-12}}{3.17\times10^{-15} + 2.06\times10^{-12}} = 1.25\end{aligned}$$

The S/N is improved by a factor of nearly 200 by inserting the EDFA preamplifier. □

13.11 NOISE FIGURE OF THE RECEIVER USING THE OPTICAL AMPLIFIER AS A PREAMPLIFIER

The signal to noise ratio S/N and the noise figure F [4,12–14] are among the most important considerations in designing an optical communication system. The optical S/N, which is the ratio between the first and second terms of Eq. (13.90), is seldom used. The output S/N normally means the ratio between the signal current square to the total noise current square if the light is detected by a photodiode detector.

Next, the noise figure of the amplifier will be derived. The noise performance of the amplifier is quantified by the noise figure F defined as the ratio of input and output signal to noise ratios:

$$F = \frac{(\mathrm{S/N})_i}{(\mathrm{S/N})_o} \tag{13.138}$$

The output signal to noise ratio $(\mathrm{S/N})_o$ of the amplifier can be obtained immediately by dividing the input signal to noise ratio $(\mathrm{S/N})_i$ by the noise figure F. The noise figure F is sometimes expressed in dB as

$$10\log(\mathrm{S/N})_o = 10\log(\mathrm{S/N})_i - 10\log F \tag{13.139}$$

10 log F is a measure of the degradation of the output $(\mathrm{S/N})_o$ from that of the input. The smaller F is, the better.

The output signal to noise ratio $(\mathrm{S/N})_o$ of the EDFA is the one that is measured at the output side of the EDFA, as shown in Fig. 13.23a, using an ideal photodiode whose shot noise is the only source of noise. The input signal to noise ratio $(\mathrm{S/N})_i$ is the one that is measured at the input side of the EDFA, as shown in Fig. 13.23b, disconnecting the EDFA preamplifier and using the same ideal photodetector. The output current square from such a detector is

$$\langle i_s^2 \rangle = \left(\frac{\eta e}{h\nu} P_s\right)^2 = I_s^2 \tag{13.140}$$

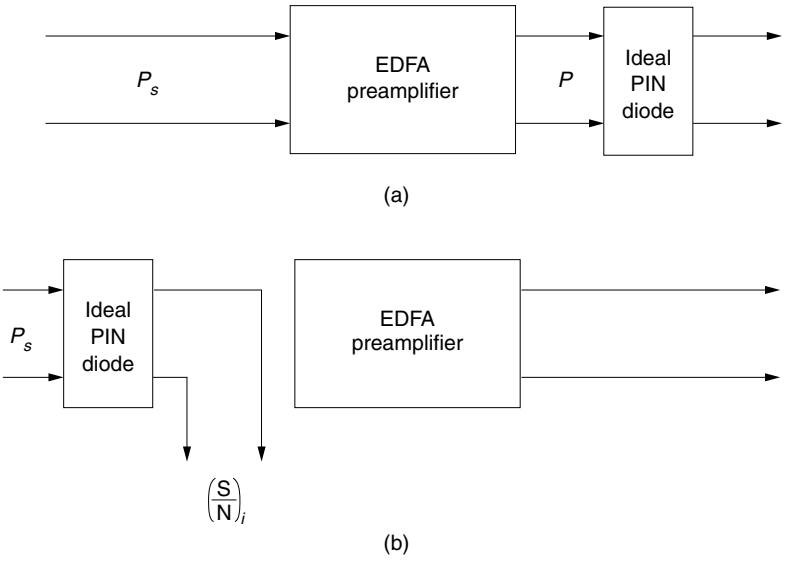


Figure 13.23 Method of measuring the noise figure of the EDFA: $F = (S/N)_i/(S/N)_o$. (a) S/N with the EDFA preamplifier. (b) S/N without the EDFA preamplifier.

The shot noise is

$$\langle i_{\text{shot}}^2 \rangle = 2e\frac{\eta e}{h\nu}P_s B = 2eI_s B \tag{13.141}$$

The input signal to noise ratio is

$$\left(\frac{\text{S}}{\text{N}}\right)_i = \frac{I_s}{2eB} \tag{13.142}$$

From Eqs. (13.131) and (13.142), F is

$$F = \frac{(\text{S/N})_i}{(\text{S/N})_o} = \left[4 + \frac{I_{\text{ASE}}}{I_s}\left(2 - \frac{B}{\Delta\nu_f}\right)\right]\frac{I_{\text{ASE}}}{2e\Delta\nu_f} \tag{13.143}$$

or in terms of the light power

$$F = \frac{\eta}{2}\left[4 + \frac{m_t n_{\text{spon}}\ h\nu\ \Delta\nu_f}{P_s}\left(2 - \frac{B}{\Delta\nu_f}\right)\right] m_t\ n_{\text{spon}} \tag{13.144}$$

The value of the second term in the square brackets can be small compared to the first term. For example, when $m_t = 1$, $n_{\text{spon}} = 3$, $\Delta\nu_f = 100$ GHz, $P_s = -30$ dBm, and $\lambda = 1.55$ μm, then

$$n_{\text{spon}}\ h\nu\ \Delta\nu_f/P_s = (3)(6.63 \times 10^{-34} \times 1.94 \times 10^{14})(10^{11})/(10^{-6}) = 0.039 \tag{13.145}$$

If this same example is computed with $P_s = -50$ dBm, the second term becomes significant compared to the first term, and hence, small values of P_s should be

avoided. As for the value of $n_{\text{spon}} = N_2/(N_2 - N_1)$, it can be made close to unity by pumping hard.

13.12 A CHAIN OF OPTICAL AMPLIFIERS

Optical amplifier chains [4,15–18] are used in optical trunk transmission cables and transoceanic optical submarine cables. Special attention should be paid to the accumulation of ASE noise from the concatenated optical amplifiers.

Figure 13.24 shows a segment of chained optical amplifiers in a trunk line together with an optical signal level diagram. It is assumed that after each amplifier the received signal is amplified to the same output level as that of the previous amplifier. This is accomplished by setting the gain of the amplifier to the same value as the fiber transmission loss between the amplifiers. The distribution of optical power along the line looks like a sawtooth function.

From Eq. (13.36), the output from the first amplifier with an optical filter with bandwidth $\Delta\nu_f$ is

$$P_1 = GP_s + (G-1)m_t n_{\text{spon}}\, h\nu\, \Delta\nu_f \tag{13.146}$$

Only the last amplifier is followed by a detector, so that the beat electrical current noise is accounted for following the last stage. After the attenuation L due to the fiber transmission loss, the input to the second amplifier is

$$P_{s,2} = LGP_s + L(G-1)m_t n_{\text{spon}}\, h\nu\, \Delta\nu_f \tag{13.147}$$

The power is assumed to recover to its original level at each amplifier:

$$GL = 1 \tag{13.148}$$

With Eq. (13.148) and $G \gg 1$, Eq. (13.147) becomes

$$P_{s,2} = P_s + m_t n_{\text{spon}}\, h\nu\, \Delta\nu_f \tag{13.149}$$

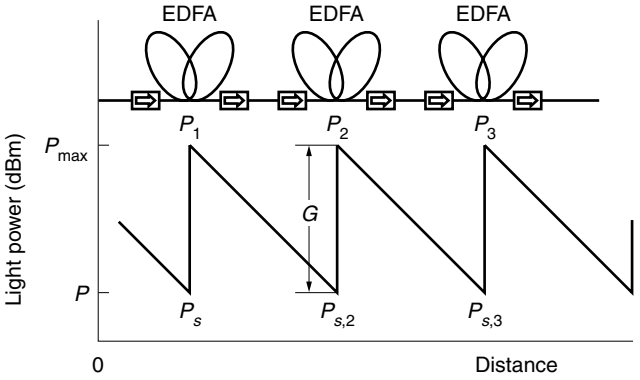


Figure 13.24 Power level diagram of an EDFA chain.

Compared with the input to the first amplifier, the input to the second amplifier contains an ASE noise of $m_t n_{\text{spon}}\ h\nu\ \Delta\nu_f$. The output from the second amplifier is

$$\begin{aligned} P_2 &= G(P_s + m_t n_{\text{spon}}\ h\nu\ \Delta\nu_f) \\ &\quad + (G-1)m_t n_{\text{spon}}\ h\nu\ \Delta\nu_f \end{aligned} \tag{13.150}$$

With the assumption $G \gg 1$, Eq. (13.150) becomes

$$P_2 = GP_s + 2Gm_t n_{\text{spon}}\ h\nu\ \Delta\nu_f \tag{13.151}$$

After transmission through the kth amplifier, the output power is

$$P_k = GP_s + kGm_t n_{\text{spon}}\ h\nu\ \Delta\nu_f \tag{13.152}$$

If the kth amplifier is followed by another transmission cable with loss L and with a detector at the end, the input power to the detector becomes

$$P_{s,d} = P_s + k\ m_t n_{\text{spon}}\ h\nu\ \Delta\nu_f \tag{13.153}$$

An important conclusion drawn from Eq. (13.153) is that the ASE noise is additive rather than multiplicative. That is, the noise term is $k\ m_t n_{\text{spon}}\ h\nu\Delta\nu_f$, and not $(m_t n_{\text{spon}}\ h\nu\ \Delta\nu_f)^k$. The reason for this is that the ASE noise power, as well as the signal power, is attenuated by L during the transmission in the cable.

Let's evaluate the performance of the chain amplifier combined with a detector at the end. The overall $(\text{S/N})_o$ and noise figure after detection are calculated by replacing n_{spon} by kn_{spon} in Eq. (13.137) and Eq. (13.144), respectively. The noise figure of the k amplifier chain is, therefore,

$$F_k = \frac{\eta}{2}\left[4 + \frac{km_t n_{\text{spon}}\ h\nu\ \Delta\nu_f}{P_s}\left(2 - \frac{B}{\Delta\nu_f}\right)\right] km_t n_{\text{spon}} \tag{13.154}$$

When designing repeater amplifiers in a trunk line, the number k of amplifiers has a significant impact on the overall performance. Suppose that the total distance D of the trunk line and the maximum output power from each amplifier are both fixed. A reduction in k means a reduction of input power P_s to each amplifier, and with a reduction of P_s, the contribution of the spontaneous–spontaneous beat noise of the second term in the square brackets of Eq. (13.154) is increased. There is an optimum number k of amplifiers that minimizes the noise figure.

Example 13.8 Compare the noise figures of the relay amplifiers for the two cases shown in Fig. 13.25. The configuration shown in Fig. 13.25a has higher gain (60 dB) with smaller input power and hence only one amplifier repeater is needed. The configuration shown in Fig. 13.25b has lower gain (30 dB) with larger input power and hence two amplifiers are needed. The maximum signal output in both cases is set at 5 dBm. Fiber transmission loss is 0.6 dB/km. The parameters for each optical amplifier are $\eta = 1$, $m_t = 2$, $n_{\text{spon}} = 3$, $\Delta\nu_f = 100$ GHz, $\lambda = 1.55$ μm, and $B = \Delta\nu_f$.

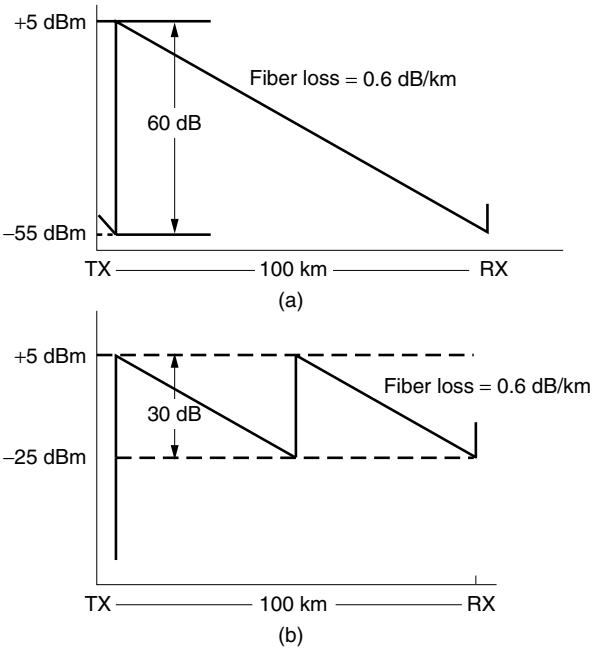


Figure 13.25 Comparison of power level diagrams for two chained amplifier configurations. (a) With 60-dB amplifier, higher gain and longer span, (b) With 30-dB amplifier, lower gain and shorter span.

Solution The input powers for the two cases are

$$P_s = 3.16 \times 10^{-9} \text{ W} \quad \text{for the 60-dB amplifier}$$
$$P_s = 3.16 \times 10^{-6} \text{ W} \quad \text{for the 30-dB amplifier}$$

(a) The noise figure F_1 for one amplifier in Fig. 13.25a with the given parameters is

$$\begin{aligned} F_1 &= 2\left[2n_{\text{spon}} + n_{\text{spon}}^2 h\nu\ \Delta\nu_f / P_s\right] \\ &= 2\left[2(3) + (3^2)(6.63 \times 10^{-34})(1.94 \times 10^{14})(10^{11})/3.16 \times 10^{-9}\right] \\ &= 2 \times 6 + 73.3 = 85.3 \end{aligned}$$

The noise figure F_2 for two amplifiers of Fig. 13.25b is

$$\begin{aligned} F_2 &= 2\left[2(2n_{\text{spon}}) + 2^2 n_{\text{spon}}^2 h\nu\ \Delta\nu_f / P_s\right] \\ &= 2\left[2(6) + 2^2(3^2)(6.63 \times 10^{-34})(1.94 \times 10^{14})(10^{11})/3.16 \times 10^{-6}\right] \\ &= 2 \times 12 + 0.29 = 24.29 \end{aligned}$$

Two distributed 30-dB amplifiers are better than one 60-dB amplifier.

In conclusion, a significant increase in the contribution of the spontaneous–spontaneous beat noise is incurred if the input power P_s becomes too low.

□

Example 13.9 An erbium-doped fiber amplifier (EDFA) chain is used in an optical communication trunk line. Each EDFA has a gain of 20 dB and maximum signal output of 5 dBm. Other parameters for each EDFA are $m_t = 2$, $n_{\text{spon}} = 3$, $\Delta\nu_f = 100$ GHz, and $\lambda = 1.55$ μm. The attenuation of the trunk fiber is 0.4 dB/km. It is assumed that after each amplifier the received signal is amplified to the same output level as that of the previous amplifier. Draw a power level diagram of the EDFA chain of *both* the signal and ASE noise power with distance.

Solution The signal power level is

$$P = P_{\max} e^{-2\alpha z}$$

In dB, the above equation is expressed as

$$10 \log P = 10 \log P_{\max} - 20\alpha z \log e$$

where

$$10 \log P_{\max} = 5 \text{ dBm}$$

$$20\alpha \log e = 0.4 \text{ dB/km}$$

Hence,

$$10 \log P = 5 - 0.4\ z$$

$$P_{\text{ASE}} = m_t n_{\text{spon}}\ h\nu\ \Delta\nu_f$$

$$= (2)(3)(6.63 \times 10^{-34}) \left(\frac{3 \times 10^{14}}{1.55}\right) (10^{11})$$

$$= 7.7 \times 10^{-8}\ W = -41.1 \text{ dBm}$$

After the kth amplifier, the ASE noise power becomes kP_{ASE}. The ASE noise (in dB) after the kth amplifier is

$$10 \log kP_{\text{ASE}} = 10 \log P_{\text{ASE}} + 10 \log k$$

$$= -41.1 \text{ dBm} + 10 \log k$$

The ASE noise decays at the same rate, 0.4 dB/km, as the signal during the transmission. The diagram in Fig. 13.26 shows the power level diagrams of the signal and the ASE noise. □

Example 13.10 An optical communication trunk line is to be built with a separation of D km between the transmitter and the receiver and connected by an optical amplifier chain with parameters $\eta = 1$ and $m_t = 2$.

The maximum signal level is set to $P_{\max}$ Watts. The transmission loss of the optical fiber is l dB/km. The amplifiers are equally spaced.

(a) Find an expression for the noise figure F_k as a function of the number k of repeater amplifiers.

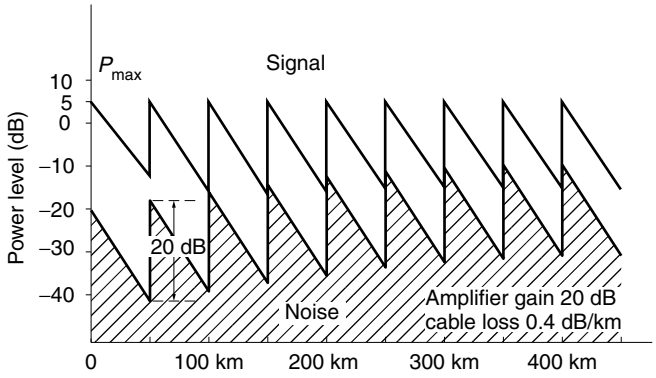


Figure 13.26 Power level diagram of signal and ASE noise of an EDFA.

(b) Derive an expression for numerically optimizing the number of repeater amplifiers.

Solution The power level diagram is shown in Fig. 13.27.

(a) The total power loss between the transmitter and the receiver is lD dB. If this loss is equally split into k relay spans, then each span has a cable loss of lD/k dB and $10\log(P_s/P_{\max}) = -lD/k$. Hence,

$$P_s = P_{\max}10^{-lD/10k} \text{ W}$$

(b) The expression for the noise figure with k number of repeaters given by Eq. (13.154) is rewritten as

$$F_k = ax + bx^2 10^{c/x}$$

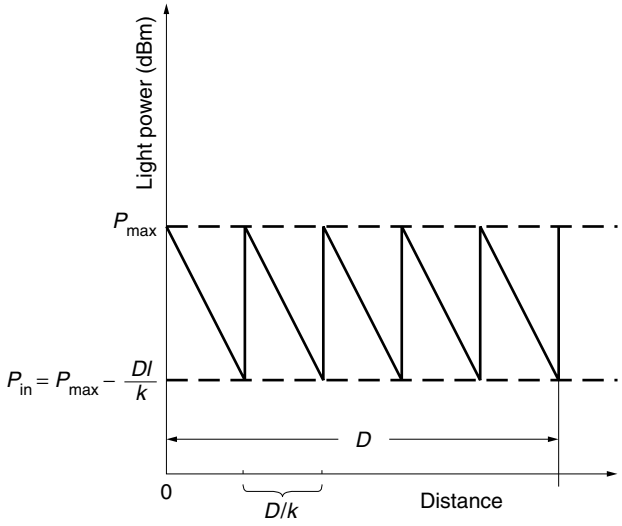


Figure 13.27 The power level diagram for k relay optical amplifiers over a distance D.

where

$$x = k$$
$$a = 2\eta m_t n_{\text{spon}}$$
$$b = \tfrac{1}{2}\eta m_t^2 n_{\text{spon}}^2\ h\nu\ \Delta\nu_f(2 - B/\Delta\nu_f)/P_{\max}$$
$$c = \frac{lD}{10}$$

The derivative of F_x with respect to x is set to zero to find the minimum value of F_x:

$$\frac{dF_x}{dx} = a + b(2x - c\ln 10)10^{c/x}$$

where the relationships

$$10^{c/x} = e^{(c/x)\ln 10}$$

and

$$\frac{d}{dx}\left(10^{c/x}\right) = -10^{c/x}\left(\frac{c\ln 10}{x^2}\right)$$

were used. The condition for $dF_x/dx = 0$ is

$$\frac{a}{b}\frac{1}{(c\ln 10 - 2x)} = 10^{c/x}$$

Unfortunately, this is a transcendental equation, but graphical solutions such as shown in Fig. 13.28 can be used. □

Example 13.11 Find the optimum number of EDFAs in the following system:

$$P_{\max} = 3\text{ mW}, \qquad \lambda = 1.55\ \mu\text{m}$$
$$D = 500\text{ km}, \qquad l = 0.3\text{ dB/km}$$
$$\eta = 0.8, \qquad m_t = 2, \qquad n_{\text{spon}} = 1$$
$$B = \Delta\nu_f = 20\text{ GHz}$$

Solution The noise figure F_k for the system with k amplifiers is, from Eq. (13.154),

$$F_k = \frac{\eta}{2}\left[4 + \frac{km_t n_{\text{spon}}\ h\nu\ \Delta\nu_f}{P_s}\left(2 - \frac{B}{\Delta\nu_f}\right)\right] km_t n_{\text{spon}}$$
$$P_s = P_{\max}10^{-lD/10k}$$

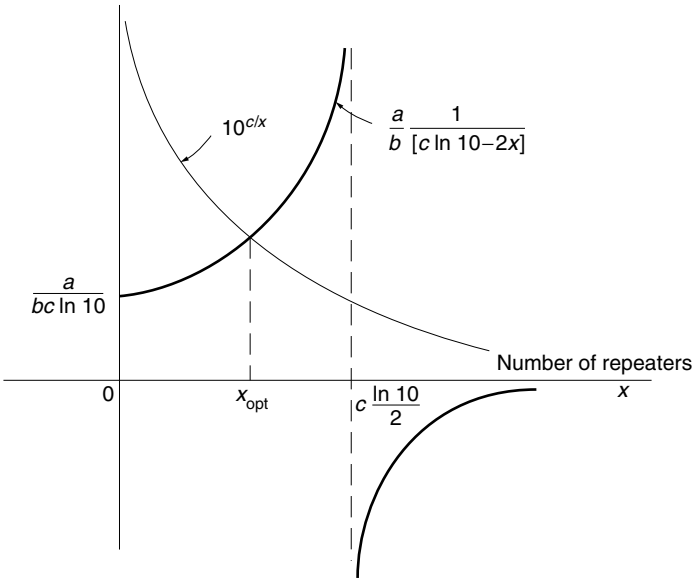


Figure 13.28 Solving a transcendental equation to find the optimum number x of repeater amplifiers.

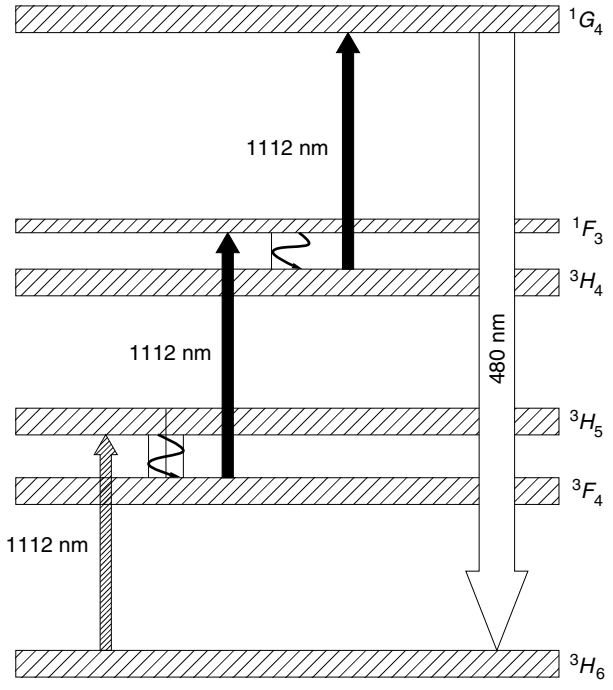


Figure 13.29 Triple-photon excitation of Tm^{3+} in ZBLAN fiber.

where

$$a = 2\eta m_t n_{\text{spon}} = 2(0.8)(2)(1) = 3.2$$

$$b = \tfrac{1}{2}\eta m_t^2 n_{\text{spon}}^2 h\nu\ \Delta\nu_f (2 - B/\Delta\nu_f)/P_{\max}$$

$$= \tfrac{1}{2}(0.8)(2^2)(1^2)(6.63 \times 10^{-34})(1.94 \times 10^{14})(2 \times 10^{10})/3 \times 10^{-3}$$

$$= 1.37 \times 10^{-6}$$

$$c = \frac{Dl}{10} = \frac{(500)(0.3)}{10} = 15$$

With $x = k$, the derivative $dF_x/dx = 0$ leads to the condition

$$q = \frac{a}{b} \cdot \frac{1}{c \ln 10 - 2x} - 10^{c/x} = 0$$

$$\frac{3.2}{1.37 \times 10^{-6}} \frac{1}{15 \ln 10 - 2x} - 10^{15/x}$$

$$q \begin{cases} <0 & \text{for } x = 3 \\ >0 & \text{for } x = 4 \end{cases}$$

and it is approximately $k = 3.3$. The distance between the amplifiers is $500/3 = 167$ km. □

13.13 UPCONVERSION FIBER AMPLIFIER

In Section 13.6, it was pointed out that the 800-nm pump for the EDFA is inefficient because of the excited state absorption (ESA). While this ESA is inefficient for amplification at 1.55 μm, the ESA can be used for amplification of shorter wavelength light at 540 nm. Referring to Fig. 13.13, the down transition from E_4 to E_1 emits 540-nm blue light; thus, a 800-nm infrared pump generates blue light in the Er^{3+}-doped fiber. This is unlike other situations where the pump light wavelength has always been shorter than that of the signal light. This particular excitation is known as upconversion by means of stepwise two-photon absorption.

The case of triple-photon excitation of thulium (Tm^{3+}) doping in a fluorozirconate glass (ZBLAN) fiber, whose energy diagram is shown in Fig. 13.29, can emit 480-nm blue light from 1112-nm infrared pump light [19]. Other types of upconversion mechanisms can be found in Ref. [20].

Upconversion is used for applications such as high-density CD recording by using a shorter wavelength light and to produce coherent light at wavelengths for which no other coherent light sources are available.

PROBLEMS

13.1 Figure P13.1a shows an experimental setup for learning the principles of operation of light amplification by stimulated emission in a neodymium YAG rod. The frequency ν of the signal light source S can be swept over a broad range of optical frequencies. The frequency ν_p of the pump light S_p, however,

is fixed rather than swept. The pump light is injected into the rod by means of a half-mirror (HM). There is a choice of detecting the signal light after one pass through the Nd YAG rod or two passes through the rod. The presence or absence of mirror M_1 determines one or two passes. The energy level diagram of the Nd YAG rod is shown in Fig. P13.1b. Even though Nd YAG has many energy levels, its function is normally explained by the four-level model, which consists of the ground level, $^4I_{11/2}$, $^4F_{3/2}$, and the pump band level.

(a) An experiment was performed with the pump light off and the filter F removed. Draw a qualitative graph of the detected light output from the Nd YAG rod with respect to frequency in terms of the wavenumber $1/\lambda$ (in 10^3 cm^{-1}) of the signal source S.

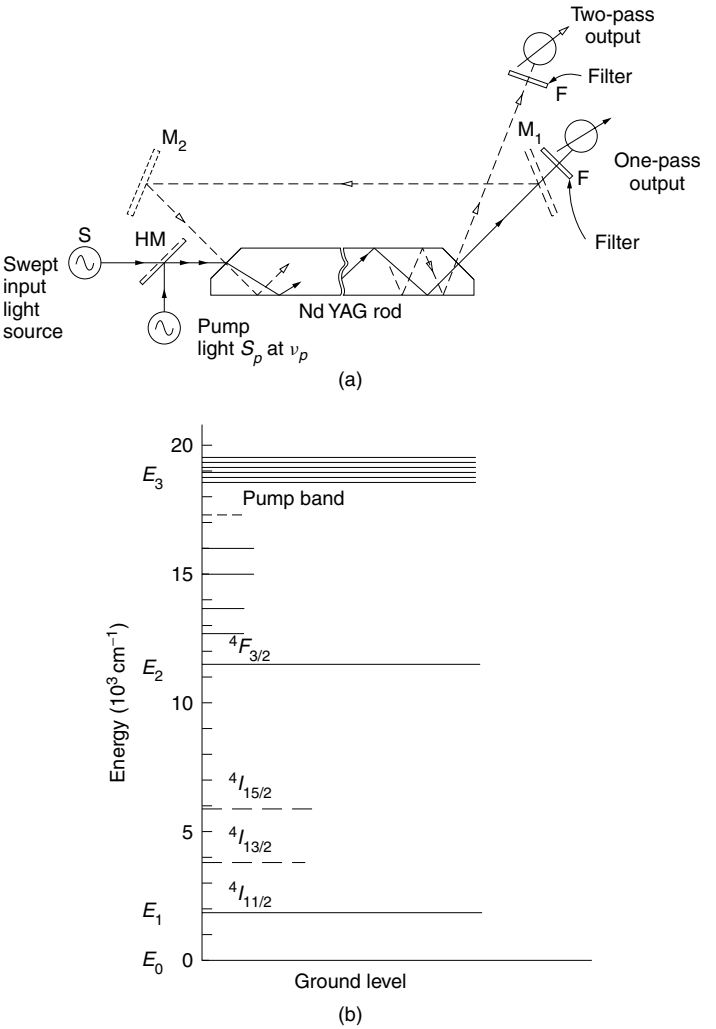


Figure P13.1 Configuration and energy levels of the Nd YAG optical amplifier. (a) Light amplification using stimulated emission. The solid line indicates one pass through the amplifier, and the dashed line indicates two passes through the amplifier. (b) Energy levels of Nd YAG.

(b) The experiment of part (a) is repeated with the pump light turned on. The energy $h\nu$ of the pump light corresponds to the difference in energy levels between the pump band and the ground levels. The filter F is installed to prevent the pump light from reaching the detector. What would be the change in the shape of the absorption lines compared to part (a)?

(c) The pump light is turned on, and signal light with frequency ν_{21} and input power of 2 mW enters the rod. After one pass, the output power was 10 mW. For the same input power, what would the output power be after two passes through the rod? The arrangement for the two passes is shown in Fig. P13.1a. Assume that M_1 and M_2 are perfect reflectors.

13.2 **(a)** What is the length of an erbium-doped fiber amplifier with 35-dB gain?

(b) What is the ASE noise power from this amplifier?

The amplifier parameters are

$N_1 = 1.8 \times 10^{17}\ \text{cm}^{-3}$
$N_2 = 4.8 \times 10^{17}\ \text{cm}^{-3}$
$\sigma_s = 7.0 \times 10^{-25}\ \text{m}^2$
$m_t = 1$
$\Delta\nu_f = 100\ \text{GHz}$
$\lambda = 1.55\ \mu\text{m}$
$h = 6.63 \times 10^{-34}\ \text{J} \cdot \text{s}$

13.3 With an erbium-doped fiber amplifier having the physical parameters described below, the threshold pump light power was found to be 20 mW. Determine the value of the metastable lifetime of Er^{3+}.

$\sigma_p = 0.42 \times 10^{-21}\ \text{cm}^2$
$\beta = 0.38$
$\Gamma = 0.4$
$r = 2\ \mu\text{m}$ (radius of the core)
$\lambda_p = 1.48\ \mu\text{m}$

13.4 **(a)** Find the optimum number of repeater amplifiers over a transmitter receiver distance of 1000 km from the viewpoint of optimizing the noise.

(b) What is the gain of each amplifier?

(c) What is the noise figure at the optimum condition?

Use the following parameters:

$m_t = 1$
$n_{\text{spon}} = 2$
$\eta = 1$
$\lambda = 1.55\ \mu\text{m}$
$P_{\max} = 10\ \text{mW}$
$l = 0.2\ \text{dB/km}$
$B = \Delta\nu_f = 100\ \text{GHz}$

REFERENCES

1. T. Miyazazaki, K. Inagaki, Y. Karasawa, and M. Yoshida, "Nd-doped double-clad fiber amplifier at 1.06 μm," *J. Lightwave Technol.* **16**(4), 562–566 (1998).
2. K. Isshiki, H. Sasamori, H. Watanabe, K. Kasahara and K. Ito, "A 980-nm band laser diode pump source with a detuned wavelength of 1000 nm for praseodymium doped fiber amplifiers," *J. Lightwave Technol.* **16**(3), 401–404 (1998).
3. A. Yariv, *Optical Electronics*, 4th ed., Saunders, Philadelphia, PA, 1991.
4. E. Desurvire, *Erbium-Doped Fiber Amplifiers: Principles and Applications*, Wiley, New York, 1994.
5. W. J. Miniscalco, "Erbium-doped glasses for fiber amplifiers at 1500 nm," *J. Lightwave Technol.* **9**(2), 234–250 (1991).
6. A. E. Siegman, *Lasers*, University Science Books, Mill Valley, CA, 1986.
7. K. Nakagawa, S. Nishi, K. Aida, and E. Yoneda, "Trunk and distribution network application of erbium-doped fiber amplifier," *J. Lightwave Technol.* **9**(2), 198–208 (1991).
8. E. Desurvire, "Analysis of transient gain saturation and recovery in erbium-doped fiber amplifiers," *IEEE Photonics Technol. Lett.* **1**(8), 196–199 (1989).
9. S. Mookherjea, "Remotely pumped optical distribution networks: A distributed amplifier model," *J. Lightwave Technol.* **19**(7), 926–932 (2001).
10. Y. Hirano, Y. Shoji, K. Kasahara, M. Yoshida, T. Araki, and Y. Hisada, "Multiwatt operation of square-shaped double-clad Nd-doped fiber amplifier," *Conference on Lasers and Electro-Optics (CLEO'98)*, pp. 427–428, 1998.
11. P. Glas, M. Naumann, A. Schirrmacher, and Th. Pertsch, "A CW diode-pumped single-silica fiber comprising 40 cores used as active elements for a high power fiber laser at $\lambda = 1050$ nm," *Conference on Lasers and Electro-Optics (CLEO'98)*, pp. 113–114, 1998.
12. S. Wen and S. Chi, "Characteristics of the gain and signal-to-noise ratio of a distributed erbium-doped fiber amplifier," *J. Lightwave Technol.* **10**(12), 1869–1878 (1992).
13. Y. H. Cheng, "Optimal design for direct-detection system with optical amplifiers and dispersion compensations," *J. Lightwave Technol.* **11**(9), 1495–1499 (1993).
14. S. Ryu, *Coherent Lightwave Communication Systems*, Artech House, Boston, (1995).
15. V. Arya and I. Jacobs, "Optical preamplifier receiver for spectrum-sliced WDM," *J. Lightwave Technol.* **15**(4), 576–583 (1997).
16. N. S. Sergano and C. R. Davidson, "Wavelength division multiplexing in long-haul transmission systems," *J. Lightwave Technol.* **14**(6) 1299–1308 (1996).
17. S. Sudo, *Optical Fiber Amplifiers: Materials, Devices and Applications*, Artech House, Boston, 1997.
18. M. J. Yadlowsky, E. M. Deliso, and V. L. Da Silva, "Optical fibers and amplifiers for WDM systems," *Proc. IEEE* **85**(11), 1765–1779 (1997).
19. P. Laperle, R. Vallée, and A. Chandonnet, "Stable blue emission from a 2500 ppm thulium-doped ZBLAN fiber laser," *Conference on Lasers and Electro-Optics (CLEO'98)*, pp. 77–78, 1998.
20. W. Lenth and R. M. Macfarlane, "Upconversion lasers," *Opt. Photonics News* **3**(3), 8–15 (1992).

14

TRANSMITTERS

The concept of a laser was first published by A. L. Schawlow and C. H. Townes in 1958. Successful operation of the first laser (ruby laser) by T. H. Maiman amazed the scientific world in 1960. Since then, laser development has progressed at a fast pace, and numerous kinds of lasers have been invented.

The principle of operation of a laser will be presented as an extension of the light amplifier presented in the previous chapter. In a light amplifier, light passes through the gain medium once. In a laser, there are partially reflecting mirrors at each end of the gain medium, forming what is called the laser cavity. On average, light travels back and forth through the laser cavity many, many times before exiting. For lasing to occur, the amplitude must grow with each round trip, and the various round-trip components must interfere constructively.

14.1 TYPES OF LASERS

Our primary interest is semiconductor lasers, but before moving on to an in-depth treatment of semiconductor lasers, let us take a moment to look at some other popular kinds of lasers. These include gas lasers, excimer lasers, dye lasers, solid-state lasers, and chemical lasers.

14.1.1 Gas Lasers

An example of a gas laser [1] is the helium–neon (He–Ne) laser. The He–Ne laser emits red light at a wavelength of 0.6328 μm with a narrow spectrum of a few kilohertz. Figure 14.1 shows the structure of the He–Ne laser. The gas pressure is a few torr with a 5:1 gas pressure ratio of He and Ne. A few kilovolts is applied to start a discharge and then the potential is reduced by about half after the discharge starts. The He–Ne laser is pumped by the impact of electrons in the discharge. The laser cavity is made of two concave end mirrors with reflectivities higher than $R = 99.8\%$.

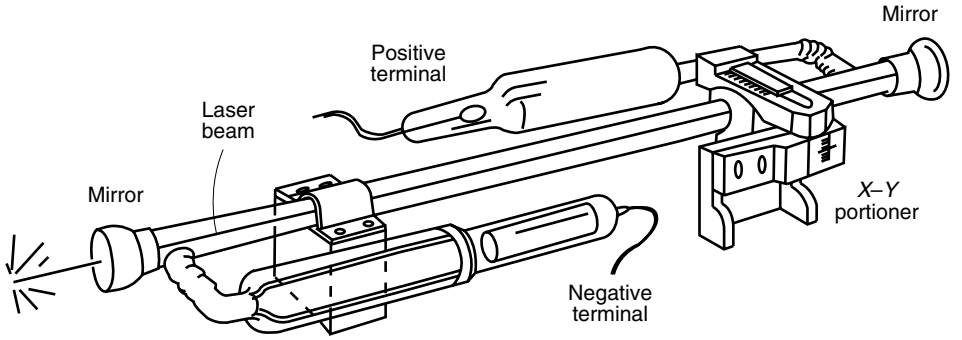


Figure 14.1 Geometry of the He–Ne gas laser.

The argon ion (Ar^+) laser is a gas laser that has several discrete spectra ranging from $\lambda = 0.488$ to 0.515 μm, of which the 0.515-μm line is the strongest.

The carbon dioxide (CO_2) laser emits high-intensity light at $\lambda = 10.6$ μm in the far-infrared region.

The excimer laser derives its name from an abbreviation for excited dimer. (A dimer is a molecule made of two identical atoms or molecules, whereas a polymer is made of many identical molecules.) Excimer lasers use such gases as Xe_2, Kr_2, or Ar_2. The output power is large and the wavelength lies in the vacuum ultraviolet (VUV) region, which is between 0.1 and 0.2 μm.

14.1.2 Solid-State Lasers

It was the solid-state laser [2] that T. H. Maiman first demonstrated at Hughes Aircraft Company. An artificial ruby (CrO_3 in Al_2O_3) was used. The wavelength of the ruby laser is 0.6934 μm.

Other popular solid-state lasers that the reader has some familiarity with from the previous chapter are the Er (erbium)-doped fiber laser, the Pr (praseodymium)-doped fiber laser, the neodymium YAG (yttrium–aluminum–garnet) laser, and the neodymium glass laser. The Er- and Pr-doped fiber lasers operate at 1.54 and 1.3 μm, respectively. The neodymium YAG and neodymium glass lasers operate at 1.064 and 1.3 μm.

The solid-state lasers are optically pumped by either a xenon or mercury flash lamp or even by a semiconductor laser array.

14.1.3 Dye Lasers

Organic dyes are used as laser media [1]. The gain curve of the rhodamine 6G dye extends from 0.57 to 0.66 μm and that of oxazine extends from 0.7 to 0.83 μm. The organic dyes are optically pumped by a laser with a wavelength shorter than the wavelength of their own emission. The wavelength of the dye laser can be scanned continuously by changing the resonance of the cavity containing the dye. If the laser media are exchanged, one after the other, the dye laser can cover a wide wavelength spectrum from 0.3 to 1.0 μm using the same physical arrangement.

14.1.4 Chemical Lasers

When hydrogen gas is injected into a high-temperature supersonic jet of SF_6, atomic fluorine (F) dissociated from SF_6 reacts with the hydrogen gas and chemical reactions take place, such as

$$H_2 + F \rightarrow HF + H \quad \text{or} \quad H_2 + 2F \rightarrow HF_2 + H \tag{14.1}$$

This chemical reaction releases 31.7 kcal/mol of energy. This energy pumps HF or HF_2 into the excited state. The transition of HF or HF_2 emits radiation at wavelengths ranging from 2.7 to 2.9 μm. It is reported that the conversion efficiency of chemical to laser energy is 12%. A laser with 20-foot diameter mirrors placed 3 feet apart emits more than 400-W laser light [3]. A special feature of the chemical laser is the possibility of creating a laser that does not require an external energy source.

14.2 SEMICONDUCTOR LASERS

Of all the lasers mentioned so far, semiconductor [4,5] and solid-state lasers are the most commonly used sources for fiber-optic communication. This is because of such advantages as:

- Long life
- High reliability
- Ruggedness, compactness, and light weight
- High efficiency of electrooptic conversion
- Low applied voltage
- Spectral purity compared to non laser sources
- Direct modulation capability up to tens of gigahertz

It is the semiconductor lasers that are predominantly used for fiber-optic communication systems. Gallium arsenide lasers doped with aluminum ($Al_xGa_{1-x}As$) cover the wavelength range from 0.85 to 0.78 μm with increasing x. The laser doped with indium and phosphorus rather than aluminum ($In_{1-x}Ga_xAs_yP_{1-y}$) covers from 0.9 to 1.67 μm, depending on the values of x and y. With the proper combination of x and y, a laser can be designed to emit at 1.3 μm corresponding to the wavelength of the nondispersive range of the glass optical fiber, or it can be designed to emit at 1.55 μm corresponding to the lowest loss range of glass optical fibers. In communications, the semiconductor laser is also referred to as a laser diode or LD for short.

14.2.1 Gain of a Semiconductor Laser Amplifier

In Chapter 13 on the optical amplifier, it was mentioned that a semiconductor laser immediately becomes a semiconductor laser amplifier (SLA) simply by removing the end reflectors of the laser cavity. The treatment in this section complements that in Sections 13.4 and 13.5.

What's cooking in the laser kitchen? A couple of novelty lasers will be added here. An edible laser was cooked by mixing the nontoxic dye fluorescene with gelatin. Emission at $\lambda = 0.657$ μm was reported.

Although one is tempted to dismiss the edible laser as being flippant, gelatin is in fact a material with a number of interesting properties and has been studied quite seriously. Gelatin is used in emulsions for photography and holography and has played an important role in the development of the sophisticated DFB laser (see Section 14.9.3).

It is also reported that a kitchen utensil can be used as a mirror in the laser cavity. The laser cavity was made by combining a spatula with a conjugate mirror. The conjugate mirror (see Section 8.1) is made by four-wave mixing in a photorefractive medium and has the special property of reflecting back the incident light along the same path that the light had originally. In other words, the reflected light is on the same side of the normal to the surface of the mirror rather than following the law of reflection, which states that the light is reflected on the other side of the normal to the mirror surface at an angle equal to the incident angle, as shown in Fig. 14.2. No matter how crudely the laser cavity mirror is oriented with respect to the other conjugate mirror, the light goes back and forth taking the same path, and an efficient feedback of light is realized. If a conjugate mirror is used as one mirror of the laser cavity, a handheld spatula can be used as the other laser cavity mirror.

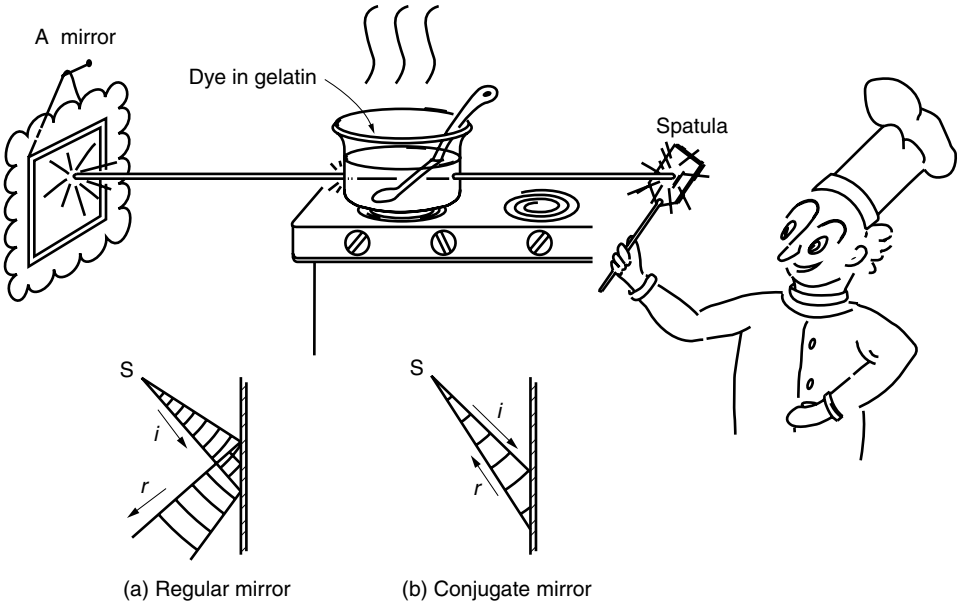


Figure 14.2 What's cooking in the kitchen? An edible laser!

One of the biggest factors that differentiates semiconductor laser materials from other materials is the existence of bands of energy levels rather than discrete energy levels, as in the case of gas or fiber amplifier materials mentioned in Chapter 13. A transition of an electron from any energy level within the conduction band into any level of the valence band emits light. Because of the spread of the energy band, the gain of the semiconductor material is spread over a wide range.

First, an explanation of the gain curve will be attempted using Fig. 14.3. Figure 14.3a shows how the electrons would populate the energy levels in a pure (intrinsic) semiconductor at a temperature of 0 K and with no current flowing through it ($T = 0$, $J = 0$). The energy level E_c is the bottom of the conduction band and E_v is the energy level of the top of the valence band and $g(E)$ represents the density of states. For Fig. 14.3a, all valence states are filled by electrons, as indicated by the hatched region, and the conduction states are empty. The Fermi level F, which defines the border between occupied electron states and unoccupied electron states at $T = 0$ K and $J = 0$, lies between E_c and E_v. Figure 14.3b illustrates what happens when an n-type impurity is introduced into the semiconductor, while keeping $T = 0$ and $J = 0$. An n-type impurity is one that has more valence electrons than the host material. In this case, all the valence band states are filled by electrons, and also the lower states of the conduction band are occupied up to the level F_c, which is the Fermi level of the n-doped semiconductor. Figure 14.3c is the same for a p-doped semiconductor. A p-type impurity is one that has less valence electrons than the host material. The Fermi level F_v of the p-doped semiconductor lies below E_v, indicating that there are unfilled states in the valence band.

If p-type and n-type semiconductors are joined, a nonequilibrium situation can be established, as shown in Fig. 14.3d, whereby the conduction band resembles that of the n-doped semiconductor and the valence band resembles that of the p-doped semiconductor. If one applies an optical pump to the junction area, or applies a bias

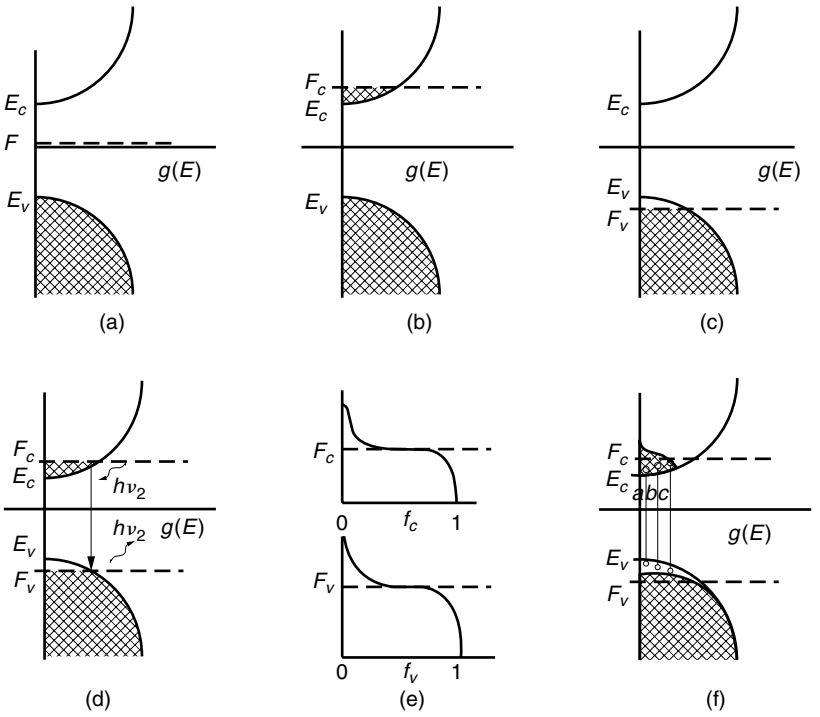


Figure 14.3 Density of state function $g(E)$ and the Fermi–Dirac distribution. (a) $T = 0$, $J = 0$, pure semiconductor. (b) $T = 0$, $J = 0$, n-doped semiconductor. (c) $T = 0$, $J = 0$, p-doped semiconductor. (d) $T = 0$, $J = 0$, p-n junction. (e) Fermi–Dirac distribution. (f) $T \neq 0$, $J = 0$, p-n junction.

voltage, electrons in the valence band absorb light and move to the conduction band, where they quickly settle into the lower states of the conduction band. Conversely, a photon of energy $h\nu_2$, as shown in Fig. 14.3d, can stimulate emission of another photon $h\nu_2$ by inducing an electron to make the transition from the conduction band to an empty state in the valence band.

Figure 14.3f depicts the situation where $T \neq 0$ and $J \neq 0$ at the p-n junction. The effect of a bias current is to create a nonequilibrium distribution similar to the impinging light in Fig. 14.3d. The effect of raising the temperature is to blur the distribution around F_c and F_v. The next step is to determine the criteria for amplification.

Before doing this, let us digress for a moment to explain semiconductor terminology. When an electron makes the transition from the valence band to the conduction band, it leaves behind a positively charged hole. The valence state vacated by the electron is said to be occupied by the positively charged hole. Furthermore, this hole is a positively charged carrier. The mobility of the hole arises from the fact that a nearby valence electron tends to fill the hole and, in doing so, leaves a hole behind in a different location, which in turn is filled by a nearby valence electron, and so forth. When an electron makes a transition from the conduction band to fill a hole in the valence band, this process is called recombination. Recombination can be viewed as the annihilation of a negative carrier (conduction electron) and a positive carrier (hole in the valence band), and energy is released. Energy is used either to radiate photons or heat the lattice of the crystal.

The electron density $n(E_2)$, which is the number of electrons per unit volume in energy levels between E_2 and $E_2 + dE_2$ in the conduction band, is given by

$$n(E_2)dE_2 = g_c(E_2)f_c(E_2)\,dE_2 \tag{14.2}$$

where

$$g_c(E_2) = \frac{\sqrt{2}m_e^{*3/2}}{\hbar^3\pi^2}(E_2 - E_c)^{1/2} \tag{14.3}$$

$$f_c(E_2) = \frac{1}{e^{(E_2-F_c)/kT} + 1} \tag{14.4}$$

In Eq. (14.3), m_e^* is the effective mass of an electron, and $\hbar = h/2\pi$.

The function $g_c(E_2)$ represents the density of states per unit energy per unit volume in the conduction band at energy level E_2 and is proportional to $(E_2 - E_c)^{1/2}$. The proof of Eq. (14.3) is presented later when discussing the case of a quantum well and it is given by Eq. (14.203). The function $f_c(E_2)$ is the Fermi–Dirac distribution function that specifies the probability that the energy level E_2 in the conduction band will be occupied by an electron [6]. The Fermi–Dirac distribution is graphed in Fig. 14.3e. F_c is the qausi-Fermi level of electrons in the conduction band. E_c is the bottom edge of the conduction band.

Similarly, at the energy level E_1 in the valence band, the hole density $p(E_1)$, which is the number of holes per unit volume that are in the energy levels between E_1 and $E_1 + dE_1$, is expressed by

$$p(E_1)dE_1 = g_v(E_1)[1 - f_v(E_1)]\,dE_1 \tag{14.5}$$

where

$$g_v(E_1) = \frac{\sqrt{2}m_h^{*3/2}}{\hbar^3\pi^2}(E_v - E_1)^{1/2} \tag{14.6}$$

$$f_v(E_1) = \frac{1}{e^{(E_1 - F_v)/kT} + 1} \tag{14.7}$$

$[1 - f_v(E_1)]$ is the probability of the absence of an electron or the presence of a hole at E_1. The quantities $g_v(E_1)$, $f_v(E_1)$, F_v, E_v, and m_h^* are similar to those defined for Eqs. (14.2), (14.3), and (14.4), but for holes in energy levels between E_1 and $E_1 + dE_1$ in the valence band.

The difference in the quasi-Fermi levels F_c and F_v is equal to the potential energy eV_b of the bias voltage V_b.

Recall that in the case of discrete energy levels, such as for an erbium-doped fiber amplifier, the emission equation, Eq. (13.20), consists of two terms. The first term is proportional to the input light whose coefficient can be considered as the gain of the amplifier. The second term is independent of the input light and can be considered as the noise of the amplifier. Let us first treat the amplifier term. Gain is the difference between stimulated emission and absorption. With a semiconductor, stimulated emission occurs when an incident photon induces an electron to make the transition from the conduction band to the valence band, and is proportional to the product $n(E_2)p(E_1)$ of the density of electrons in the conduction band and the density of holes in the valence band. Absorption is the transition in the reverse direction and is proportional to the product of the density of electrons in the valence band expressed by

$$n(E_1)dE_1 = g_v(E_1)f_v(E_1)dE_1 \tag{14.8}$$

and the density of vacant states in the conduction band expressed by

$$p(E_2)dE_2 = g_c(E_2)[1 - f_c(E_2)]\,dE_2 \tag{14.9}$$

Finally, the net increase in the photon density S (number of photons per unit volume) is thus

$$\frac{dS}{dt} = \iint W[n(E_2)p(E_1) - n(E_1)p(E_2)]\,dE_1\,dE_2 \tag{14.10}$$

W is the stimulated transition probability and is expressed by

$$W = BE_d \tag{14.11}$$

where B is Einstein's B coefficient and E_d is the energy density of the light that is stimulating the transition. E_d is given by

$$E_d = h\nu S \tag{14.12}$$

From Eqs. (14.10), (14.11), and (14.12), the increase in the energy density of the light is

$$\frac{dE_d}{dt} = \iint Bh\nu E_d[n(E_2)p(E_1) - n(E_1)p(E_2)]\,dE_1\,dE_2 \tag{14.13}$$

E_d is related to the light intensity I_s by means of Eq. (13.11) as

$$E_d = \frac{I_s}{v}$$

where v is the velocity of light.

$$\left(\frac{dI_s}{dz}\right)_{\text{stim}} = I_s \iint B\frac{h\nu}{v}[n(E_2)p(E_1) - n(E_1)p(E_2)]\,dE_1\,dE_2 \tag{14.14}$$

where

$$\frac{d}{vdt} = \frac{d}{dz}$$

was used.

Next, the spontaneous emission will be calculated. The spontaneous emission is proportional to the product $n(E_2)p(E_1)$ of the density of electrons in the conduction band and the density of holes in the valence band. The increase in the photon density S by spontaneous emission is

$$\left(\frac{dS}{dt}\right)_{\text{spon}} = \iint An(E_2)p(E_1)\,dE_1\,dE_2 \tag{14.15}$$

where A is Einstein's A coefficient. Equation (14.15) can be converted into the change in the light intensity variation due to spontaneous emission in a similar manner as Eq. (14.14) was obtained from Eq. (14.10).

$$\left(\frac{dI_s}{dz}\right)_{\text{spon}} = \iint A\frac{h\nu}{v}n(E_2)p(E_1)\,dE_1\,dE_2 \tag{14.16}$$

The resultant light intensity variation is

$$\frac{dI_s}{dz} = \left(\frac{dI_s}{dz}\right)_{\text{stim}} + \left(\frac{dI_s}{dz}\right)_{\text{spon}} \tag{14.17}$$

and combining Eqs. (14.14) and (14.15) and rewriting in a manner similar to Eq. (13.20), the final differential equation of the semiconductor laser amplifier is

$$\frac{dI_s}{dz} = g(h\nu)I_s + h(h\nu) \tag{14.18}$$

where

$$g(h\nu) = B\iint \frac{h\nu}{v}[n(E_2)p(E_1) - n(E_1)p(E_2)]\,dE_1\,dE_2 \tag{14.19}$$

$$h(h\nu) = A\iint \frac{h\nu}{v}n(E_2)p(E_1)\,dE_1\,dE_2 \tag{14.20}$$

Let us examine Eq. (14.19). Insertion of Eqs. (14.2), (14.5), (14.8), and (14.9) into Eq. (14.19) leads to

$$g(h\nu) = B \iint \frac{h\nu}{v} g_c(E_2) g_v(E_1) [f_c(E_2) - f_v(E_1)] \, dE_1 \, dE_2 \tag{14.21}$$

The condition for positive gain is obtained by inserting Eqs. (14.4) and (14.7) into Eq. (14.21) as

$$E_2 - E_1 < F_c - F_v \tag{14.22}$$

This means the upper limit on the energy of $E_2 - E_1 = h\nu$ is $F_c - F_v$. When a forward bias current is applied, the difference $F_c - F_v$ is identical to the potential energy eV_b established by the external bias voltage V_b. The lower limit of the gain curve is determined approximately by the energy gap and, hence, the frequency range of the stimulated emission is

$$E_g < h\nu < eV_b \tag{14.23}$$

where

$$eV_b = F_c - F_v \tag{14.24}$$

Next, the gain curve will be obtained using a graphical method. E_1 in Eq. (14.21) is replaced by $E_1 = E_2 - h\nu$ to obtain

$$g(h\nu) = B \frac{h\nu}{v} \int g_c(E_2) g_v(E_2 - h\nu) [f_c(E_2) - f_v(E_2 - h\nu)] \, dE_2 \tag{14.25}$$

Figure 14.4 shows the distribution of the integrand with respect to energy E_2. $T = 0$ is assumed and the part of the integrand in the square brackets becomes either ± 1 or zero. Note that mathematically $g(h\nu)$ is the cross-correlation between $g_c(E_2)$ and $g_v(E_2 - h\nu)$, multiplied by the factor of $[f_c(E_2) - f_v(E_2 - h\nu)]$. This means $g(h\nu)$ is obtained by shifting one of the functions by $h\nu$ and integrating the product.

In order to calculate $g(h\nu)$, g_v and f_v are shifted toward the right. As soon as the amount $h\nu$ of shift exceeds $E_c - E_v$, the value of $g(h\nu)$ becomes nonzero because of the overlap of g_v and g_c. However, $g(h\nu)$ becomes zero and then negative as soon as the amount $h\nu$ of shift exceeds $F_c - F_v$. The value of $g(h\nu)$ becomes negative because of the change in the sign of $[f_c(E_2) - f_v(E_2 - h\nu)]$. The $g(h\nu)$ curve has a bell shape with the maximum near the center.

The calculated gain spectrum $g(h\nu)$ for $Ga_{0.47}In_{0.53}As$ [7] is shown in Fig. 14.5. The lower frequency cutoff does not match the bandgap, but slightly extends beyond it; this will be ignored for the time being.

Finally, an attempt will be made to recast the gain curve $g(h\nu)$. Such a recast makes the analysis of the semiconductor laser simpler. Figure 14.6 is a replot of the peak values of the gain curves in Fig. 14.5 as a function of the carrier density N. The replot is quite linear with $N - N_\alpha$, so that it is justifiable to write

$$g = a(N - N_\alpha) \tag{14.26}$$

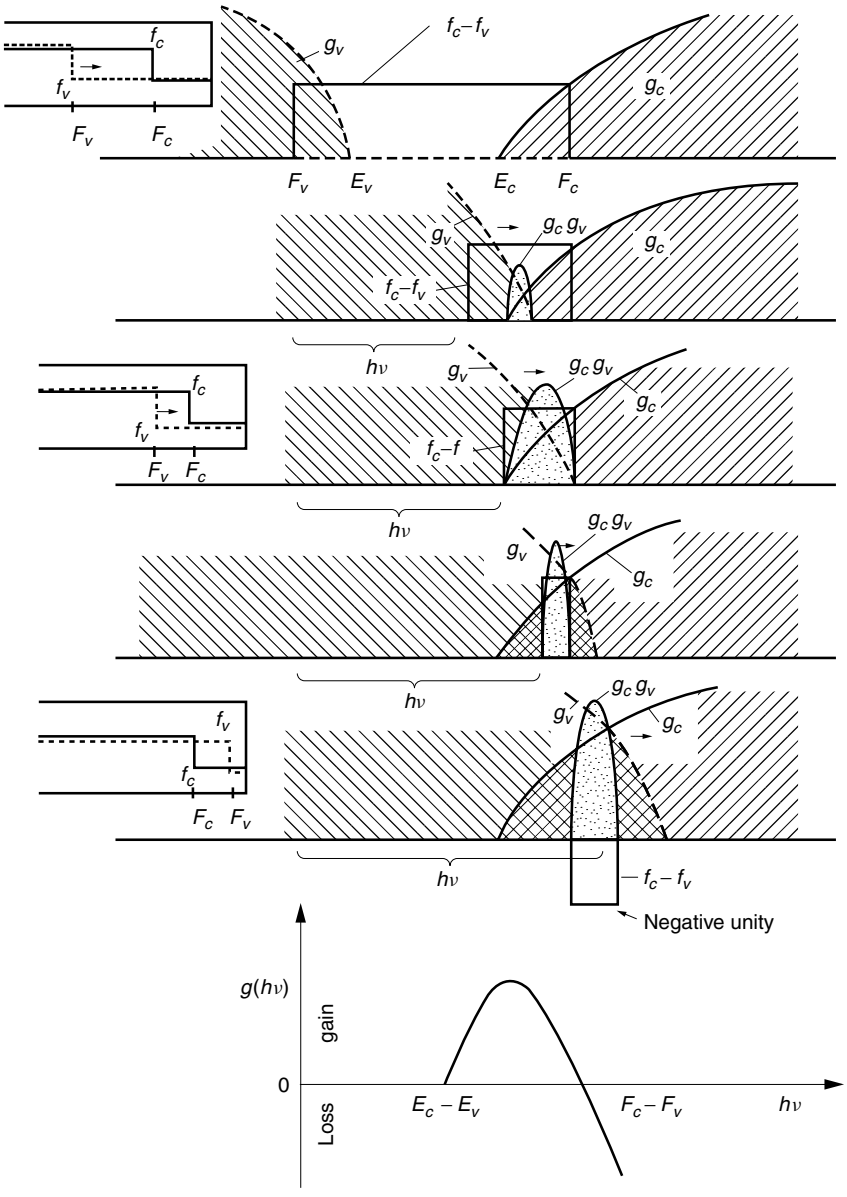


Figure 14.4 Graphical explanation of the calculated gain spectrum.

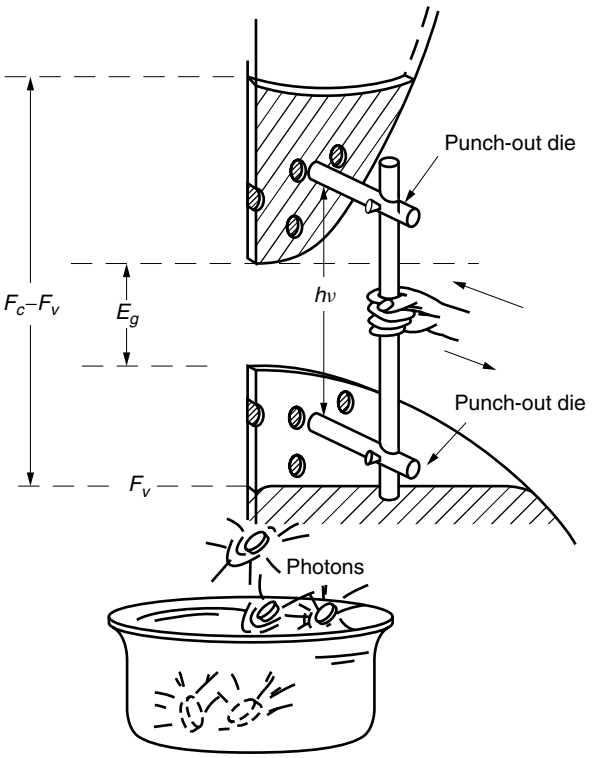
where g without the argument $(h\nu)$ is used to designate the replot of the peaks of $g(h\nu)$. The quantity a is the newly defined gain constant. N_α is the electron density that intercepts the horizontal axis; it is called the zero gain electron density of the medium. For those applications where the frequency behavior is not a prime concern, Eq. (14.26) is a good approximation. The power equation of the semiconductor laser becomes

$$\frac{dI_s}{dz} = gI_s + h \tag{14.27}$$

The general shape of the gain curves in Fig. 14.5 will be appreciated using the model of a "photon punch press" shown here. This machine is made of a pair of punch-out dies. In order to generate a photon by the recombination of an electron and a hole, the machine has to punch out an electron and a hole at the same time. The spacing between the punch-out dies can be adjusted for photons of a specific frequency.

Such a machine may help us to better visualize the following behavior of the gain curves in Fig. 14.5:

1. The lower and upper limits of $h\nu$ are E_g and $F_c - F_v$.
2. The fan shape for the material to be punched means the gain curves have their maxima slightly above the center of the positive gain region.
3. When the bias current is increased, and F_c is raised and F_v is lowered, the gain shifts toward a larger value of $h\nu$, and as well, the value of the maximum is increased.



Photon punch press.

14.2.2 Laser Cavity

If mirrors are installed at the ends of the light amplifier and are oriented so that the output from the amplifier reenters the amplifier and follows the same path as the first pass, this is called *optical feedback*. A small amount of light, initially generated by spontaneous emission, is fed back to generate stimulated emission, which in turn is fed back to further generate stimulated emission and the oscillation starts. This is a laser.

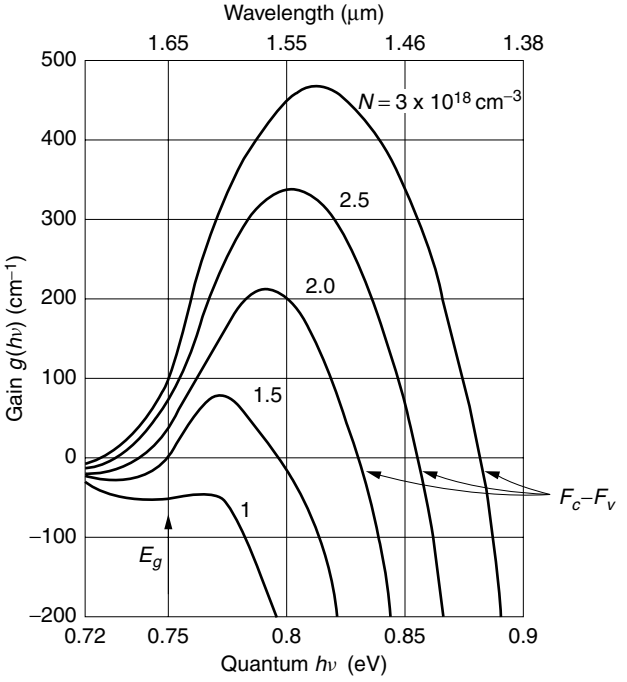


Figure 14.5 Gain spectrum of $Ga_{0.47}In_{0.53}As$. (After M. Asada and Y. Suematsu [7].)

The laser whose feedback light path is in a ring form is called a ring laser. The more common laser type uses end mirrors to reflect the light back and forth as a means of feedback. This laser can be very compact and is called a Fabry–Pérot (FP) type laser.

The amplifying medium of the laser is also referred to as the active medium. A structure that uses feedback to increase optical field intensities is known as an optical resonator. The laser is an optical resonator containing an active medium. Figure 14.7 shows the typical structure of a semiconductor laser. The active layer is the semiconductor p-n junction.

The forward bias current injects holes into the p-type layer on the top through a narrow metal stripe electrode, and electrons to the n-type layer through a metal electrode at the bottom. Recombination of the electrons and holes takes place in the narrow thin active layer. The generated light is confined inside the active layer because the index of refraction of the active layer is raised from that of the adjacent p- and n-type cladding layers.

Partially reflecting mirrors, which are simply cleaved surfaces or deposited mirrors, are placed on both ends of the active layer. The light that is reflected back and forth between these mirrors enhances the stimulated emission and becomes significantly larger than spontaneous emission. For each pass through the active layer, part of the light is reflected at the mirror, and the rest is transmitted through the mirror. The transmitted light is the laser output.

Artificial diamond whose heat conductivity is about three times as high as copper is often used as a heat sink.

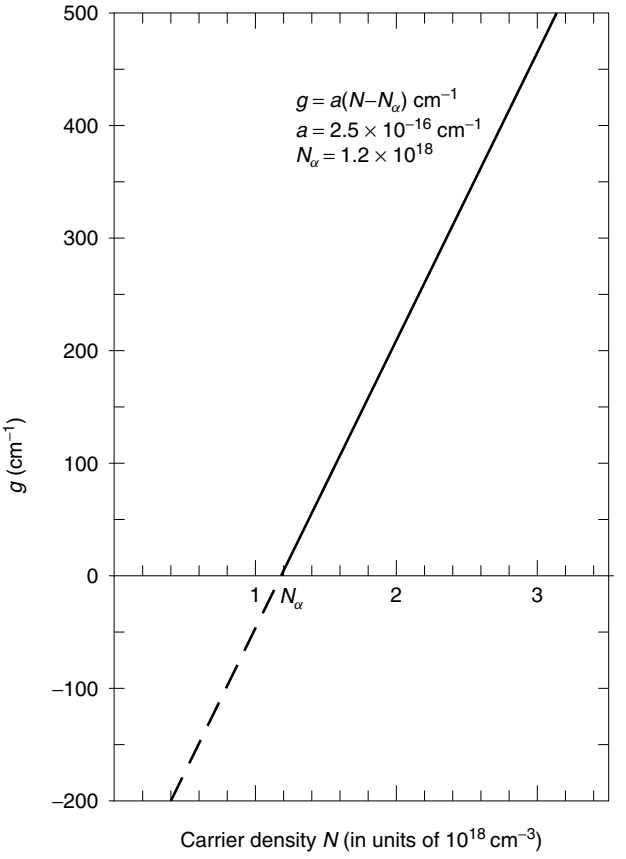


Figure 14.6 Linearized gain with respect to the carrier density N for $Ga_{0.47}In_{0.53}As$.

14.2.3 Conditions for Laser Oscillation

Figure 14.8 gives a heuristic illustration of a Fabry–Pérot type laser. As the light bounces back and forth in the active region, the light is amplified. From Fig. 14.8, the conditions for the wave to grow and finally oscillate are (1) the amplitude after each round trip has to be larger than before, and (2) the phase after each round trip has to be the same as the phase of the original trip.

14.2.3.1 Amplitude Condition for Laser Oscillation

The light travels back and forth in the active region between the end mirrors. The power of the amplification with distance will be obtained by solving Eq. (14.27) with $h \doteqdot 0$.

$$P = P_0 e^{gz}$$

In the real world, power is lost due to attenuation in the medium and escape at the cavity end mirrors. Taking these into account and writing in terms of amplitude rather than power, the amplification for one round trip is given by

$$E = E_0 R e^{(1/2)(g-\alpha)2L + j2\beta L} \tag{14.28}$$

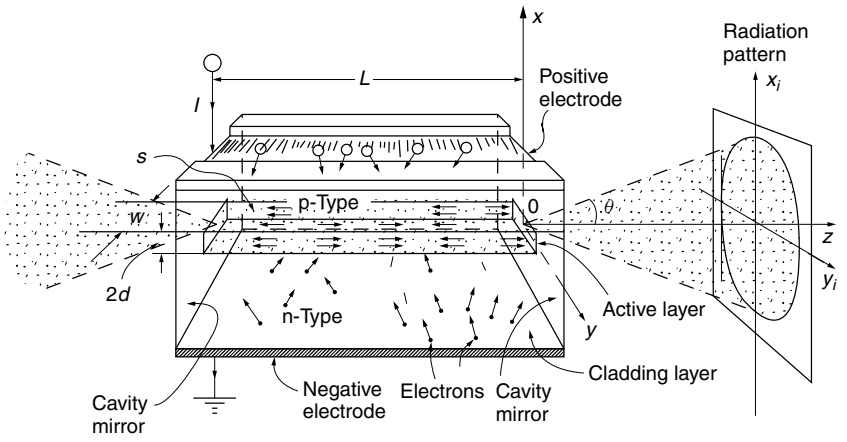


Figure 14.7 Semiconductor laser.

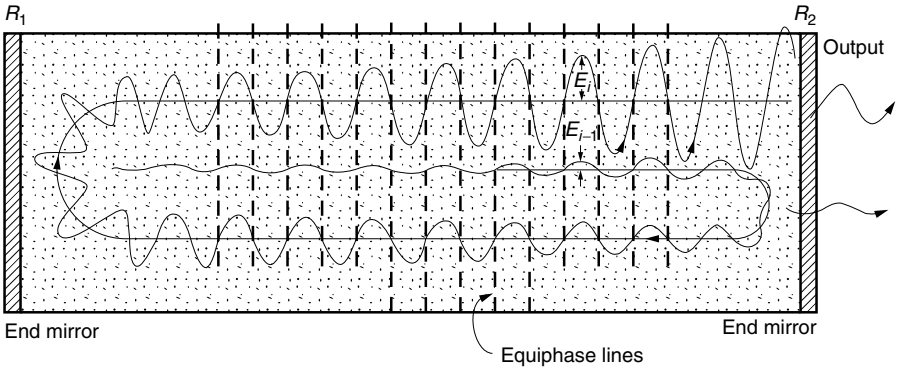


Figure 14.8 Two conditions for lasing action: (1) gain $E_i > E_{i-1}$ and (2) phase match $\phi_i = \phi_{i-1} + 2\pi n$.

where E_0 is the amplitude of the light before the trip, L is the length of the active region, α is the power attenuation constant, and β is the propagation constant. If the amplitude reflection coefficients of the mirrors are r_1 and r_2, then $R = r_1 r_2$.

Equation (14.28) is the amplitude E of the light after each round trip. The reason for the expression in terms of amplitude is that both amplitude and phase are necessary when analyzing the conditions for lasing. Note that both g and α are defined in terms of power rather than amplitude, and they are halved for expressing the amplitude.

There are two important conditions for the field to grow after each round trip. The first condition is that the gain has to be larger than unity

$$Re^{(g-\alpha)L} \geq 1$$

and

$$g \geq \frac{1}{L} \ln \frac{1}{R} + \alpha \tag{14.29}$$

In a medium with small gain, R has to be large. For instance, the He–Ne laser has a small gain $g = 10^{-4}$–10^{-3}cm^{-1}, so that the reflectivity R of the cavity mirrors has to be

$R = 0.998$. A GaAs semiconductor laser whose gain is as large as $g = 100–300\ \text{cm}^{-1}$ can afford a smaller reflectivity, and simple cleaved surfaces at the ends will suffice. The reflectance from a cleaved surface is

$$R = \left(\frac{n-1}{n+1}\right)^2 \tag{14.30}$$

where n is the index of refraction. For GaAs, $n = 3.5$ and $R = 0.31$.

The second condition for the field to grow concerns the phase.

14.2.3.2 Phase Condition for Laser Oscillation

So that the lightwave will constructively interfere, the phase after each round trip has to be an integral multiple of 2π, as shown in Fig. 14.8.

$$2\beta L = 2p\pi \tag{14.31}$$

where $p = 1, 2, 3, \ldots$, and β is the propagation constant in the active layer. Since the active layer is so thin that only the lowest order transverse mode is excited (see Section 9.4), the propagation constant β can be approximated as

$$\beta = \frac{2\pi}{\lambda_p} n \tag{14.32}$$

where n is the index of refraction of the active layer. Combining Eqs. (14.31) and (14.32) gives

$$\lambda_p = \frac{2nL}{p} \tag{14.33}$$

The wavelength λ_p is the wavelength of the pth cavity mode or pth longitudinal mode. The separation between adjacent longitudinal modes is

$$\Delta\lambda = -\frac{\lambda_p^2}{2nL} \tag{14.34}$$

where

$$\Delta\lambda = \lambda_p - \lambda_{p-1}$$

Equation (14.34) is an important quantity to be considered in fabricating a single longitudinal mode (SLM) laser.

Expressed in terms of frequency, Eq. (14.33) becomes

$$\nu_p = \frac{c}{2nL} p \tag{14.35}$$

and

$$\Delta\nu = \frac{c}{2nL} \tag{14.36}$$

where

$$\Delta\nu = \nu_p - \nu_{p-1}$$

The mode spacing decreases with an increase in cavity length.

14.2.4 Qualitative Explanation of Laser Oscillation

Figure 14.9 explains the same fact in a more qualitative manner. Figure 14.9a illustrates the frequencies that satisfy the condition for the round trips to be in phase. These frequencies are the resonant frequencies ν_p given by Eq. (14.35) of the cavity, as indicated by the notches along the horizontal axis in Fig. 14.9a. Figure 14.9b is a graph of the gain curves with respect to the frequency ν for the semiconductor laser with the injection current as a parameter. The envelopes of the gain curves are similar to those in Fig. 14.5, with a comb like structure that will be explained later.

The following describes the output of the laser as the injection current is increased. When the injection current is very small, the gain is low and the amplitude after each

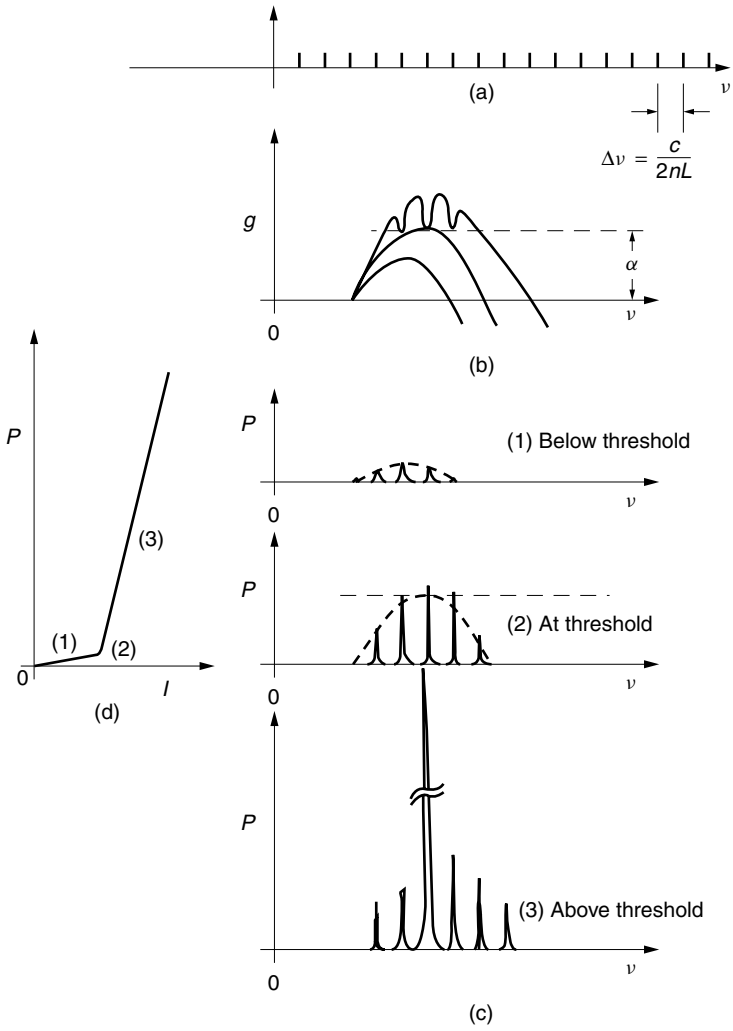


Figure 14.9 Gain curves and laser oscillation. (a) Resonance frequencies of the cavity. (b) Gain curves. (c) Output from the laser. (d) Power versus current characteristic of a laser.

round trip is smaller than the amplitude of the preceding trip because of the attenuation. Light with small amplitude is observed at the resonant frequencies of the cavity, as indicated in Fig. 14.9c(1), but the radiation is primarily due to spontaneous emission. As the injection current is increased, the height of the gain curve rises until it reaches the threshold value, namely, the gain is just canceled by the attenuation. At threshold, the amplitude after each round trip is the same as before. The output spectrum at threshold is shown in Fig. 14.9c(2). With a further increase in the injection current beyond the threshold, the gain g exceeds the attenuation α, and the amplitude of the light grows. What happens above threshold depends on the type of laser. The resonant frequencies are also called longitudinal modes. If the laser is a "quasi" single longitudinal mode laser, several modes may reach threshold, but only the mode that reaches threshold first keeps growing, as shown in Fig. 14.9c(3), and this mode clearly dominates the other lasing modes. In a truly multimode laser, the resonant frequencies are closely spaced and several will reach threshold at about the same time. These modes will begin to oscillate, and there is no large dominance of one mode over its neighbors. In a true single-mode laser, only one mode oscillates.

The output light power P versus bias current is shown in Fig. 14.9d. The output light is practically null until the bias current reaches the threshold current. Thereafter, the output light almost linearly increases with the bias current.

14.3 RATE EQUATIONS OF SEMICONDUCTOR LASERS

The rate equations for the semiconductor laser [8] are slightly different from those developed for the EDFA in the previous chapter, where both pump and signal powers were light powers. With the semiconductor laser, electrons are used for the pump and photons for the signal.

Electrons injected into the semiconductor junction in the form of a current are eventually converted into the emission of photons. The rate equations are differential equations that relate these two quantities: injected carrier density and photon density. The rate equations are used to obtain quantities such as the threshold current, the turn-on delay, and the upper limit on modulation frequency.

In previous sections that dealt with trying to find the power of amplification, the derivative with respect to length was used. However, in a laser, one is usually not interested in keeping track of how many times the light goes back and forth for amplification. The derivative with respect to time is often the more practical approach in laser systems.

$$\underbrace{\frac{dN}{dt}}_{\substack{\text{Rate of electrons introduced into unit volume per unit time}}} = \underbrace{\frac{J}{2ed}}_{\substack{\text{Number of electrons injected into unit volume per unit time}}} - \underbrace{Sb(N - N_\alpha)}_{\substack{\text{Number of electrons depleted by stimulated emission of photons per unit volume per unit time}}} - \underbrace{\frac{N}{\tau_n}}_{\substack{\text{Number of electrons depleted by spontaneous emission or escaping from the action region per unit volume per unit time}}} \tag{14.37}$$

$$\underbrace{\frac{dS}{dt}}_{\substack{\text{Rate of photons generated per unit volume per unit time}}} = \underbrace{Sb(N - N_\alpha)}_{\substack{\text{Number of photons generated by stimulated emission per unit volume per unit time}}} - \underbrace{\frac{S}{\tau_s}}_{\substack{\text{Number of photons escaping from unit volume per unit time}}} + \underbrace{\beta_s \frac{N}{\tau_n}}_{\substack{\text{Spectrum of spontaneous emission that overlaps with that of the stimulated emission}}} \tag{14.38}$$

where

$2d$ = thickness of the active region
N = electron density
S = photon density
g = gain with respect to distance
a = gain constant with respect to distance
b = gain constant with respect to time and equal to va
J = injection current density
τ_n = lifetime of the electrons in the conduction band before being lost by spontaneous emission or escape from the active region
τ_s = lifetime of photons inside the cavity before going out of the cavity or being absorbed inside the cavity
β_s = fraction of spontaneously emitted photons whose spectrum coincides with that of the stimulated emission

Equation (14.37) is the rate of increase in conduction band electrons per unit volume per unit time. The rate is the contribution made by the injection current minus electrons recombined by stimulated emission minus the electrons lost by spontaneous emission and by escaping from confinement. Equation (14.38) is the rate of photons generated per unit volume per unit time, that is, the stimulated emission minus the photons that have escaped from the cavity plus the contribution of the spontaneous emission whose spectrum coincides with the stimulated emission. The last term is usually very small and is normally ignored ($\beta_s \approx 10^{-4}$–10^{-6}). In the following sections, the various quantities that can be derived from the rate equations will be presented.

The quantities τ_s, b, and J appearing in Eqs. (14.37) and (14.38) need further explanation. The lifetime τ_s of the photons inside the laser cavity is the inverse of the rate of disappearance of photons out of the laser cavity. For t seconds in the cavity, the light makes $vt/2L$ round trips. For each round trip, the light power is reduced by R^2 due to reflections at both end mirrors. For t seconds in the cavity, the light travels a distance vt, so that the power loss due to attenuation in the medium is $e^{-\alpha vt}$. Both losses may be represented by one parameter τ_s when $vt/2L$ is large. The lifetime τ_s satisfies

$$e^{-t/\tau_s} = (R^2)^{vt/2L} e^{-\alpha vt}$$

and hence

$$\frac{1}{\tau_s} = v\left(\frac{1}{L}\ln\frac{1}{R} + \alpha\right) \tag{14.39}$$

where v is the velocity in the active region and the identity $R = e^{\ln R}$ was used.

Next, b is explained. Using Eq. (14.26) and the relation $z = vt$, the first term of Eq. (14.27) can be rewritten as

$$\frac{dP}{dt} = b(N - N_\alpha)P \tag{14.40}$$

where

$$b = va \tag{14.41}$$

Next, a few words are added about the current density J. As shown in Fig. 14.7, the injection current I (not light intensity) is fed from the top electrode into the active region whose cross-sectional area is s and thickness is $2d$. The number of electrons brought into the active region per unit time is I/e, and the number N of electrons injected into unit volume during τ_n seconds therefore is $(I/es2d)\ \tau_n$ and

$$N = \frac{J}{2ed} \cdot \tau_n \tag{14.42}$$

where $J = I/s$ is the current per unit cross section or the current density in the active region and where the thickness of the active region is so short that the electron density was assumed uniform.

14.3.1 Steady-State Solutions of the Rate Equations

The steady-state solutions of Eqs. (14.37) and (14.38) will be obtained. The steady-state condition is

$$\frac{dN}{dt} = \frac{dS}{dt} = 0 \tag{14.43}$$

The threshold electron density and current and the output light power from the laser will be found.

14.3.2 Threshold Electron Density and Current

The electron density and the corresponding current that just initiates lasing action will be found. Before lasing, the photon density is

$$S_s \doteqdot 0$$

and from Eqs. (14.37) and (14.43) the electron density is

$$N = \frac{\tau_n}{2ed} J \tag{14.44}$$

The electron density linearly increases with the injection current density before lasing. As soon as the lasing starts and $S \neq 0$, from Eq. (14.38), the lasing steady-state condition implies

$$\left[b(N - N_\alpha) - \frac{1}{\tau_s} \right] S_s = 0 \tag{14.45}$$

where β_s was considered small and S_s is the steady-state photon density. The condition that Eq. (14.45) be satisfied even after lasing, when $S \neq 0$, is that the quantity inside the square brackets in Eq. (14.45) stays zero and

$$N = N_\alpha + \frac{1}{b\tau_s} \tag{14.46}$$

Note that Eq. (14.46) is independent of injection current. This means that, before lasing, N grows with J as in Eq. (14.44), but as soon as the lasing starts and then the steady state is reached, N stops growing and stays at this value, as indicated in Fig. 14.10. Physically speaking, after lasing starts, no sooner are the electrons injected into the active region than they recombine with holes and release photons. The electrons never have a chance to pile up, and the electron density stays at the steady-state value as in Eq. (14.46). This fact is explained by the comb-shaped gain function in Fig. 14.9b. The threshold electron density N_{th} is therefore

$$N_{\text{th}} = N_\alpha + \frac{1}{b\tau_s} \tag{14.47}$$

and the threshold current J_{th} that is needed for N to reach N_{th} is, from Eq. (14.44),

$$\frac{J_{\text{th}}}{2ed} = \frac{N_{\text{th}}}{\tau_n} \tag{14.48}$$

or, with Eq. (14.47),

$$\frac{J_{\text{th}}}{2ed} = \frac{1}{\tau_n}\left(N_\alpha + \frac{1}{b\tau_s}\right) \tag{14.49}$$

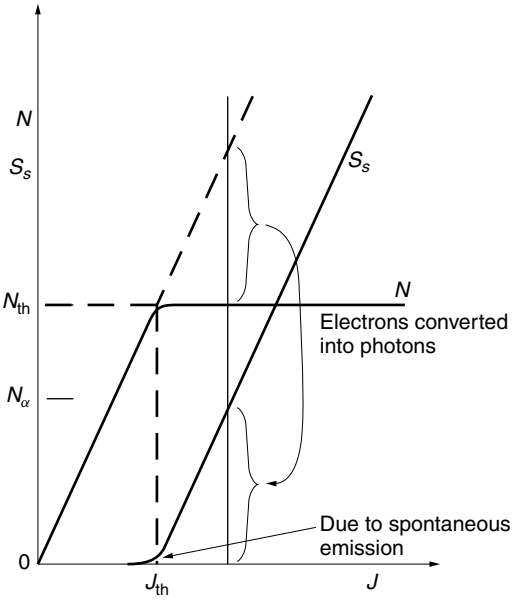


Figure 14.10 Photon and electron densities with respect to injection current J.

The right-hand side of Eq. (14.48) is associated with τ_n, which means that N_{th}/τ_n is the number of electrons either leaked out of the active region or converted into photons by spontaneous emission. The quantity J_{th} is the current needed to supply these lost electrons. It will soon become clear that it is important to keep J_{th} as small as possible for better laser performance.

14.3.3 Output Power from the Laser

The steady-state light output from the laser for a given injection current will be calculated. The steady state of Eqs. (14.37) and (14.38) gives

$$\frac{J}{2ed} = \frac{S_s}{\tau_s} + \frac{N}{\tau_n} \tag{14.50}$$

The right-hand side consists of the number of electrons converted into photons by stimulated emission and the number of electrons leaked out of the active region. As soon as lasing starts, N stops growing and stays at the value specified by Eq. (14.47) and

$$N = N_{\text{th}} \tag{14.51}$$

With Eqs. (14.48) to (14.51), S becomes

$$S_s = \frac{\tau_s}{2ed}(J - J_{\text{th}}), \quad J \geq J_{\text{th}} \tag{14.52}$$

S_s increases linearly with $(J - J_{\text{th}})$. If Eq. (14.52) is rewritten as

$$\frac{S_s}{\tau_s} = \frac{J - J_{\text{th}}}{2ed} \tag{14.53}$$

then the left-hand side is the rate of photon emission, and the right-hand side is the rate of electrons injected by the current above J_{th}. The electron density and the photon density are plotted as a function of injection current in Fig. 14.10. While the injection current is below the threshold, only the electron density increases with the injection current, and the photon density stays at zero. Once the injection current exceeds the threshold value, the situation reverses, and the photon density increases linearly with the injection current, while the electron density stays put at $N = N_{\text{th}}$.

It is interesting to check what happens to N if Eqs. (14.37) and (14.38) are used to calculate N beyond the threshold:

$$N = \tau_n \left(\frac{J}{2ed} - \frac{S_s}{\tau_s} \right) \tag{14.54}$$

The second term in the parentheses is the portion that is used for the stimulated emission. The rate of generation of S is replaced by Eq. (14.52), and Eq. (14.54) becomes

$$N = N_{\text{th}}$$

where Eq. (14.48) was also used. This confirms the earlier result of $N = N_{\text{th}}$ after lasing.

Next, an expression for the output power will be found. Equation (14.38) with $\beta_s = 0$ and $dS/dt = 0$ means that the number of photons generated inside the laser by stimulated emission (first term on the right-hand side) balances the number of photons leaked out and lost in the cavity (second term). The second term has two components, one component being the light leakage from the mirrors, that is, the laser output itself, and the other component being light lost due to attenuation in the cavity. The contribution S_s/τ_s for only the portion that leaked through the mirrors is found from Eq. (14.39) with $\alpha = 0$ as

$$\frac{1}{\tau_c} = \frac{v}{L} \ln \frac{1}{R} \tag{14.55}$$

If V is the volume of the active region, then VS_s/τ_c is the number of photons per second that escape from the active region of the laser into the outside world. The output light power from the cavity is therefore

$$P = h\nu V \frac{S_s}{\tau_c} \tag{14.56}$$

Insertion of Eq. (14.52) into (14.56) and rewriting the current density J in terms of the current I gives

$$P = \frac{h\nu}{e} \frac{\tau_s}{\tau_c} (I - I_{\text{th}}) \tag{14.57}$$

If $h\nu$ is approximated by the bandgap energy E_g in electron volts, then the output power is

$$P = E_g \frac{\tau_s}{\tau_c} (I - I_{\text{th}}) \tag{14.58}$$

Thus, the light output increases linearly with $I - I_{\text{th}}$ as indicated in Fig. 14.9d.

14.3.4 Time-Dependent Solutions of the Rate Equations

The physical quantities associated with the time-dependent solutions of Eqs. (14.37) and (14.38) will be found.

14.3.4.1 Turn-On Delay

A laser does not start lasing until the electron density reaches the threshold value, so that there is a turn-on delay. The turn-on delay time t_d for N to reach N_{th} will be calculated. Before lasing $S = 0$, and Eq. (14.37) becomes

$$\frac{dN}{dt} = \frac{J}{2ed} - \frac{N}{\tau_n} \tag{14.59}$$

and the turn-on delay is

$$t_d = \int_0^{N_{\text{th}}} \frac{dN}{\dfrac{J}{2ed} - \dfrac{N}{\tau_n}} \tag{14.60}$$

Integration gives

$$t_d = \tau_n \ln\left(\frac{J}{J - J_{\text{th}}}\right) \tag{14.61}$$

where Eq. (14.48) was used. When a step injection current such as shown in Fig. 14.11a is applied to the laser diode, N starts increasing as shown in Fig. 14.11b, and the laser diode does not start lasing until t_d seconds later as shown in Fig. 14.11c. From Eq. (14.61), one way of shortening the turn-on delay time is to raise the injection current. Equation (14.61) is often used to experimentally determine the value of τ_n. The logarithmic curve of $\ln[J/(J - J_{\text{th}})]$ versus t_d gives τ_n as its slope. The quantity τ_n is normally a few nanoseconds.

Pulsed modulation of the light is achieved by switching the bias current on and off. As the pulse rate is increased, the turn-on delay becomes a problem and limits the maximum bit rate. The laser with a lower threshold current has a shorter turn-on delay time, and this is one of the reasons for the desirability of a laser with a low threshold current mentioned earlier.

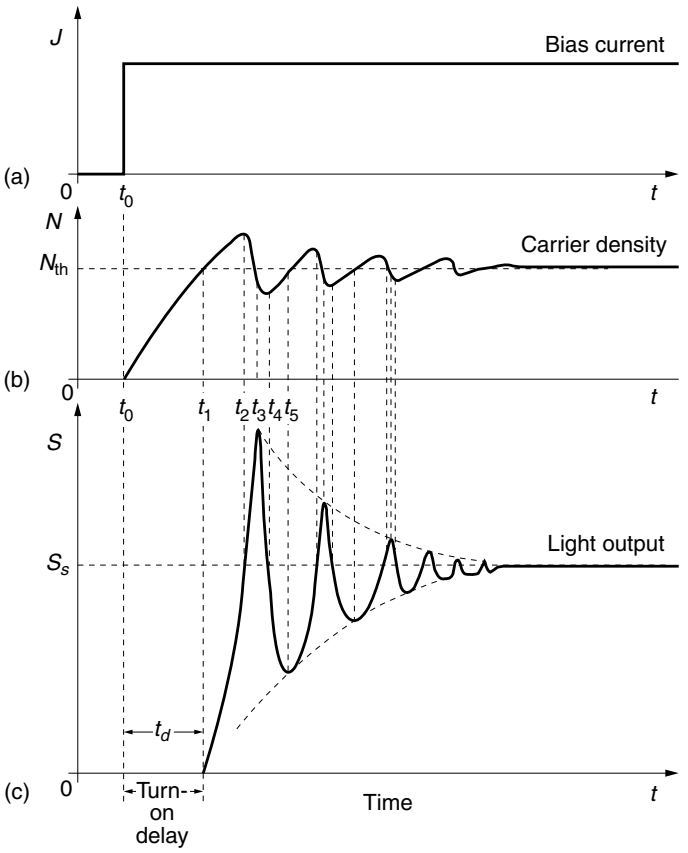


Figure 14.11 Explanation of the turn-on delay and the relaxation oscillation of a semiconductor laser. (a) Step injection current. (b) Carrier density. (c) Light output.

Another way to reduce t_d is biasing the injection current to $J = J_b$. The turn-on delay for such a case is calculated by changing the lower limit of the integral in Eq. (14.60):

$$t_d = \tau_n \ln\left(\frac{J - J_b}{J - J_{\text{th}}}\right) \tag{14.62}$$

A disadvantage of the biasing, however, is that as J_b approaches the threshold, the amount of emission (see Fig. 14.9 below threshold) for the off state of the pulse increases and the extinction ratio, which is the ratio of the light intensities for on and off states, becomes worse.

14.3.4.2 Relaxation Oscillation

When a laser diode is turned on, a transient decaying sinusoidal oscillation occurs as shown in Fig. 14.11. This oscillation is called *relaxation oscillation*. Along with the turn-on delay time, the relaxation oscillation limits the bit rate of the pulse modulation of the laser.

Referring to Fig. 14.11, let us examine the sequence of changes in the electron as well as the photon densities when a step injection current is applied.

1. As soon as the injection current is turned on, the electron density N starts to increase in accordance with Eq. (14.37) with $S = 0$. No lasing takes place until N reaches N_{th}.
2. When N reaches N_{th}, S begins to respond; however, the response of S is not instantaneous. It takes a few picoseconds for the field to build up in the cavity, and during this build-up time, N overshoots N_{th}.
3. Because N overshoots N_{th}, the rate of increase in S is further accelerated. As soon as S overshoots the steady-state value S_s, J cannot catch up with the consumption of N, and N starts to decay at $t = t_2$. As long as N is larger than N_{th}, S keeps on increasing but at a slower rate. However, as soon as N drops below N_{th} at $t = t_3$, S starts decreasing.
4. When S decays to smaller than S_s at $t = t_4$, N starts increasing because fewer electrons are participating in the stimulated emission.
5. As soon as N reaches N_{th}, S starts increasing and the second cycle of the oscillation begins.
6. In the second cycle of oscillation, S does not have to start from zero, and the time it takes to reach the steady state is less than in the first cycle. Thus, N has less time to go beyond N_{th} than in the first cycle. The amplitude of oscillation in the second cycle is therefore less than that of the first cycle.

14.3.5 Small Signal Amplitude Modulation

One of the biggest advantages of using a semiconductor laser is the ease with which the output is amplitude modulated. It can be modulated by simply modulating the injection current. In this section the modulation characteristics will be investigated using the rate equations.

A steady injection current J_s, set above the threshold current, is applied to the laser diode. A modulation current j is superimposed on the steady value J_s:

$$J = J_s + j \tag{14.63}$$

where

$$j << J_s \tag{14.64}$$

Both N and S deviate from their steady-state values.

$$N = N_{\text{th}} + n \tag{14.65}$$

$$S = S_s + s \tag{14.66}$$

The steady-state values satisfy these two equations:

$$\frac{J_s}{2ed} - S_s b(N_{\text{th}} - N_\alpha) - \frac{N_{\text{th}}}{\tau_n} = 0 \tag{14.67}$$

$$b(N_{\text{th}} - N_\alpha) - \frac{1}{\tau_s} = 0 \tag{14.68}$$

Insertion of Eqs. (14.63) through (14.66) into Eqs. (14.37) and (14.38) and then simplification with Eqs. (14.67) and (14.68) leads to

$$\frac{dn}{dt} = \frac{j}{2ed} - \left(bS_s + \frac{1}{\tau_n}\right) n - b(N_{\text{th}} - N_\alpha)s \tag{14.69}$$

$$\frac{ds}{dt} = bS_s n \tag{14.70}$$

where Eq. (14.68) was used twice for S_s as well as for s and the term with the product ns was ignored. By taking the derivative, the equations for n and s can be separated as

$$\left[\frac{d^2}{dt^2} + \gamma\frac{d}{dt} + \omega_r^2\right]\begin{bmatrix} n \\ s \end{bmatrix} = \begin{bmatrix} \dfrac{1}{2ed}\dfrac{dj}{dt} \\ \dfrac{bS_s}{2ed} j \end{bmatrix} \tag{14.71}$$

where

$$\gamma = \left(bS_s + \frac{1}{\tau_n}\right) \tag{14.72}$$

$$\omega_r^2 = b^2 S_s (N_{\text{th}} - N_\alpha) \tag{14.73}$$

Useful quantities will be obtained by solving Eq. (14.71) in the next few sections.

14.3.5.1 Time Constant of the Relaxation Oscillation

If the relaxation oscillation is to be analyzed properly, then the differential equation (14.71) has to be solved with the proper initial conditions. But if one is simply interested in finding the frequency or decay time of the free oscillation, then one need

only find the solution of the differential equation with the right-hand side of Eq. (14.71) set to zero:

$$\left(\frac{d^2}{dt^2} + \gamma\frac{d}{dt} + \omega_r^2\right) s = 0 \tag{14.74}$$

Let us assume a solution of the form

$$s = s_0 e^{-j\omega t} \tag{14.75}$$

Inserting Eq. (14.75) into (14.74) gives

$$(-\omega^2 - j\gamma\omega + \omega_r^2) s_0 e^{-j\omega t} = 0 \tag{14.76}$$

and the value of ω that satisfies Eq. (14.76) is

$$\omega = -j\frac{\gamma}{2} \pm \sqrt{\omega_r^2 - \left(\frac{\gamma}{2}\right)^2} \tag{14.77}$$

The final solution is

$$s = s_0 \exp\left[-\frac{\gamma}{2}t \pm j\omega_r\sqrt{1 - \left(\frac{\gamma}{2\omega_r}\right)^2}\, t\right] \tag{14.78}$$

The relaxation oscillation decays as $e^{-(\gamma/2)t}$, where γ is given by Eq. (14.72).

Equation (14.72) for γ will be rewritten in terms of the operational and physical constants associated with the laser. If $N_{\text{th}} \gg N_\alpha$ is assumed, Eq. (14.68) becomes

$$bS_s = \frac{S_s}{\tau_s N_{\text{th}}} \tag{14.79}$$

Inserting Eqs. (14.48) and (14.53) into Eq. (14.79) gives

$$bS_s = \frac{1}{\tau_n}\frac{J - J_{\text{th}}}{J_{\text{th}}} \tag{14.80}$$

Insertion of Eq. (14.80) into (14.72) finally gives

$$\gamma = \frac{1}{\tau_n}\left(\frac{J}{J_{\text{th}}}\right) \tag{14.81}$$

This result indicates that the relaxation oscillation can be shortened by choosing a larger bias current J and by choosing $\tau_n J_{\text{th}}$ small. Let's further rewrite $\tau_n J_{\text{th}}$ by using parameters involving the geometry of the laser. From Eq. (14.49) the quantity $\tau_n J_{\text{th}}$ becomes

$$\tau_n J_{\text{th}} = 2ed\left(N_\alpha + \frac{1}{b}\frac{1}{\tau_s}\right) \tag{14.82}$$

where τ_s in Eq. (14.39) is repeated here,

$$\frac{1}{\tau_s} = v\left(\frac{1}{L}\ln\frac{1}{R} + \alpha\right) \tag{14.83}$$

Thus, $\tau_n J_{\text{th}}$ can be reduced, or equivalently the relaxation oscillation can be shortened by reducing any of the following: the thickness $2d$ of the active layer, N_α, $(1/L)\ln(1/R)$, or α.

14.3.5.2 Amplitude Modulation Characteristics

Let the laser be modulated at frequency ω_m by modulating the injection current density around the bias current J_0,

$$j = j_0 e^{-j\omega_m t} \tag{14.84}$$

The quantity bS_s on the right-hand side of Eq. (14.71) can be rewritten as $bS_s = \omega_r^2/b(N_{\text{th}} - N_\alpha)$ using Eq. (14.73). Combining this result and Eq. (14.68), Eq. (14.71) leads to

$$\frac{d^2 s}{dt^2} + \gamma\frac{ds}{dt} + \omega_r^2 s = \frac{\tau_s \omega_r^2}{2ed} j_0 e^{-j\omega_m t} \tag{14.85}$$

which is the forced vibration equation. The steady-state solution (a particular solution) can be found similar to Eq. (14.75) by assuming a solution

$$s = s_0 e^{-j\omega_m t} \tag{14.86}$$

Inserting Eq. (14.86) into (14.85) gives

$$s = \frac{\tau_s \omega_r^2 j_0}{2ed} \cdot \frac{1}{\sqrt{(\omega_r^2 - \omega_m^2)^2 + (\gamma\omega_m)^2}} e^{-j(\omega_m t + \phi)} \tag{14.87}$$

The magnitude of the modulated light given by Eq. (14.87) has a resonance. The resonance peak appears where the denominator of Eq. (14.87) is at the minimum, which is found by setting the derivative with respect to ω_m to be zero.

$$\omega_m^r = \omega_r\sqrt{1 - \frac{1}{2}\left(\frac{\gamma}{\omega_r}\right)^2} \tag{14.88}$$

By comparing Eq. (14.89) with (14.81), it is seen that $(\gamma/\omega_r)^2$ is much smaller than unity because τ_s is of the order of 10^{-12}s and τ_n is of the order of 10^{-9}s. Thus, the usable modulation bandwidth is more or less the same as ω_r. From a practical point of view, it is important to make ω_m^r or ω_r as large as possible.

As was done for γ, let us express ω_r in terms of the laser operational and physical parameters. The ω_r^2 in Eq. (14.73) can be split into the product of $bS_s \cdot b(N_{\text{th}} - N_\alpha)$. The bS_s term was already found in Eq. (14.80). The $b(N_{\text{th}} - N_\alpha)$ term is found from Eq. (14.68). Thus,

$$\omega_r^2 = \frac{J - J_{\text{th}}}{\tau_s \tau_n J_{\text{th}}} \tag{14.89}$$

The product $\tau_s \tau_n$ appearing in the denominator of Eq. (14.89) is the product of the time for the injected electrons to disappear by spontaneous emission (of the order of nanoseconds) and the time for the photons created by the stimulated emission to disappear after bouncing back and forth between the end mirrors (of the order of picoseconds). The maximum modulation frequency can be achieved by choosing J large, and making both τ_s and $\tau_n J_{\text{th}}$ small. The photon lifetime τ_s and the product $\tau_n J_{\text{th}}$ become small if $2d$, N_α, $(1/L)\ln(1/R)$, or α are made small, as mentioned in the previous section.

Figure 14.12 gives an example of the measured results [9]. The solid lines correspond to a laser with a 280-μm long cavity, and the dashed lines correspond to a laser with an 80-μm long cavity. The amplitudes of both curves are normalized by the value at $\omega_m = 0$. The laser with the shorter cavity has the shorter photon lifetime and hence a higher limit on the modulation frequency. With both lasers, the modulation frequency limits always increase with an increase in the injection current, as predicted by Eq. (14.89).

14.3.5.3 Comparisons Between Theoretical and Experimental Results

The theories presented thus far will be tested by two practical examples.

Example 14.1 Figure 14.13a is the vendor's specifications for an $Al_x\ Ga_{1-x}\ As$ laser diode, and Figs. 14.13b and 14.13c are the experimental results for the turn-on characteristics and the spectrum of this same laser diode. Calculate the theoretical bias voltage to injection current characteristics (V–I curve) and compare it with the experimental curve. The cavity mirrors are cleaved surfaces. The power attenuation constant is $\alpha = 175\text{cm}^{-1}$. The terminal contact resistance is $R = 0.5\ \Omega$. The following step-by-step instructions are intended to lead to the solution.

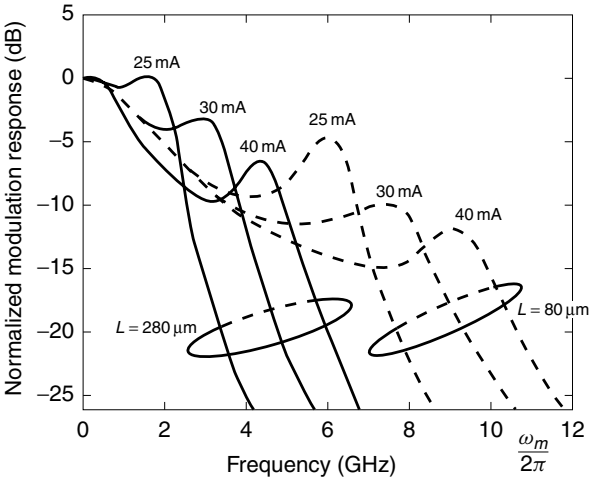


Figure 14.12 Amplitude modulation characteristics of the direct modulation of a laser diode. Solid line $L = 280$ μm. Dashed line $L = 80$ μm. (After R. Tucker, C. Lin, and C. A. Burrus [9].)



LAS DIODE

TYPE LCW-10
TYPE LCW-10F
with FIBER PIGTAIL
MULTIMODE
GaAlAs CW
INJECTION LASER DIODE

FEATURING:
- CW OPERATION
- UP TO 14 mW OUTPUT POWER ON SELECTED UNITS
- CHOICE OF WAVELENGTH
- HIGH TEMPERATURE OPERATION
- LOW THRESHOLD CURRENT

DESCRIPTION: The LCW-10 is a multimode CW GaAlAs double heterostructure laser diode. This device features low threshold current and high output power. The diodes are passivated for long life and reliability. They offer good performance at elevated temperatures. The LCW-10F is the fiber coupled version of the LCW-10. This unit is available with standard fiber as noted below. Custom fiber attachment available upon request.

Pom. I.. & I.. Information furnished with each device — Data at 27°C unless otherwise noted.

		Symbol	Min.	Typ.	Max.	Units
Total Radiant	LCW-10	Pom	4	7		mW
Flux at Rated I_1	LCW-10F*	Pom	1	2.0		mW
Peak Wavelength		λ_0	See Below	830	See Below	nm
Spectral Width @ 50% points		$\Delta\lambda$		2.5		nm
Source Size				0.2×7.0		μm
Rise Time of Radiant Flux		T_1		100	800	ps
Far Field Beam Divergence (LCW-10) @ 50% Points		$\theta_1 \times \theta_n$		10×35		degrees
Threshold Current		I_n		30	150	mA
Operating Current		I_v		115	$I_m + 50$	mA
Differential Quantum Efficiency				0.3		mW/mA
Forward Voltage at I_1		V_v		2.0		volts
Operating Temperature		T_n	0		60**	°C
Storage Temperature		T_v	–55		125	°C

*Standard fiber Digtail is all glass. 50 μm graded index core. 125 μm O.D. with N.A. = .20
**Selections to higher operating temperatures are available.

WAVELENGTH SELECTION

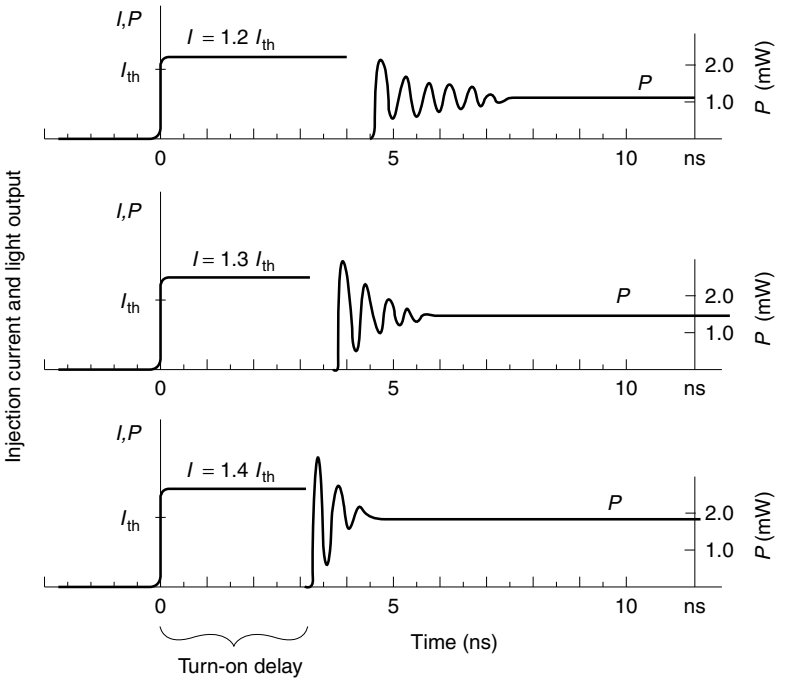
OPTION	WAVELENGTH RANGE
A	800 to 810 nm
B	810 to 840 nm
C	840 to 860 nm
D	860 to 880 nm

ABSOLUTE MAXIMUM RATINGS (CW)

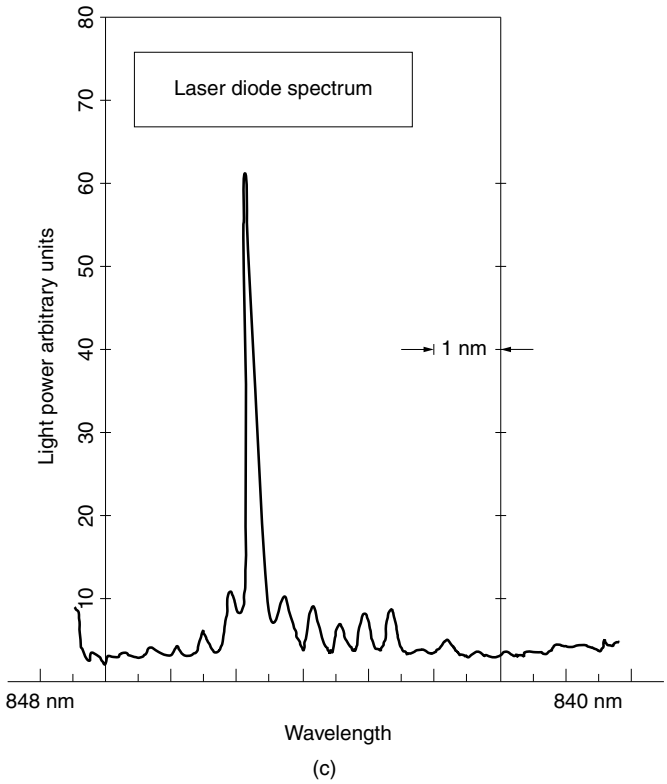
Maximum Forward Current = lower of $I_m + 60$ or 210 mA
Maximum Reverse Voltage = 2.0 volts
Maximum Operating Temperature = 70°C
Maximum Storage Temperature = 140°C
Minimum Storage Temperature = 75°C
Maximum Radiant Output Power = 14 mW

(a)

Figure 14.13 Comparison among vendor's specification, experimental results, and theoretical calculations for an $Al_xGa_{1-x}As$ laser diode. (a) Laser diode specifications. (b) Turn-on characteristics of the laser diode. (c) Measured spectrum of the laser diode. (d) Gain curve of $Al_xGa_{1-x}As$. (After H. C. Casey and M. B. Panish [10]). (e) Properties of $Al_xGa_{1-x}As$ crystal.

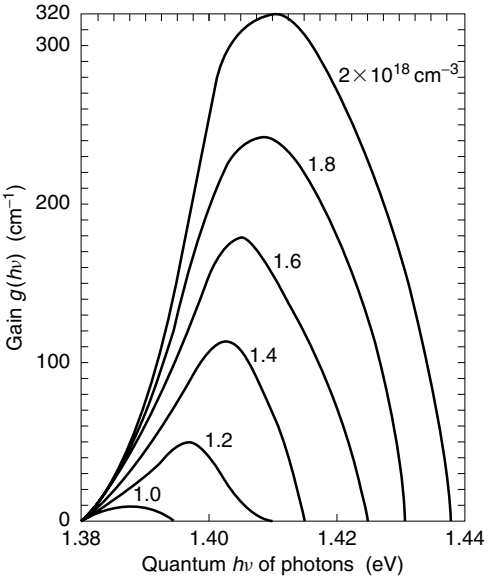


(b)

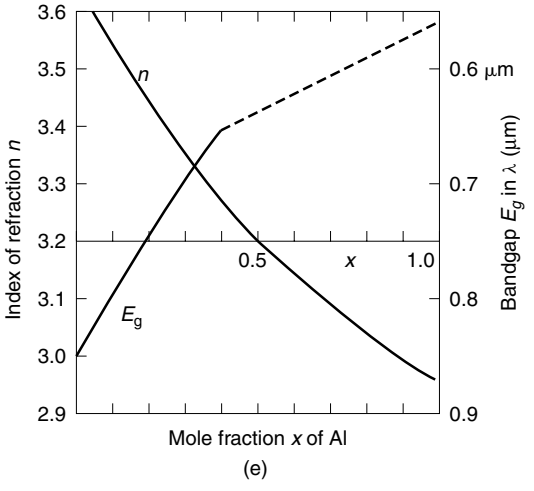


(c)

Figure 14.13 (*Continued*)



(d)



(e)

Figure 14.13 (*Continued*)

(a) Derive an expression that relates the carrier density and the quasi-Fermi levels. Approximate the Fermi–Dirac distribution by the Boltzmann distribution

$$f_c(E_2) = e^{-(E_2 - F_c)/kT} \tag{14.90}$$

so that the integral of Eq. (14.2) can be calculated in a closed form using

$$\int_0^\infty x^{1/2} e^{-x}\, dx = \frac{\sqrt{\pi}}{2} \tag{14.91}$$

(b) The calculations have to be done separately for carrier densities above and below the threshold carrier density. The threshold carrier density is determined by the spontaneous emission lifetime, which is obtainable from the turn-on delay time and the cavity length. The cavity length is obtained from the laser spectrum. The physical and optical properties of $Al_x\ Ga_{1-x}\ As$, which are controlled by the molar fraction x of Al, are

$$m_e^* = (0.067 + 0.082x)m_0 \tag{14.92a}$$

$$m_h^* = (0.5 + 0.29x)m_0 \tag{14.92b}$$

$$E_g = 1.424 + 1.247x \text{ eV} \cdot \text{s} \tag{14.92c}$$

$$n = (13.1 - 3x)^{1/2} \tag{14.92d}$$

where m_0 is the electron mass 9.109×10^{-31} kg and n is the index of refraction. The terminal voltage V is related to the difference between the quasi-Ferm levels.

$$eV = F_c - F_v \quad \text{(in electron volts)} \tag{14.93}$$

Solution

(a) With the Boltzmann distribution as an approximate quasi-Fermi distribution, the integrals for the carrier densities become

$$N = \frac{\sqrt{2}}{\hbar^3}\frac{m_e^{*3/2}}{\pi^2}\int_{E_c}^{\infty}(E_2 - E_c)^{1/2}e^{-(E_2-F_c)/kT}\,dE_2 \tag{14.94}$$

$$P = \frac{\sqrt{2}}{\hbar^3}\frac{m_h^{*3/2}}{\pi^2}\int_{-\infty}^{E_v}(E_v - E_1)^{1/2}e^{-(F_v-E_1)/kT}\,dE_1 \tag{14.95}$$

where Eqs. (14.2) to (14.7) were used, and N and P are the negative and positive carrier densities, respectively. These integrals can be performed by putting

$$\frac{E_2 - E_c}{kT} = x_2$$

and

$$\frac{E_v - E_1}{kT} = x_1$$

The results are

$$\begin{aligned} N &= N_c e^{(F_c-E_c)/kT} \\ P &= N_v e^{(E_v-F_v)/kT} \end{aligned} \tag{14.96}$$

where

$$\begin{aligned} N_c &= 2\left(\frac{\sqrt{2\pi k m_0}}{h}\right)^3\left(\frac{m_e^*}{m_0}T\right)^{3/2} \\ N_v &= 2\left(\frac{\sqrt{2\pi k m_0}}{h}\right)^3\left(\frac{m_h^*}{m_0}T\right)^{3/2} \end{aligned} \tag{14.97}$$

Inserting the physical constants gives

$$N_c = 4.817 \times 10^{15} \left(\frac{m_e^*}{m_0} T\right)^{3/2} \text{cm}^{-3}$$
$$N_v = 4.817 \times 10^{15} \left(\frac{m_h^*}{m_0} T\right)^{3/2} \text{cm}^{-3} \tag{14.98}$$

The densities N_c and N_v are called the effective density of states for the negative and positive carriers, respectively. Comparing with Eq. (14.2) in format, Eq. (14.96) can be interpreted as if the energy states of the carriers are discretely concentrated at E_c and E_v with the density of states N_c and N_v rather than spread across the bands. The carrier densities N and P are simply obtained by multiplying N_c or N_v by $e^{(F_c - E_c)/kT}$ or $e^{(E_v - F_v)/kT}$, which are the quasi-Fermi distribution functions indicating the probability that the E_c and E_v levels are occupied [8]. Electrical neutrality requires

$$N + N_0 = P + P_0 \tag{14.99}$$

where N_0 and P_0 are the densities of n and p dopants, respectively. The active layer is practically undoped so as to lower the loss of the cavity (increasing the doping increases the loss inside the cavity) and, hence,

$$N = P \tag{14.100}$$

Thus, from Eq. (14.96), the product NP gives

$$F_c - F_v = E_g + kT \ln\left(\frac{N^2}{N_c N_v}\right) \tag{14.101}$$

This is the desired relationship between the electron density and the Fermi level gap, which gives the terminal voltage V by Eq. (14.93).

In order to calculate m_e^*, m_h^*, and n, the molar fraction x has to be determined. The longest wavelength limit $\lambda = 0.846$ μm of the wavelength spectrum in Fig. 14.13c approximately corresponds to the energy bandgap and

$$E_g = h\nu = 1.467 \text{ eV}$$

where Planck's constant $h = 4.136 \times 10^{-15}$ eV · s in electron volts was used.

From Eq. (14.92c), the molar fraction x is determined to be

$$x = 0.0345$$

and the effective masses become

$$\frac{m_e^*}{m_0} = 0.0698$$
$$\frac{m_h^*}{m_0} = 0.510$$

The effective densities of states at $T = 293$ K are, from Eq. (14.98),

$$N_c = 0.446 \times 10^{18}\text{cm}^{-3}$$
$$N_v = 8.80 \times 10^{18}\text{cm}^{-3}$$

(b) The spontaneous emission lifetime τ_n is found by fitting the curve of Eq. (14.61) with the graphs in Fig. 14.13b. The values used for fitting are tabulated in Table 14.1. The average value is $\tau_n = 2.65$ ns.

Next, the length L of the cavity will be found using Fig. 14.13c. From Eq. (14.34), the cavity length is

$$L = \frac{\lambda^2}{2n\ \Delta\lambda}$$

Inserting $x = 0.0345$ into Eq. (14.92), the index of refraction is found to be $n = 3.6$. The wavelength λ and the separation between longitudinal modes $\Delta\lambda$ are read from Fig. 14.13c. Thus, the cavity length is

$$L = 248\ \mu\text{m}$$

First, the V–I curve below threshold will be found. Using Eq. (14.48), and letting w denote the width of the laser cavity as shown in Fig. 14.7, the relationship between injection current I and N is

$$\begin{aligned} N &= \tau_n \frac{I}{wL}\frac{1}{2ed} \\ &= 2.65 \times 10^{-9}\frac{I}{(7 \times 10^{-6})(248 \times 10^{-6})(1.6 \times 10^{-19})(0.2 \times 10^{-6})} \\ &= 4.77 \times 10^{25} I/\text{m}^3 \end{aligned} \tag{14.102}$$

Hence, the carrier density (in cm^3) is

$$N = 4.77 \times 10^{19} I/\text{cm}^3$$

Inserting this result and N_c and N_v into Eq. (14.101) gives

$$F_c - F_v = (1.628 + 50.57 \times 10^{-3} \ln I)\ \text{eV} \tag{14.103}$$

where $I = 0$ is excluded and where Boltzmann's constant $k = 8.63 \times 10^{-5}$ eV/mol · K in electron volts was used.

Table 14.1 Turn-on delay time and output light power versus J/J_{th}

$\frac{J}{J_{\text{th}}}$	t_d (ns)	τ_n (ns)	P (mW)
1.2	4.8	2.67	1.1
1.3	3.9	2.66	1.47
1.4	3.3	2.63	1.92

> The amount of time for one man to finish a workload of 5 man-months is 5 months. Similarly, if one electron possesses 5 eV of energy, then that electron has a potential of 5 volts.

Since $F_c - F_v$ is in electron volts, the bias voltage V_b that can give $(F_c - F_v)$ eV is, from Eq. (14.93),

$$eV_b = (F_c - F_v) \text{ eV}$$

The potential V of each electron is

$$V_b = (F_c - F_v) \text{ volts}$$

Keep in mind that N does not exceed N_{th}. Thus, from Eq. (14.101), V_b stops increasing as soon as N reaches N_{th}, no matter how I is increased, as the measured curve in Fig. 14.14 indicates.

Next, the threshold carrier density and the corresponding threshold current will be found. From Eq. (14.29) at threshold, the gain satisfies

$$g_{\text{th}} = \frac{1}{L} \ln \frac{1}{R} + \alpha$$

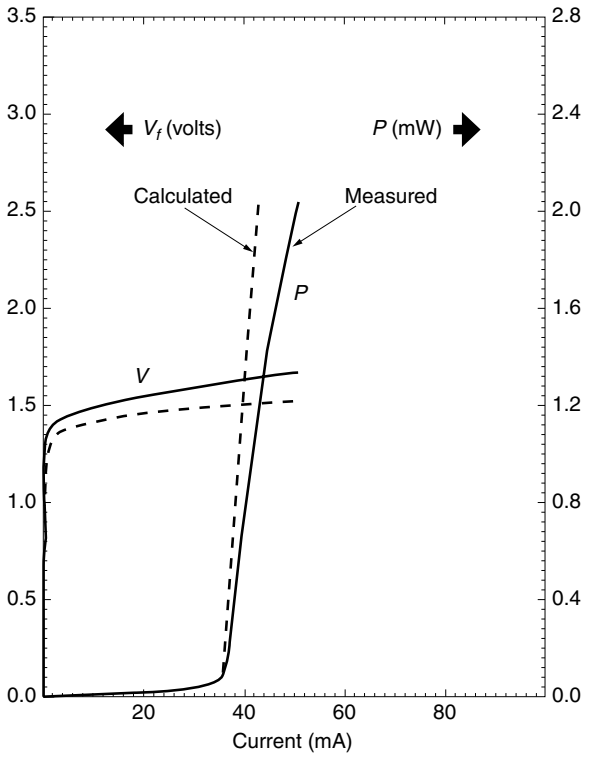


Figure 14.14 Measured (solid line) and calculated (dashed line) V–I curve and P–I curve of an $Al_xGa_{1-x}As$ laser diode.

With the given value $\alpha = 175\text{cm}^{-1}$, the threshold gain becomes

$$g_{\text{th}} = 221 \text{ cm}^{-1}$$

The gain spectrum in Fig. 14.13d [10] is replotted as a function of the carrier density in Fig. 14.15. The curve is quite linear. The threshold electron density for $g = 221 \text{ cm}^{-1}$ is

$$N_{\text{th}} = 1.7 \times 10^{18} \text{ cm}^{-3}$$

The current needed to supply N_{th} is

$$I_{\text{th}} = \frac{wL2ed}{\tau_n} N_{\text{th}}$$

From Eq. (14.102),

$$I_{\text{th}} = 35.6 \text{ mA}$$

which is close to the measured value. The corresponding terminal voltage is 1.459 volts from Eq. (14.103). The external terminal potential V_e is

$$V_e = \begin{cases} IR + (F_c - F_v) \text{ volts} & \text{below the threshold, } I < 35.6 \text{ mA} \\ IR + 1.459 \text{ volts} & \text{above the threshold, } I > 35.6 \text{ mA} \end{cases}$$

Using Eq. (14.103), the calculated and measured V_e are compared in Table 14.2. □

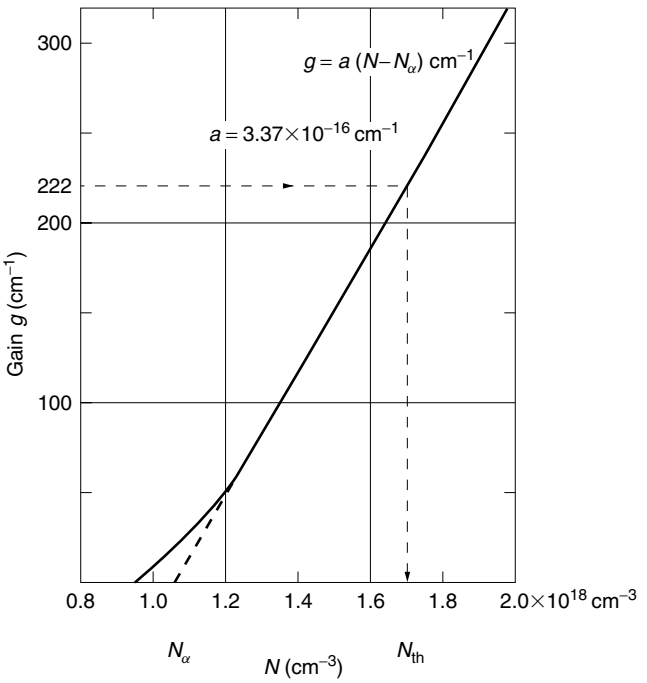


Figure 14.15 Linearized gain with respect to the carrier density N for GaAs.

Table 14.2 Terminal voltage compared to injection current

I (mA)	IR (volts)	$F_c - F_v$ (volts)	Calculated V_e (volts)	Measured V_e (volts)
10^{-1}	0.5×10^{-4}	1.159	1.159	1.15
1	0.5×10^{-3}	1.275	1.275	1.35
2	10^{-3}	1.310	1.311	1.37
5	2.5×10^{-3}	1.357	1.360	1.42
35	17.5×10^{-3}	1.459	1.477	1.58
50	25×10^{-3}	1.459	1.484	1.62

Example 14.2

(a) With the same laser diode as studied in Example 14.1, obtain the theoretical light output versus injection current curve.

(b) Calculate ω_r and $1/\gamma$ at $I = 1.4I_{\text{th}}$. Compare the theoretical rise time t_r with the value in the specification sheet. The rise time t_r is related to the cutoff frequency f_m^r of the modulation by Eq. (16.85) as

$$t_r f_m^r = 0.35$$

Solution

(a) The output power is given by Eq. (14.58). From Eqs. (14.39) and (14.55) the output power is

$$P_{\text{out}} = E_g \frac{\frac{1}{L}\ln\frac{1}{R}}{\frac{1}{L}\ln\frac{1}{R} + \alpha}(I - I_{\text{th}}) \text{ W} \tag{14.104}$$

Insertion of the parameters gives

$$P_{\text{out}} = 0.31(I - I_{\text{th}}) \text{ W}$$

The calculated results are plotted with the experimental curve in Fig. 14.14. The theoretical curve puts out slightly more power than the experimental curve.

(b) The parameter γ is given by Eq. (14.81) and τ_n is obtained from Table 14.1.

$$\frac{1}{\gamma} = 1.9 \times 10^{-9}\text{s}$$

The frequency ω_r is given by Eq. (14.89) and τ_s is given by Eq. (14.83):

$$\omega_r = 1.657 \times 10^{10} \text{ rad/s}$$

$$f_r = 2.63 \text{ GHz}$$

From Eq. (14.88),

$$f_m^r = 2.629 \text{ GHz}$$

The rise time is

$$t_r = \frac{0.35}{2.629 \times 10^9} = 0.133 \times 10^{-9}\text{s} = 133 \text{ ps}$$

The datasheet in Fig. 14.13a gives a rise time of 100 ps. □

14.4 CONFINEMENT

The overall dimensions of the laser diode shown in Fig. 14.7 are typically 200 × 200 μm^2 in base and 75 μm in height with an active region typically 0.1–0.2 μm thick, 200 μm long, and 10 μm wide. Compared to most types of lasers, the dimensions of the laser diode are indeed very small. In this small region, one has to confine carriers, injection current, and photons.

For best performance, it is essential to confine all three quantities effectively. Proper design of the confinement is one of the most decisive factors in laser diode performance. Effective confinement lowers the threshold current. As discussed earlier, lowering the threshold current J_{th} shortens both the turn-on delay time and the period of the relaxation oscillation and raises the upper limit of the modulation frequency. A lower threshold current also means that less thermal heat is generated, thereby reducing heat sink problems and the temperature dependence of the performance. The sections to follow look at ways to achieve the desired confinement of carriers, injection current, and photons.

14.4.1 Carrier Confinement

In order to build a high-efficiency laser diode, it is important to maintain maximum carrier density with minimum injection current. The double heterojunction is formed by sandwiching the active layer by layers with a wider energy bandgap. Potential barriers formed by the wider energy bandgap material bounce back the carriers once injected into this region and confine the injected carriers inside the active layer. Thus, a high carrier concentration is established in the active layer. In Fig. 14.16, three kinds of junctions are compared, and these are (a) the homojunction, (b) the single heterojunction, and (c) the double heterojunction. The top row diagrams are the energy bands before the p-n junction is formed, those in the middle are the energy bands after the junction is formed, and those on the bottom are the energy bands when a forward bias is applied.

The operation of the homojunction, which is the basis of all others, will be explained first. As soon as the n- and p-type layers are joined, electrons that are majority carriers in the n-type layer start diffusing into the p-type layer as minority carriers. Conversely, holes in the p-type layer start diffusing into the n-type layer.

The diffusion causes a redistribution of the carriers, and as a consequence, a potential difference is established between p- and n-type layers. To understand the origin of this potential difference, let us once again consider the n-type material before it is joined. For every electron that breaks loose from the n-type atom to become a negative carrier, there is a positively charged immobile ion left behind. Thus, overall, the n layer is electrically neutral. In the same way, the p layer is also electrically neutral before it is joined. For every positive carrier (hole) that is created, a negatively charged immobile

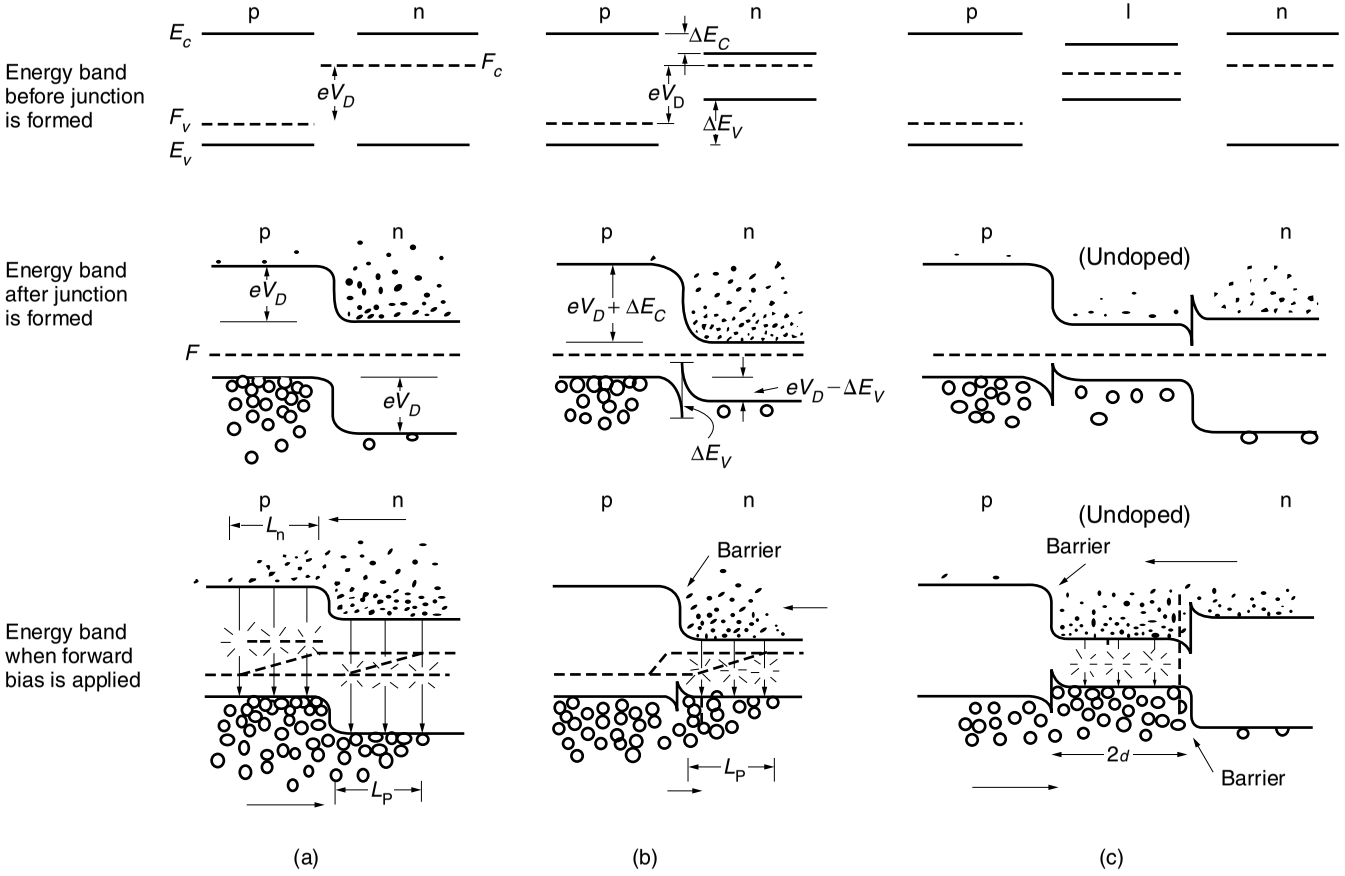


Figure 14.16 Comparison of three kinds of p-n junctions. (a) Homojunction. (b) Single heterojunction. (c) Double heterojunction.

ion is left behind in the crystal lattice. As soon as the contact is made, electrons in the n-type layer diffuse into the p-type layer, leaving immobile positively charged ions behind. As a result, the n-type layer starts to have a positive potential. For the same reason, the diffusion of the holes from the p-type layer into the n-type layer results in charging the p-type layer to a negative potential. The potential difference V_D between the two layers is called the *diffusion potential* or *contact potential*.

The diffusion of the electrons does not go on forever. As the diffusion progresses, the diffusion potential also keeps increasing. The polarity of the diffusion potential is such that the associated Coulomb force on the carriers is in the opposite direction to the direction of the diffusion force. At the equilibrium of the two forces, the movement of the electrons and holes ceases. At equilibrium, the Fermi levels in the two layers line up as shown in the middle row of Fig. 14.16, but steps are created between the energy levels. When a forward bias is applied, the steps of the energy levels are decreased. Electrons in the n-type layer start flowing into the p layer, and the holes in the opposite direction. As the flow of electrons penetrates into the p region, the electrons have a chance to recombine with a sea of holes and to emit light by spontaneous emission. The density of electrons decays as $e^{-x/L_{\mathrm{n}}}$ (exponentially with distance of penetration) where L_{n} is the diffusion length of electrons. The region of emission is spread over approximately L_{n} from the junction. Similarly, the diffusion length of the holes in the n-type layer is L_{p}. In the case of the homojunction, both p- and n-type layers near the junction emit light, and the region of emission is spread over approximately $L_n + L_p$.

Next, the function of the heterojunction will be explained. The heterojunction is made of p- and n-type layers with different energy bandgaps as shown in Fig. 14.16b. The middle diagram shows the configuration of the energy bands after the junction is made. The Fermi levels of p and n layers line up with each other. The band structure of the p layer moves up. The conduction band of the p layer becomes higher than that of the n layer.

The final shape of the conduction band can be drawn by smoothly connecting the conduction bands of the two layers. The shape of the valence band near the junction can be obtained by drawing lines parallel to those of the conduction band. As a consequence, the step between the conduction bands of the p- and n-type layers is $eV_D + \Delta E_c$ where ΔE_c is due to the original difference of the conduction bands between the p and n layers. This step is larger than eV_D of the homojunction. The step between the valence bands of the p- and n-type layers is $eV_D - \Delta E_v$, which is smaller than eV_D of the homojunction. When a forward bias is applied to the junction, there is an imbalance between the flow of electrons and holes because of the imbalance between the height of steps in the conduction and valence bands as indicated in the bottom diagram. The electrons hardly penetrate into the p-type layer because of the barrier, but holes are able to penetrate the n-type layer. Hence, the recombination takes place only in the n-type layer in a region within L_{p} from the junction, as shown in Fig. 14.16b. The region of recombination is reduced to roughly half, compared to the homojunction, and the carrier concentration can be doubled for the same injection current.

Finally, the double heterojunction will be explained. The double heterojunction consists of two single heterojunctions on both sides of the active region, as shown in Fig. 14.16c. As shown in the middle figure, the junction between the p-type layer and the center layer makes a barrier against electrons while the junction between the n-type layer and the center layer makes a barrier against holes. When a forward bias is

applied, holes migrate from the p-type layer to the center layer, and electrons migrate from the n-type layer to the center layer. However, penetration of carriers from the center layer to the outer layers is effectively blocked by the barriers. It is only in the center layer that both electrons and holes are present for recombination. The region of emission is concentrated in this center region of thickness $2d$. The carrier concentration $N = \tau_n J/2ed$ in Eq. (14.44) can be raised by choosing a small thickness $2d$. The value of $2d$ is normally 0.1–0.2 μm. Compared to the homojunction $L_n + L_p = 6$ μm and single heterojunction $L_p = 3$ μm, the carrier density of the double heterojunction can be 30–60 times greater for the same current.

It is worthwhile at this point to look at some specific examples of commonly used materials for p-n junctions. Examples of host semiconductor materials include crystals of valence 4 atoms such as germanium (Ge) or silicon (Si), and compounds made of equal numbers of valence 3 and 5 atoms such as gallium arsenide (GaAs) and indium phosphide (InP). A variation of the latter is to use a mixture of valence 3 atoms in combination with the appropriate number of valence 5 atoms, as in $Al_xGa_{1-x}As$, where x is the molar fraction of Al atoms. With a Ge or Si crystal, an n-type semiconductor is formed by adding an impurity to the host of valence 5 or greater like P, As, or Sb, and a p-type semiconductor is formed by adding an impurity of valence 3 or less like B or Ga. A configuration that is often used for the double heterostructure is to use GaAs as the center layer, with p-doped $Al_xGa_{1-x}As$ on one side, and n-doped $Al_xGa_{1-x}As$ on the other side. In the case of the double heterojunction formed by $Al_xGa_{1-x}As$, the energy bandgap is changed by the molar fraction x. Aside from the increase in the energy bandgap, an increase in x results in a decrease in the refractive index of the layer.

It may be added that if the thickness $2d$ of the active layer in the double heterojunction is further reduced to the order of 10 nm, the motion of the electrons becomes restricted and the quantized nature of the energy levels must be taken into account. In this case, the double heterojunction becomes a quantum well laser that displays different features in threshold current, bit rate, and temperature dependency and will be described in more detail in Section 14.12.

14.4.2 Confinement of the Injection Current

Any injection current that flows in a region other than the designed active region increases the apparent threshold current. In order to keep the threshold current low, the injection current has to be confined to the desired region.

Various ingenious methods of funneling the current into the active region of the laser diode have been reported [11]. Five examples will be presented:

1. Narrow stripe electrode.
2. Raised resistivity by proton bombardment.
3. Barricade by a back-biased p-n junction layer.
4. Dopant-diffused channel.
5. Modulation of layer thickness.

A brief description of each method will be presented with illustrations. In all the illustrations, the coding for the shaded areas will be unified. The n-type region will be

indicated by slashing down to the left, and the p-type region by slashing down to the right.

14.4.2.1 Narrow Stripe Electrode

Figure 14.17a shows the structure of the electrode stripe laser. This structure is the simplest of all types. The narrowness of the electrode is a means of confinement. Lasers whose active region lateral dimension (width) ranges from 10 to 20 μm can use this configuration. An insulating layer of SiO_2 is deposited except for the center stripe region. A metal layer is then deposited that serves as the 10–20 μm wide electrode. Even though the deposition of SiO_2 as the insulator is simple, a drawback is the low heat conductivity.

14.4.2.2 Raised Resistivity by Proton Bombardment

Electrons are funneled into a narrow region by a resistivity wall. The resistivity of a semiconductor can be raised locally by bombarding the semiconductor with high-energy

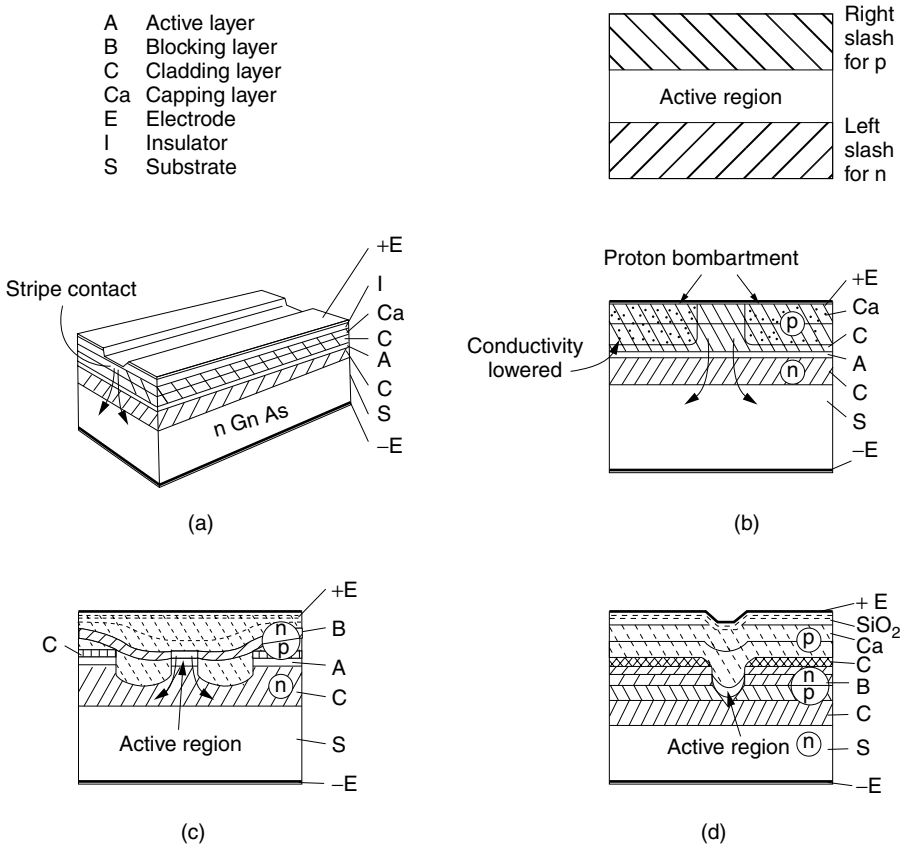


Figure 14.17 Geometries of laser diodes. (a) Electrode stripe laser. (b) Proton bombarded stripe laser. (c) Double-channel planar buried heterostructure (DC-PBH) laser. (d) Buried crescent (BC) laser. (e) Channeled substrate planar (CSP) laser. (f) V-grooved substrate buried heterostructure (VSB) laser. (g) Terraced substrate (TS) laser. (h) Stripe substrate laser. (i) Transverse junction stripe (TJS) laser. (j) Constricted double heterojunction (CDH) laser.

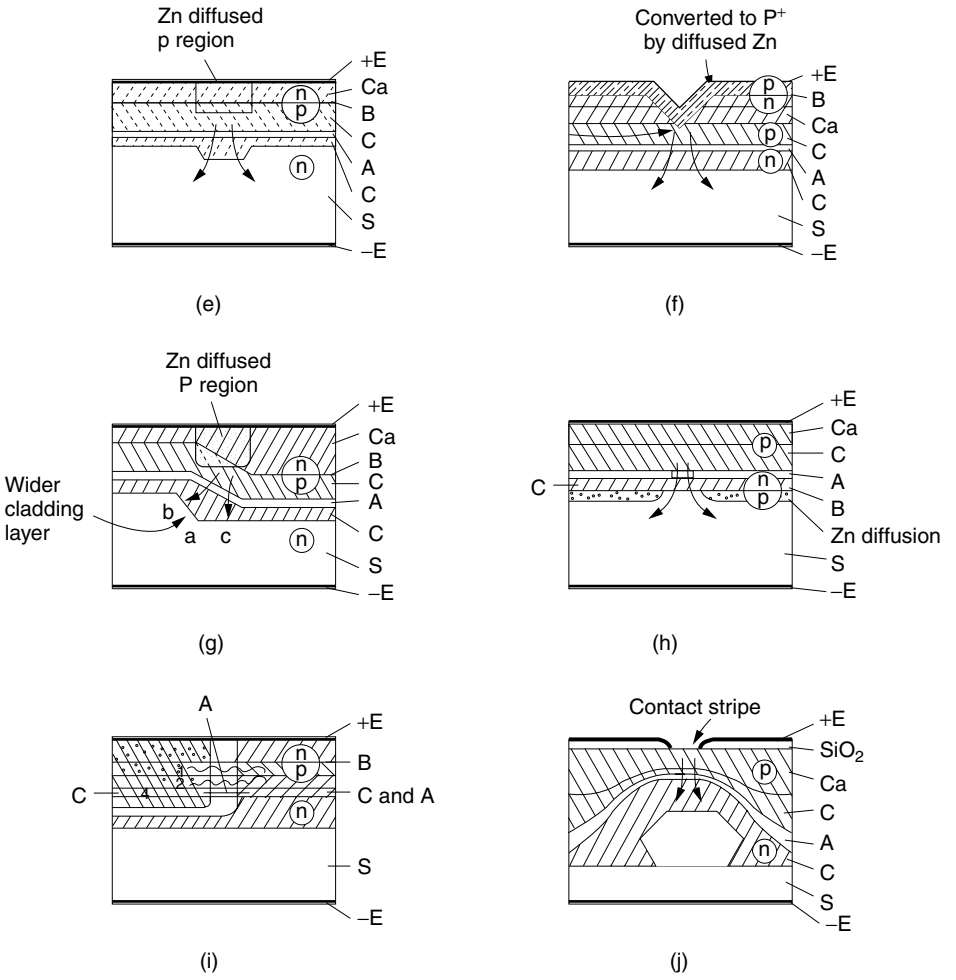


Figure 14.17 (*Continued*)

protons. Figure 14.17b shows the cross section of a proton-bombarded stripe laser. The proton-bombarded region has increased resistivity due to the damage caused by the bombardment. The injected current is funneled into the region that has escaped the bombardment. In some designs, the bombardment is allowed to penetrate as far as the active layer for even tighter confinement. Another advantage of this method is better heat conduction than the SiO_2 insulator used in the previous method.

Test results [12] show that the 10-Ω resistance of the n-type GaAs becomes 10^8 Ω after bombardment with a 3-MeV proton beam of dose $10^{13}/cm^2$. The resistivity increases 10^7 times.

14.4.2.3 Barricade by a Back-Biased p-n Junction Layer

When a p-n junction is back-biased, the current is blocked by the high impedance across the junction. The aperture made on the barricade provides confined electrons. The simplest example is the case when the SiO_2 layer in Fig. 14.17a is replaced by an

n-type semiconductor layer so that the top two semiconductor layers form a back-biased p-n junction. The technique of the back-biased p-n junction is widely used, and a few types of lasers that use this technique are included here.

The structure of the double-channel planar buried heterostructure (DC-PBH) laser is fabricated as shown in Figs. 14.17c. The current cannot pass through the blocking p-n junction layer indicated by B and flows through the top surface of the mesa.

Another laser that uses the blocking p-n junction is the buried crescent (BC) laser [13] shown in Fig. 14.17d. Only the active layer inside the V groove will lase, as the current cannot get to the active layer on the sides because of the blocking p-n junction underneath indicated by B. A stripe electrode is used to reinforce the confinement of the injection current. This laser gets its name from the fact that the active layer is shaped like a crescent rather than a rectangle.

The method of the back-biased p-n junction is used more often in combination with the method of the diffusion of dopants than by itself. Windows for the current are created by locally converting the p-n junction into either a p-p junction or an n-n junction by diffusing dopants, as will be explained in the next section.

14.4.2.4 Dopant-Diffused Channel

The conversion is done locally by diffusing the dopants. Diffusion of either zinc (Zn) or cadmium (Cd) (group II) is commonly used to convert the semiconductor to p-type, and sulfur (S) (group VI) to convert to n-type. The channeled substrate planar (CSP) laser shown in Fig. 14.17e uses this method. In the center region, Zn is diffused through the p-n blocking layer. The n layer is converted into a p-type semiconductor and the underlying p layer becomes more strongly p type. The current channel is thus created in the center. This current channel is sometimes extended close to the active region for better confinement of the current.

Another type of laser that uses the same diffusion technique is the V-grooved substrate buried heterostructure (VSB) laser shown in Fig. 14.17f. The V groove is made so that only the tip of the Zn-diffused p-type region reaches the p-type layer of the laser. Thus, the size of the current window is small for better confinement.

The terraced substrate (TS) laser shown in Fig. 14.17g is another example. A substrate with a step in the center is used. On the substrate, the layers are grown in the same order as other types. The Zn-diffused p region directly connects the positive electrode to the p layer to form a current channel. The active region is tilted so that the cladding layer is wider near the center of the active region. As will be explained in the next section, the tapered width of the cladding layer improves the confinement of the photons inside the active layer.

The stripe substrate laser shown in Fig. 14.17h uses the dopant layer differently. The Zn-diffused p layer is not used to open a channel; rather, it is used as a component layer of the blocking p-n junction. The surface of the substrate is first treated by Zn diffusion to create a p-type layer except for a narrow region in the center. An n-type layer, followed by an active layer, followed by a p-type layer, is grown to form a laser with a current channel in the center.

The transverse junction stripe (TJS) laser [14] shown in Fig. 14.17i uses not only Zn diffusion but also the differences in the forward-biased p-n junction impedance. The top two layers form a back-biased p-n junction and no current flows vertically. The extensively Zn-diffused upper left section becomes a p^+ region. Between the left and right regions in the third and fifth layers, p-n junctions are created.

All possible paths below the blocking p-n layer are indicated by the wavy arrows in Fig. 14.17i. Even though all paths are forward-biased p-n junctions, the active region indicated by a straight arrow has the least impedance because of the narrower energy bandgap of the C layer material, and the injection current is confined to flow horizontally through the active layer.

14.4.2.5 Modulation of the Layer Thickness

The constricted double heterojunction (CDH) laser shown in Fig. 14.17j takes a slightly different approach to confine injection current. A difference in resistivity is established by the differences in layer thickness. The cladding–active–cladding layers are grown over a substrate that has a protrusion on the top surface. Both the cladding and the active layers are made thinnest in the center section so that the resistivity is the lowest, and the current selectively flows in the center.

14.4.3 Light Confinement

Since the laser light is amplified by going back and forth in the active region, it is important for the light to be properly confined in the active region in both the transverse (vertical) and lateral (horizontal) directions. Let us first deal with transverse confinement.

In designing the transverse dimension (thickness) of the active layer, the following considerations are made:

1. Because the higher order modes create side lobes in the radiation pattern of the laser that make the coupling of the laser output into an optical fiber more difficult, the fundamental mode is used exclusively. The upper limit on the thickness $t = 2d$ is determined from the cutoff of the first higher order mode. From Eq. (9.32), the thickness must satisfy

$$t \leq \frac{\lambda}{2\sqrt{n_1^2 - n_2^2}} \tag{14.105}$$

2. Deciding the thickness of the active layer is a compromise between a higher efficiency of the active layer and a narrow beamwidth of the output light.

From these two considerations the thickness of the active layer is selected to be around 0.1–0.2 μm for most lasers in the range of 0.85–1.55 μm.

Light confinement in the vertical (transverse) direction is achieved by means of refractive index guiding, which is based on total internal reflection from the boundaries between the core and cladding layers. Refractive index guiding can be used to confine the light in the lateral (horizontal) direction as well. An alternate approach for lateral confinement is to use gain guiding, a technique that is unique to lasers. The following few sections will concentrate on lateral light confinement.

14.4.3.1 Gain Guiding

The stripe electrode laser will be used for the sake of explanation. Figure 14.18 shows the carrier concentration distribution $N(y)$, the gain distribution $g(y)$, and the refractive index distribution $n(y)$ also with respect to the y (lateral) axis for the case that J is

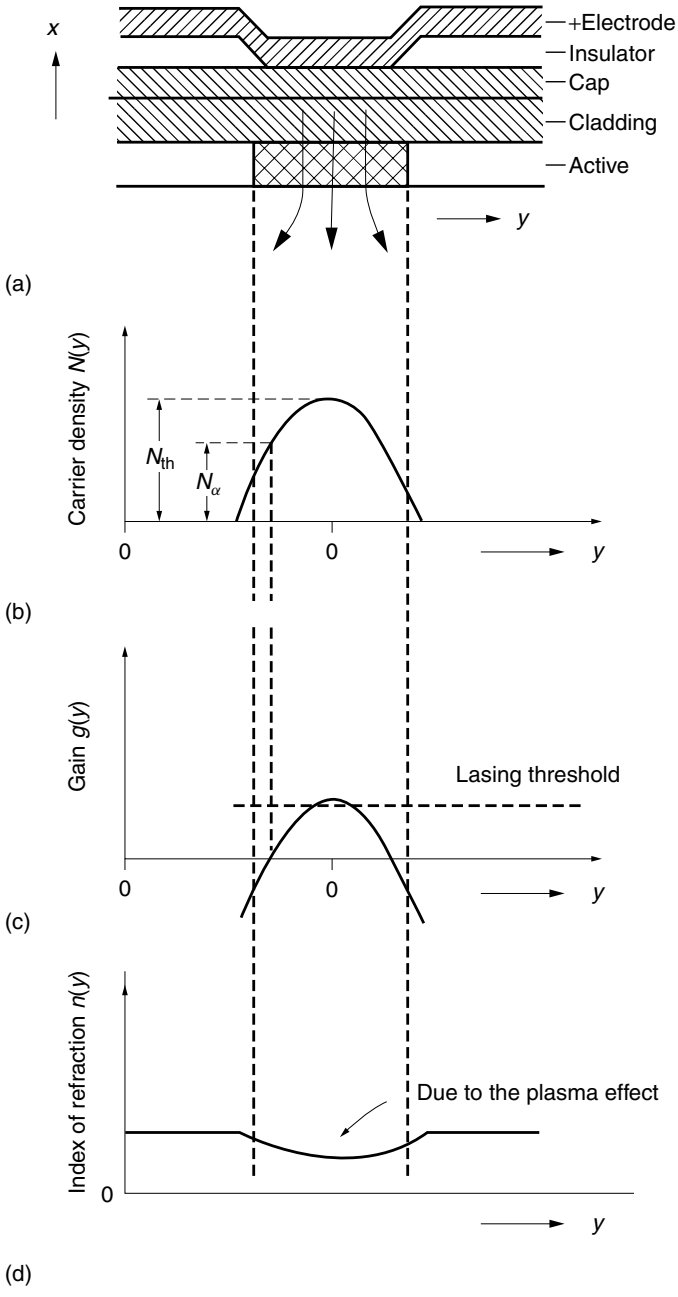


Figure 14.18 Stripe electrode laser and the distribution function of parameters: carrier concentration in the center creates an index dip due to the plasma effect. (a) Active region. (b) Carrier density distribution. (c) Gain distribution. (d) Distribution of the index of refraction.

slightly above J_{th} at $y = 0$. The injection current density J is at its maximum in the center and almost zero at both sides of the active region. The distribution of the carrier concentration roughly follows the current distribution. The gain $g(y)$, in turn, follows the distribution of the carrier concentration in accordance with

$$g(y) = a[N(y) - N_\alpha] \tag{14.106}$$

In short, the $g(y)$ curve more or less follows that of J, except that the curve is lowered by aN_α. As the junction current is increased, $g(y)$ increases and the maximum of $g(y)$ located in the center first reaches the threshold and initiates lasing at the center. The photons now become available for stimulation. The intensity of generated photons is also highest in the center region. The combination of high gain and high stimulated emission further accelerates the stimulating emission near the center. Thus, once the lasing starts near the center where the gain is maximum, the location of the lasing is pinned down at that location even without other light-guiding mechanisms. Such a phenomenon that lasing stays in the region of the highest gain is called *gain guiding*. The gain-guiding principle can be utilized as the sole guiding mechanism, but one has to be aware of potentially undesirable side effects, which include the carrier-induced plasma effect, hole burning, and self-focusing. These are discussed in the sections to follow.

14.4.3.2 Plasma-Induced Carrier Effect

Being charged particles, the carriers are subjected to Coulomb's force by the electric field **E** of the lightwave. Oscillation of the charges creates an oscillating polarization. The phase of the oscillation, however, is shifted from the applied field due to the mass inertia of the electrons. The oscillating electrons with the shifted phase decrease the index of refraction (see boxed note). This reduction of the refractive index is called the *plasma effect* of the electrons. The quantitative expression for the decrease in the refractive index due to the plasma effect is

$$\Delta n = \frac{e^2 N}{2\omega^2 \epsilon_0 \epsilon_r m_e^*} n \tag{14.107}$$

where N is the electron concentration, n is the refractive index in the absence of the plasma effect, $\epsilon_0\epsilon_r$ is the dielectric constant of the active region, and m_e^* is the effective mass of the electron. The coefficient of Eq. (14.107) becomes $\sim 10^{-2}$ with $N = 10^{25}$ m^{-3}, and $\Delta n/n$ is about 1%. Considering the fact that the refractive index step that confines the light in the core of the optical fiber is only 0.5%, the plasma effect is of a comparable magnitude.

The refractive index distribution has a shallow minimum at the center due to the plasma effect, as shown at the bottom of Fig. 14.18. This index dip tends to diverge the light beam, and the divergent beam moves out of the region and does not lase. The light beam that remains in the center does lase, as long as threshold conditions continue to be satisfied. The net result is that the gain guiding keeps the light beam in the center. However, the situation becomes more complex when J is increased well above J_{th}, as will be shown next.

An extremely simplified proof of Eq. (14.107) will be presented here. Let the **E** field of the light be represented by

$$\mathbf{E} = \mathbf{E_0} e^{-j\omega t} \tag{14.108}$$

The equation of motion due to the Coulomb force exerted by **E** on a particular electron with charge e and effective mass m_e^* is

$$m_e^* \frac{dv}{dt} = eE_{ox} e^{-j\omega t} \tag{14.109}$$

where the direction of the x axis is taken in the same direction as **E**. It is also assumed that there are no collisions between the electrons that impede the motion of the electron. Insertion of Eq. (14.108) into (14.109) followed by integration gives

$$v = j \frac{eE_{ox}}{\omega m_e^*} e^{-j\omega t} \tag{14.110}$$

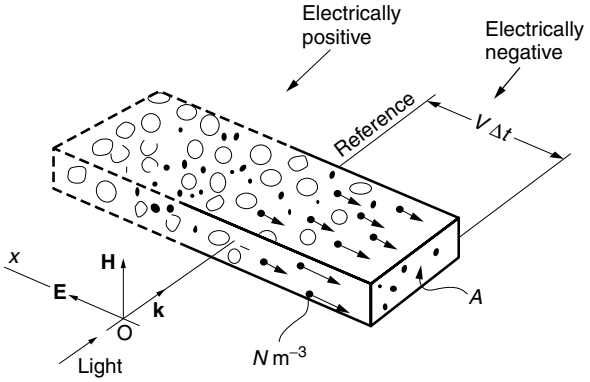


Figure 14.19 The movement of charge across a reference cross section is the plasma current.

In a short time Δt this particular electron moves a distance $v\Delta t$. In the absence of collisions, all electrons move by the same amount $v\Delta t$, leaving the heavy ions behind. With reference to a particular cross section, the total volume of electrons that are pushed out is $Av\Delta t$, as shown in Fig. 14.19. The total charge ΔQ that has moved through the reference cross section is

$$\Delta Q = eNAv\Delta t \tag{14.111}$$

The density J_{px} of the plasma current, which is the movement of the charges across a unit area of the reference cross section, is

$$J_{px} = \frac{1}{A} \frac{\Delta Q}{\Delta t} \tag{14.112}$$

Inserting Eqs. (14.110) and (14.111) into Eq. (14.112) gives

$$J_{px} = j \frac{e^2}{\omega m_e^*} NE_{ox} e^{-j\omega t} \tag{14.113}$$

More generally,

$$\mathbf{J_p} = j \frac{e^2}{\omega m_e^*} N\mathbf{E}$$

The oscillation of the free electrons contributes to the total current, and Maxwell's equation becomes

$$\nabla \times \mathbf{H} = j\frac{e^2}{\omega m_e^*}N\mathbf{E} + \frac{d\mathbf{D}}{dt} \tag{14.114}$$

Using Eq. (14.108) in (14.114) leads to

$$\nabla \times \mathbf{H} = -j\omega\epsilon_0\epsilon_r\left(1 - \frac{e^2N}{\omega^2\epsilon_0\epsilon_r m_e^*}\right)\mathbf{E} \tag{14.115}$$

The value in the large parentheses is unity when $N = 0$. One way of looking at Eq. (14.115) is that it is Maxwell's equation with a modified dielectric constant. In terms of the refractive index $n = \sqrt{\epsilon_r}$, the modified index of refraction n' is

$$n' = n\sqrt{1 - \frac{e^2N}{\omega^2\epsilon_0\epsilon_r m_e^*}}$$

Finally, $\Delta n = n' - n$ is approximated as

$$\Delta n = \frac{e^2N}{2\omega^2\epsilon_0\epsilon_r m_e^*}n \tag{14.116}$$

14.4.3.3 Kink in the Characteristic Curve

Another phenomenon closely related to the lateral mode confinement is a kink appearing in the curve of the light output versus injection current. As the injection current is increased, the distributions of the refractive index, carrier concentration, and light intensity all pass through various states. The transitions of these changes are shown in Fig. 14.20. The curves in Fig. 14.20a are with the least injection current and Figs. 14.20b and 14.20c are with larger injection currents. The gain guiding explained in Fig. 14.18 was for an injection current slightly above threshold and corresponds to the case in between Figs. 14.20a and 14.20b.

As the injection current is increased above the threshold, the number of electron–hole pairs that are annihilated by stimulated emission increases nonlinearly. The supply of electrons from the injection current cannot keep up with the rate of annihilation, and the distribution curve of the carrier concentration starts to have a dip in the center, as indicated in Fig. 14.20b. Another way of explaining this is to draw upon the results of Section 14.3.2, where it was shown that, in the region of lasing, the carrier concentration stays clamped at N_{th} no matter how high J is increased above J_{th}. The dip in the carrier concentration curve is called *hole burning*. Where the hole burning takes place, the index of refraction is raised because of the reduction in the plasma effect and a local maximum is formed, as indicated by the bottom graph in Fig. 14.20b. This local maximum of the refractive index forms a light-guiding structure that steers the light toward the center and creates a further shortage of carriers in the center, which in turn raises the index of refraction and steers even more light toward the center. These repeated actions confine the light tightly in the center. This phenomenon is called *self-focusing*. As self-focusing sets in, there is a certain range in which an increase in injection current does not effectively increase the light output because the lasing is confined only to the place where the carrier concentration is low.

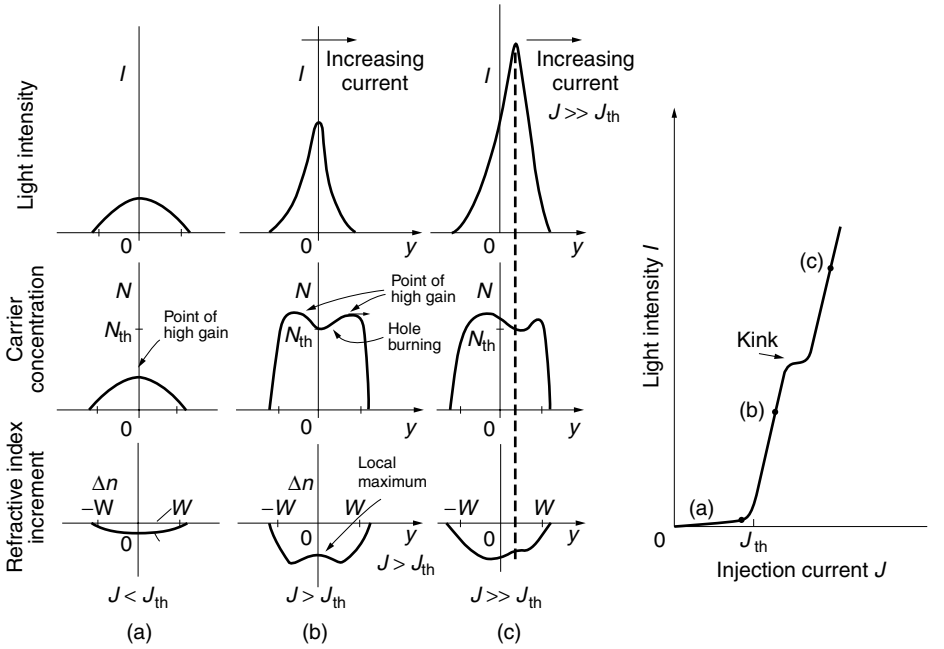


Figure 14.20 Lateral light confinement in a diode laser. (a) Below threshold. (b) Self-focusing. (c) Competition between self-focusing and gain guiding.

This ineffective increase in the light output is displayed as a kink in the curve of light output versus injection current, as shown in the right-hand side of Fig. 14.20.

Because of the hole burning in the center, the distribution of the carrier concentration builds up the maxima on both edges of the active region, as shown in the middle figure in Fig. 14.20b. These maxima of the carrier concentration have the potential to become gain guiding, because a higher carrier concentration is more susceptible to gain guiding. Now the laser beam has a choice to stay in the center by the self-focusing or to move toward the edge by the gain guiding, as shown in Fig. 14.20c. Thus, this distribution of the carrier concentration creates an instability. Even a slight fluctuation triggers the shift of the light beam to the two peaks of the carrier concentration curve, and the light beam will be divided into two beams inside the active region. If the active region is wide enough to support the higher order modes, the next higher order mode that has two intensity peaks will start to be excited. The division of the light intensity creates undesirable side lobes in the radiation pattern of the laser.

14.4.3.4 Stabilization of the Lateral Modes

The kink in the characteristic curve can be removed by preventing the hole burning of the carriers. The first method to prevent the hole burning is to decrease the width of the active region to a value less than the diffusion length of the carriers. The width of the active region should be less than $1-2\ \mu m$.

Another method is to use dielectric waveguiding instead of gain guiding, namely, to use the refractive index step between the active region and its adjacent media in the horizontal direction as well as in the vertical direction. The buried heterostructure type lasers such as the DC-PBH laser in Fig. 14.17c and the BC laser in Fig. 14.17d

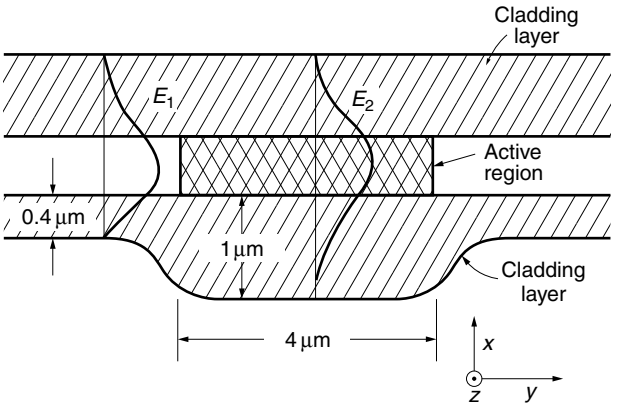


Figure 14.21 Waveguide for laterally confining the light in the active layer of a CSP laser. (After H. Yonezu [8].)

are good examples. The active regions of all these lasers are fenced in by a medium of lower refractive index on all four sides.

An example of a waveguiding method of a special kind is the CSP laser shown in Fig. 14.17e. An enlarged view is shown in Fig. 14.21. The lightwave is confined by a waveguide structure. The width of the bottom cladding layer is constricted to about 0.4 μm all along the bottom except for the active region in the center. The thickness of the cladding layer in the center is about 1 μm, and the width of the thicker center region is about 4 μm. In the region of the constricted cladding layer, the tail of the evanescent wave is pushed up toward the active layer, and the effective index of refraction in the region of the thinner cladding layer is smaller than that of the thicker central cladding region, just like the strip-loaded guide mentioned in Section 10.5.3, thus forming a dielectric waveguide effect in the horizontal direction.

The next few sections will be devoted to the description of the properties of semiconductor lasers.

14.5 WAVELENGTH SHIFT OF THE RADIATION

There are many reasons why wavelength purity and stability are important light source attributes. Material and waveguide dispersion deteriorates the transmission capacity of fiber-optic communication, and a narrow bandwidth of the wavelength of the laser source can minimize these dispersion effects. The requirement for wavelength stability of the source becomes even more severe when the communication system employs superheterodyne detection. A slight drift in the wavelength of the local oscillator laser diode causes the IF frequency to walk out of the IF amplifier bandwidth (see Section 12.6.1).

Wavelength division multiplexing is another example of a system requiring a stable frequency source (see Section 16.3.1). Wavelength division multiplexing is used to expand the capacity of fiber-optic transmission, and an unstable source wavelength creates problems of crosstalk between channels. Yet another problem can arise when the source is modulated. When a laser is amplitude modulated by directly modulating

the injection current, an undesirable frequency modulation results. This is because the wavelength of the laser is dependent on the injection current.

The wavelength of the laser oscillation is determined by $\lambda = 2nL/p$ from Eq. (14.33); therefore, any wavelength shifts are either due to n, p, or L. It so happens that for semiconductor lasers the amount of heat expansion in L is a negligible effect, and n and p are the only two parameters that have to be considered. They are influenced by injection current and temperature. While the injection current lowers the refractive index n due to the plasma effect, the temperature raises n. The mode number p shifts due to the temperature dependence of the gain curve.

The wavelength shift due to n is continuous, while that due to p is discontinuous with respect to the injection current. The two cases will be treated separately.

14.5.1 Continuous Wavelength Shift with Respect to Injection Current

Figure 14.22 illustrates the change in wavelength versus injection current in the horizontal plane and the corresponding output light spectrum in the vertical axis. While the injection current is low and below the threshold current, the carrier concentration increases linearly with the injection current. As the carrier concentration increases, the plasma effect becomes important. The plasma effect lowers the refractive index of the active region by Δn_p. Thus, the resonance wavelength of the cavity *decreases* with injection current in the region below threshold according to Eq. (14.33). This effect continues until the injection current reaches the threshold.

As the injection current is increased beyond the threshold, the carrier concentration no longer increases but remains clamped at N_{th} (See Section 14.3.2), and hence the

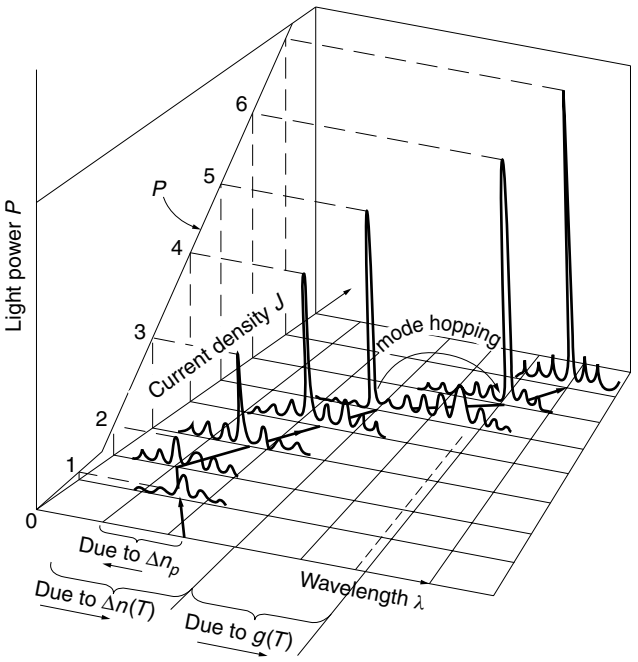


Figure 14.22 Shift of the power spectrum of a laser diode with respect to the injection current.

plasma effect stops increasing. With a further increase in the injection current, the temperature effect that increases the refractive index takes over, and the resonant wavelength of the cavity *increases*. This reverses the direction of the spectrum shift with respect to the injection current in Fig. 14.22.

Now, with a further increase in the injection current, the maximum of the gain curve starts to shift toward the longer wavelengths because the bandgap energy $E_g(T)$ is reduced with an increase in temperature T [15].

$$E_g(T) = E_0 - \frac{\alpha T^2}{\beta + T} \tag{14.117}$$

where E_0 is the bandgap energy at 0 K, T is the temperature in kelvin units, α is 4.5×10^{-4} eV/K, and β is 204 K for GaAs and 327 K for $Ga_xIn_{1-x}As_yP_{1-y}$.

The wavelength of the peak gain *increases* approximately 0.23 nm per degree for a multimode laser and 0.1 nm per degree for a single-mode laser.

14.5.2 Mode Hopping

As far as the longitudinal modes are concerned, there are three basic types of laser diodes. The first type is a true single-mode laser, where only one longitudinal mode oscillates. The second is a "quasi" single-mode laser, where several longitudinal modes oscillate, but only one mode dominates, as shown in Fig. 14.9c. The third type is truly multimode, where several modes oscillate, and no one mode has a large dominance over its neighboring modes. This discussion will concern the second type. Once a laser diode starts lasing at a particular wavelength, the stimulated emission tends to stay at that wavelength until the center of the gain curve shifts by a large amount. When the wavelength does shift due to the temperature rise, it hops one or more longitudinal modes, even with less than a few milliampere additional increase in injection current. This shift is a discrete jump, as opposed to the previous shifts, which were continuous. This phenomenon is called *mode hopping*.

The mode hopping in the single-mode laser will be illustrated in more detail using Fig. 14.23. Figure 14.23a shows the shape of the gain curve when oscillating in the p_nth longitudinal mode at one temperature. The gain curve has a dip around the wavelength of emission. This is because the heavy consumption of the carriers that participate in the transition of $h\nu_{p_n}$ affects the supply of carriers that participate in the transition near $h\nu_{p_n}$. This dip is called *spectral hole burning*. Because of the spectral hole burning near the center of the gain curve, even though the maximum of the gain curve shifts gradually with temperature, the wavelength of the oscillation hops with increasing injection current in such a sequence as shown in Figs. 14.23b and 14.23c [8].

In the high injection current region, a hysteresis phenomenon is observed in the curve that represents the wavelength versus injection current. The light wavelength follows a different path for increasing current than for decreasing current. This is partially explained by the fact that once the laser starts lasing at a particular wavelength, the stimulated emission tends to stay at that wavelength.

One of the effective countermeasures against mode hopping is mode locking that uses a cladding layer doped with a nonlinear saturable absorber like a Te compound [16] or a multi-quantum-well layer deposited on one of the end reflectors [17]. At lower light intensities the absorber absorbs light, but at higher intensities, the absorber becomes saturated and no longer absorbs. Once one mode establishes a pattern at a high power level, it stays in that mode because of the lower loss.

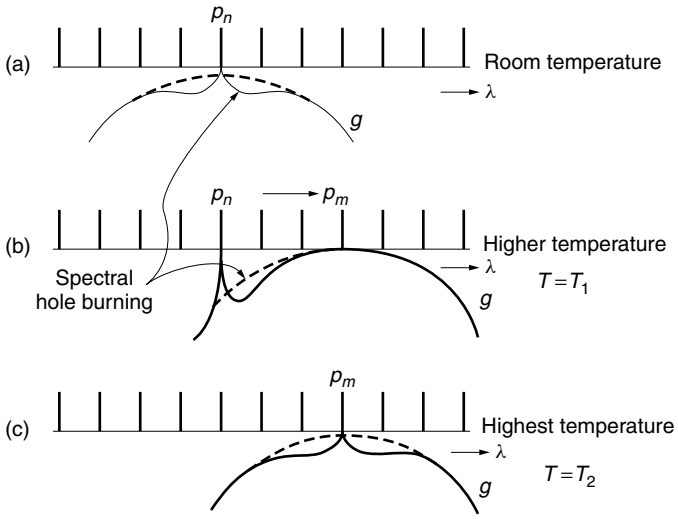


Figure 14.23 Illustration of mode hopping. (After H. Yonezu [8].)

Another method of reducing the wavelength shift due to the temperature effect is to use an externally servocontrolled thermoelectric cooling element.

14.6 BEAM PATTERN OF A LASER

It is important to design the beam pattern of a laser so as to efficiently couple its output power into an optical fiber or any other optical component. For now, the radiation pattern in only the vertical direction (in the x direction in Fig. 14.7) will be considered. As explained in Section 1.2, the expression for the beam pattern is essentially the Fourier transform of the aperture field of the laser.

$$E(\theta) = \frac{1}{j\lambda z}\int_{-\infty}^{\infty} E(x)e^{-j2\pi fx}\,dx\Big|_{f=(\sin\theta)/\lambda} \tag{14.118}$$

where $E(x)$ is the aperture field of the laser in the x direction and $\exp[jk(z_i + x_i^2/2z_i)]$ is suppressed.

Figure 14.24 summarizes the beam patterns for typical one-dimensional aperture field distributions $E(x)$. Figure 14.24a is the beam pattern of a rectangular aperture field distribution

$$E(x) = \Pi\left(\frac{x}{w}\right) \tag{14.119}$$

where w is the width of the aperture field. Insertion of Eq. (14.119) into (14.118) gives the beam pattern in the vertical plane.

$$E(\theta) = \frac{1}{j\lambda z} w\,\mathrm{sinc}\left(w\frac{\sin\theta}{\lambda}\right) \tag{14.120}$$

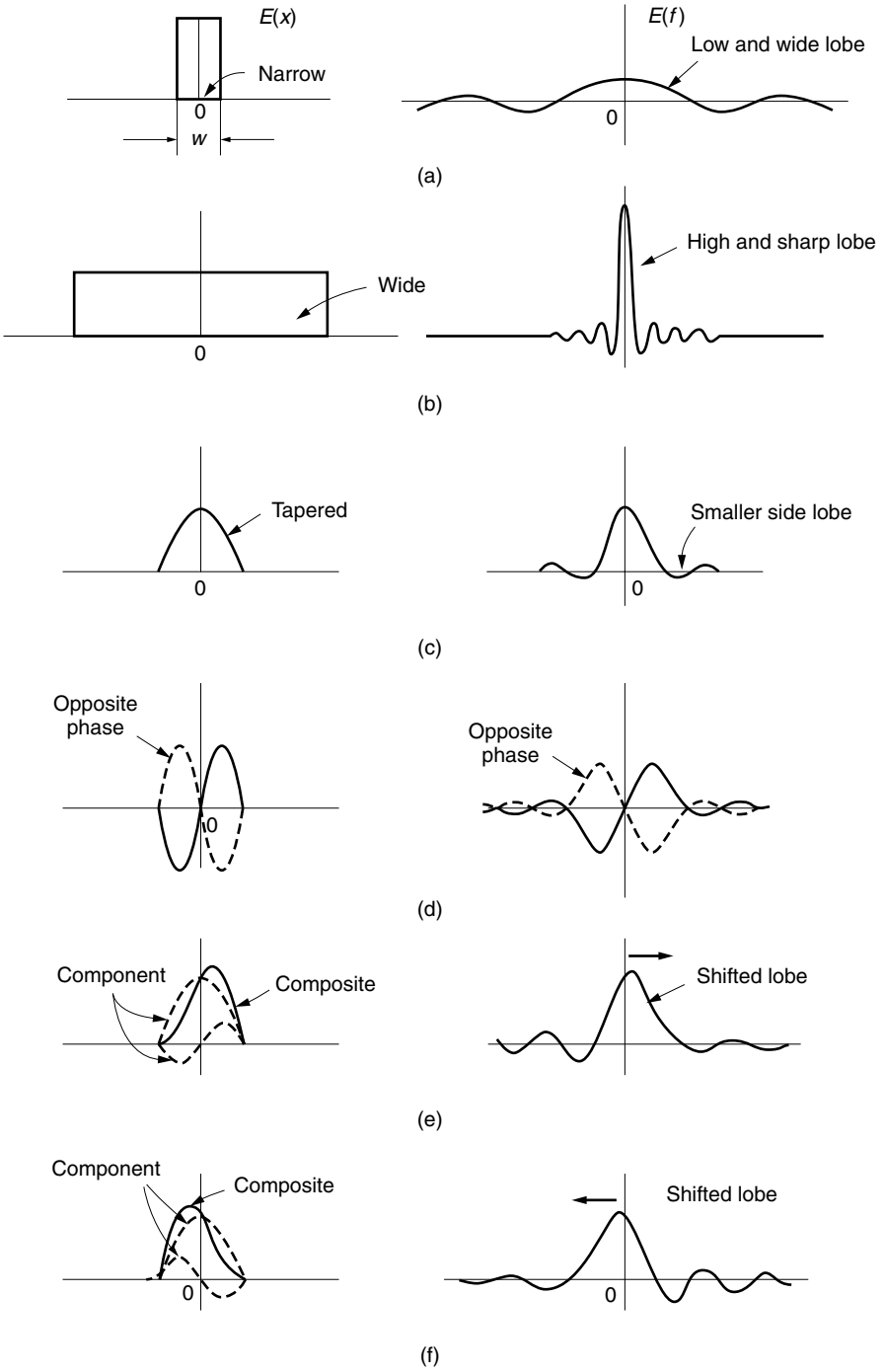


Figure 14.24 Aperture field $E(x)$ and corresponding far-field distribution $E(f)$. (a) Narrow aperture. (b) Wide aperture. (c) Single sinusoidal hump. (d) Double sinusoidal hump. (e) Single + double humps. (f) Single + double (opposite phase) humps.

The beamwidth, which is defined as the width between the nulls of the main lobe, is

$$\Delta\theta = 2\sin^{-1}\left(\frac{\lambda}{w}\right) \tag{14.121}$$

Recognize from Eqs. (14.120) and (14.121) that if a narrow beam is desired, the aperture field w has to be wide, as shown in Fig. 14.24.

It must be cautioned that the width of the aperture field is in general different from the thickness of the active layer of the laser because of the contribution of the evanescent field in the cladding layers. In fact, the spread of the evanescent field is significantly beyond the thickness of the active layer, as will be illustrated in Example 14.3.

Next, the sinusoidal hump aperture field shown in Fig. 14.24c is taken as an example to explain the case when the amplitude is tapered from the center toward the edges of the aperture.

$$E(x) = \Pi(x)\cos(\pi x) \tag{14.122}$$

$$E(\theta) = \frac{1}{2j\lambda z_i}\left[\text{sinc}\left(\frac{\sin\theta}{\lambda} - \frac{1}{2}\right) + \text{sinc}\left(\frac{\sin\theta}{\lambda} + \frac{1}{2}\right)\right] \tag{14.123}$$

For the derivation of Eq. (14.123), see Problem 1.4.

Compared to the case of the rectangular aperture field $\Pi(x)$, the beam of the tapered aperture field has much lower side lobe levels, but the beamwidth is slightly broadened. The lower side lobe levels usually mean better coupling to the optical fiber.

Next, the aperture intensity distribution with two humps, as shown in Fig. 14.24d, will be examined in order to gain some insight into the case when the laser output contains higher transverse modes. The aperture amplitude distribution is

$$E(x) = \Pi(x)\sin(2\pi x) \tag{14.124}$$

The radiation field is

$$E(\theta) = \frac{1}{2\lambda z_i}\left[\text{sinc}\left(\frac{\sin\theta}{\lambda} + 1\right) - \text{sinc}\left(\frac{\sin\theta}{\lambda} - 1\right)\right] \tag{14.125}$$

As shown in Fig. 14.25d, the null appears in the center of the main lobe. This is because the contribution of the aperture field of positive sign exactly cancels the contribution of the negative sign along the centerline. As a matter of fact, any mode with an antisymmetric aperture distribution creates a null in the center of the main lobe, whereas modes with symmetric distributions have a maximum in the center.

Next, the superposition of Figs. 14.24c and 14.24d is considered, as shown in Figs. 14.24e and 14.24f. This situation approximates the case when the active layer of the laser is excited by the fundamental and the next higher order mode. The resultant radiation pattern is distorted and shifted from the center. The direction of the shift depends on the phase of the higher order mode with respect to the fundamental mode. Disturbances such as temperature changes in the active layer affect the phase relationship between the modes so that the central radiation lobe appears to jitter. If the optical fiber does not accept the whole radiation pattern from the laser, this jitter causes a fluctuation in the input power to the fiber and creates noise in the fiber-optic

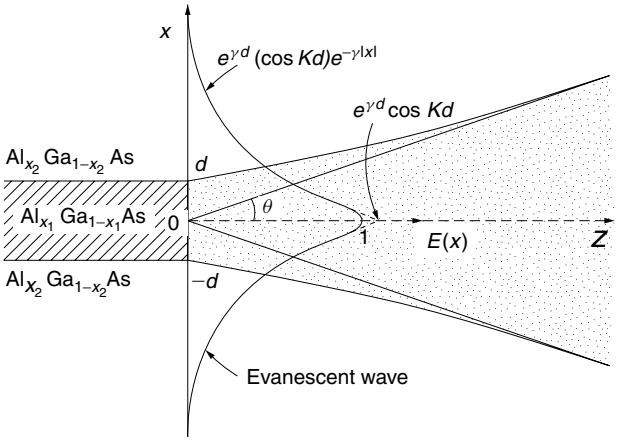


Figure 14.25 Distribution of aperture field intensity.

system. This kind of noise is called the *modal noise* of the laser. Lasers in which only the fundamental mode is excited are naturally free from modal noise.

Example 14.3 Al_xGa_{1-x} As is used for both the active and cladding layers of a laser diode but with different values of fractional molar values of x. If the thickness $2d$ of the active layer is 0.1 μm, what should be the values of x for the active layer and cladding so that the laser oscillates with a half-power beamwidth of $\pm 17^\circ$? The wavelength of oscillation is $\lambda = 0.84$ μm and the bandgap energy for the active layer is 1.462 eV. Assume that the normalized thickness V of the active layer is smaller than $\pi/8$ radians, and assume that the aperture field can be approximated by the expression

$$E(x) = e^{\gamma d}(\cos Kd)e^{-\gamma|x|} \tag{14.126}$$

for both active and cladding layers, as indicated in Fig. 14.25. The TE mode will be used.

Solution The beam pattern is found by inserting Eq. (14.126) into (14.118), but one has to be careful about the absolute value sign. The integral has to be performed for the positive and negative regions separately.

$$|x| = \begin{cases} x & \text{for } x > 0 \\ -x & \text{for } x < 0 \end{cases}$$

$$E(\theta) = C\left[\int_{-\infty}^{0} e^{\gamma x - j2\pi f x}\,dx + \int_{0}^{\infty} e^{-\gamma x - j2\pi f x}\,dx\right]_{f=(\sin\theta)/\lambda} \tag{14.127}$$

where

$$C = \frac{1}{j\lambda z_i}e^{\gamma d}\cos Kd$$

The result of the integration is

$$E(\theta) = C\frac{2\gamma}{\gamma^2 + (k\sin\theta)^2} \tag{14.128}$$

Normalizing Eq. (14.128) by $E(0)$ gives

$$\frac{E(\theta)}{E(0)} = \frac{1}{1 + \left(\dfrac{k \sin\theta}{\gamma}\right)^2} \tag{14.129}$$

The angle θ_b for the half-power is

$$\left(\frac{E(\theta_b)}{E(0)}\right)^2 = \tfrac{1}{2} \tag{14.130}$$

Equations (14.129) and (14.130) give

$$\sin\theta_b = \frac{\gamma}{k}(\sqrt{2} - 1)^{1/2} \tag{14.131}$$

Equation (14.131) means that as γ is decreased the laser beam pattern sharpens. This is because the depth of penetration $1/\gamma$ of the evanescent wave into the cladding layer is increased and the source size is increased.

Next, an approximate expression for γ will be found. Since the second term on the left-hand side of Eq. (9.27), $(Kd)^2 + (\gamma d)^2 = V^2$, is nonnegative, and since $V < \pi/8$ as stated in the problem, then

$$Kd < \frac{\pi}{8} \tag{14.132}$$

which enables us to approximate

$$\tan Kd \doteqdot Kd \tag{14.133}$$

with less than 5% error. This approximation in Eq. (9.61), $\gamma d = Kd \tan Kd$, gives

$$\frac{\gamma}{K} \doteqdot Kd \tag{14.134}$$

Equation (9.27) can be written as

$$V^2 = (Kd)^2 \left[1 + \left(\frac{\gamma}{K}\right)^2\right] \tag{14.135}$$

The second term on the right-hand side in square brackets can be neglected because $Kd < \pi/8$ in Eq. (14.134). Approximating Eq. (14.135) as $V^2 = (Kd)^2$ and using Eq. (9.28), $V = kd\sqrt{n_1^2 - n_2^2}$, leads to

$$K \doteqdot k\sqrt{n_1^2 - n_2^2} \tag{14.136}$$

Finally, putting Eq. (14.136) back into (14.134) gives

$$\gamma = k^2(n_1^2 - n_2^2)d \tag{14.137}$$

Now, let us calculate the fractional molar value x_1 for the active layer. The value of x for $\lambda = 0.84$ μm and $E_g = 1.464$ eV is, from Eq. (14.92c),

$$x_1 = 0.03 \tag{14.138}$$

and n_1 is, from Eq. (14.92d),

$$n_1 = 3.61 \tag{14.139}$$

The necessary equations and numbers are now available to allow us to determine x_2. Equation (14.131) determines the value of γ for a given θ_b. With $\theta_b = 17^\circ$, $\gamma = 3.40\ \mu\text{m}^{-1}$ or the penetration depth $1/\gamma = 0.29\ \mu\text{m}$, which is about three times the thickness of the active layer. The value of n_2 is calculated by introducing the parameters found so far into Eq. (14.137):

$$n_2 = 3.44 \tag{14.140}$$

From Eq. (14.92d), the fractional molar value for the cladding layer is finally $x_2 = 0.42$. The answers are

$$x_1 = 0.03$$
$$x_2 = 0.42 \qquad \square$$

14.7 TEMPERATURE DEPENDENCE OF *L–I* CURVES

One of the weakest points of the semiconductor laser is the strong dependence of the output power on temperature. Figure 14.26 shows the general pattern of the change in the curves of the output power versus bias current with temperature as a parameter [18]. As the temperature increases, the threshold current increases, and the slope quantum efficiency η (or external quantum efficiency) decreases. The slope quantum efficiency η is the derivative of the light output with respect to the injection current at an operating point. Almost all physical parameters depend on the temperature and an analytic expression is too cumbersome to be practical.

An amazing fact, however, is that simple empirical formulas match quite well with the measured characteristics of almost all types of lasers. These empirical formulas are

$$I_{\text{th}} = I_{\text{th0}} e^{T/T_0} \tag{14.141}$$

$$\eta = \eta_0 e^{-T/T_0'} \tag{14.142}$$

where T_0 and T_0' are called the characteristic temperatures of the threshold current and the external quantum efficiency, respectively. I_{th0} and η_0 are the threshold current and the external quantum efficiency projected to zero temperature.

The fluctuation of the output power due to temperature can be estimated immediately for a given I_{th0}, η_0, T_0, and T_0' by noting the fact

$$P = \eta(I - I_{\text{th}}) \quad \text{for } I > I_{\text{th}} \tag{14.143}$$

Insertion of Eqs. (14.141) and (14.142) into Eq. (14.143) gives the temperature fluctuation of the output power of the laser:

$$\frac{P(T + \Delta T)}{P(T)} = e^{-\Delta T/T_0'} \left(1 + \frac{1 - e^{\Delta T/T_0}}{m - 1}\right) \tag{14.144}$$

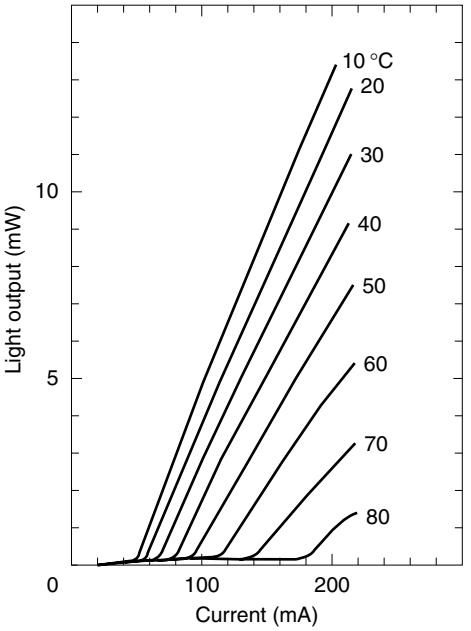


Figure 14.26 Temperature dependence of the L–I curve. (After M. Kitamura et al. [18].)

where

$$m = \frac{I}{I_{\text{th0}} e^{T/T_0}} = \frac{I}{I_{\text{th}}} \tag{14.145}$$

The factor m is called the *overdrive parameter*. The conclusion of Eq. (14.144) is that the larger m is, or equivalently, the smaller the I_{th} of the laser is, the less the temperature-induced power fluctuation of the laser is. This is another important reason why the threshold current of the semiconductor laser has to be kept as low as possible. (Other reasons were a fast turn-on time in Section 14.3.4.1, a faster decay time of the relaxation oscillation in Section 14.3.5.1, and a higher modulation frequency in Section 14.3.5.2.)

The power fluctuation of a GaAs laser is around 2 dB for a temperature excursion of 25 °C. It is much larger than the fluctuation of a light-emitting diode (LED) for the same temperature excursion. Most fiber-optic communication systems require a power stability of better than 0.5 dB. To meet these stability requirements, an electronic circuit for automatic power control (APC) is necessary.

14.8 SEMICONDUCTOR LASER NOISE

Fluctuations in amplitude (AM noise), in frequency (FM noise), and in phase (PM noise) accompany almost all radiating sources to varying degrees. The laser diode is no exception. The importance of the noise depends on how the laser diode is used. What is significant for one application may be insignificant for another application. The same holds true for different techniques used within the same application. For example,

when digital intensity modulation is employed in an optical communication system, the laser noise has little influence on the performance of the system. However, when analog modulation is employed, a much higher signal to noise ratio is demanded, and the laser noise has to be carefully considered. When the laser diode is used as a local oscillator laser, the noise of the laser diode becomes an important entity regardless of whether digital or analog modulation is used. In short, the significance of the noise consideration depends on the usage of the laser.

The following are the major causes of noise from a laser diode:

1. Noise due to external optical feedback.
2. Noise associated with relaxation oscillation.
3. Noise due to mode hopping.
4. Partition noise.
5. Noise due to spontaneous emission.
6. Noise due to fluctuations in temperature and injection current.

14.8.1 Noise Due to External Optical Feedback

The reflectivity R of the mirror of the semiconductor laser is normally small for the benefit of a shorter photon lifetime τ_s, which leads to a higher limit on the modulation frequency ω_0. The smaller R, however, means that the light reflected from the load easily enters inside the active region and the operation of the laser is critically influenced by the reflected wave.

Let us first consider the case where the external reflector [19] is within a few centimeters of the laser, and externally and internally reflected waves are coherent. With the optical feedback, ripples appear in the light power P versus injection current I curve, as shown in Fig. 14.27. The period of the ripple becomes shorter as the reflector moves away from the laser. The cause of the ripple is the external optical feedback. As the injection current is increased, the wavelength of oscillation is shifted (see Section 14.5.1). The phase relationship between the externally reflected and internally reflected waves alternates between in phase and out of phase. If the reflected wave enters the laser in phase with the internal wave that has just reflected from the laser cavity mirror, both waves enhance each other in the active region of the laser and the reflectivity of the mirror is effectively increased. The increase in the effective reflectivity of the mirror increases the lifetime τ_s of photons inside the cavity and, according to Eq. (14.53), increases the output light. Thus, the ripples appear in the P–I curve.

Now consider what happens when the distance between the external reflector and the laser is increased well beyond a few centimeters. Assume for the time being that the light remains coherent. For large distances, the noise appears even for a nominally fixed injection current. The phase ϕ of the wave reflected from the distance L is $\phi = 4\pi L/\lambda$. The change in phase $\Delta\phi$ for a change in wavelength $\Delta\lambda$ is

$$\Delta\phi = \frac{d\phi}{d\lambda}\Delta\lambda$$

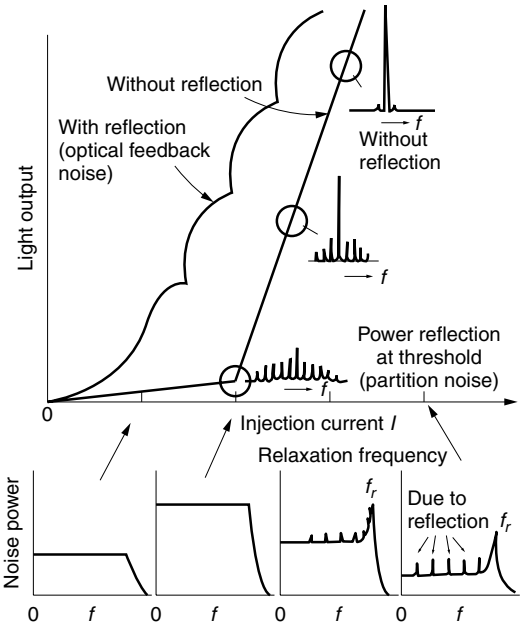


Figure 14.27 Laser diode noise.

and

$$\Delta\phi = -4\pi \frac{L}{\lambda} \frac{\Delta\lambda}{\lambda} \tag{14.146}$$

$\Delta\phi$ is accentuated by a factor of L/λ. This is called the *long-line effect*. For a kilometer-long fiber, the value of L/λ is as large as 10^9 and the wavelength fluctuations due to the current supply or temperature become large enough to change the condition of the interference between the internally and externally reflected waves. This noise appears clearly as notches in the frequency spectrum shown at the bottom of Fig. 14.27.

The countermeasures against the noise due to external optical feedback are:

1. The use of an isolator that prevents the reflected wave from entering the laser diode. The only drawback of this method is that it adds complexity to the fabrication.
2. An increase in the reflectivity R of the end mirrors reduces the noise. The increase in R, however, increases τ_s by Eq. (14.83) and this method is at the sacrifice of lowering the limit of the frequency of amplitude modulation, Eq. (14.89).
3. Removal of the coherency of the output light reduces the noise, at the cost of a reduction in the useful frequency bandwidth. Choose a multilongitudinal mode laser or single longitudinal mode laser whose spectral purity is intentionally degraded by introducing a few hundred megahertz modulation to the injection current [8].
4. Avoid external reflectors wherever possible. If it is not possible to avoid them, reduce the reflectivities of the external reflectors by such methods as antireflective coatings or index-matching fluids.

14.8.2 Noise Associated with Relaxation Oscillation

Relaxation noise occurs when the laser diode is directly pulse modulated by modulating the current. Every time the current pulse is on, the carrier density oscillates in such a way as shown in Fig. 14.11b. This oscillation modulates the plasma effect in the active region of the laser, and the resonance frequency of the cavity is modulated accordingly. This phenomenon is known as the *chirp noise* associated with direct pulse modulation. The countermeasure is to operate the laser at a higher current in order to increase γ in Eq. (14.81) and quickly attenuate the relaxation oscillation.

14.8.3 Noise Due to Mode Hopping

Mode hopping was explained in Section 14.5.2. Every time the transition from one mode to another occurs, the frequency as well as the amplitude of the oscillation jumps. Mode hopping creates both AM and FM noise. As mentioned in Section 14.5.2, the countermeasures are the use of a saturable cladding layer or tight temperature control by a servolooped thermoelectric cooling element.

14.8.4 Partition Noise

Partition noise is quite similar to the mode hopping noise. In the case of the multilongitudinal mode laser, several modes are excited at the same time. The relative phase and amplitude of the modes are fluctuating, even though the total output power looks steady. The partition noise is a source of FM noise and markedly increases near the threshold current. As the injection current is increased, the number of modes that are excited at the same time decreases, and the partition noise decreases.

The countermeasure is to use a single-longitudinal mode laser, such as will be mentioned in Section 14.9. Some multimode lasers when operated at high injection current have only one dominant longitudinal mode.

14.8.5 Noise Due to Spontaneous Emission

Inherent unwanted spontaneous emission is another source of noise. This noise spreads almost over the entire frequency spectrum. The noise spectrum excited by the spontaneous emission, however, has a peak at the relaxation frequency f_r as shown in Fig. 14.27.

One way to minimize the effect of spontaneous emission noise is to increase the injection current, thereby increasing the ratio of stimulated emission preferentially to the spontaneous emission. The ratio A/B of Einstein's coefficients is proportional to ν^3 and higher frequency lasers are more subject to spontaneous emission noise.

14.8.6 Noise Due to Fluctuations in Temperature and Injection Current

Temperature and current fluctuations are likely to influence any radiating source, but the laser diode even more so. The FM noise is the most significant. Typical values of the frequency shift due to temperature is 12 GHz per degree and that due to the injection current is 2 GHz per mA. It is important to control both of these by using an automatic power control circuit and a thermoelectric feedback loop. In the event that a laser diode is used as the local oscillator of a coherent detection system, there is a good

chance that the IF will walk out of the band of the IF amplifier if the thermoelectric feedback is not provided.

In the following sections, lasers with special features will be presented.

14.9 SINGLE-FREQUENCY LASERS

Some applications require true single-frequency lasers. The laser has to be designed to operate in a single longitudinal mode, and also the waveguide of the laser cavity has to be designed to support only the lowest order lateral as well as transverse mode. This is necessary to ensure a clean, single frequency. Figure 14.28 and Example 14.5 illustrate the spectrum splitting that is observed if higher order lateral modes begin to lase.

Let us assume that both the thickness and the width of the active layer have already been designed to have a single mode in these two directions; our attention will be concentrated only on achieving single-mode operation in the longitudinal direction. Laser diodes that operate in the single longitudinal mode (SLM) include the following:

1. Surface emitting (SE) laser.
2. Distributed feedback (DFB) laser, and distributed Bragg reflector (DBR) laser.
3. Diode laser pumped solid-state laser.

14.9.1 Surface Emitting Laser

If the frequency spacing between adjacent longitudinal modes is wide compared to the frequency bandwidth of the laser gain curve, the laser can be operated in a single longitudinal mode. As seen from Eq. (14.34), the spacing between longitudinal cavity modes is inversely proportional to the length of the cavity. A surface emitting laser

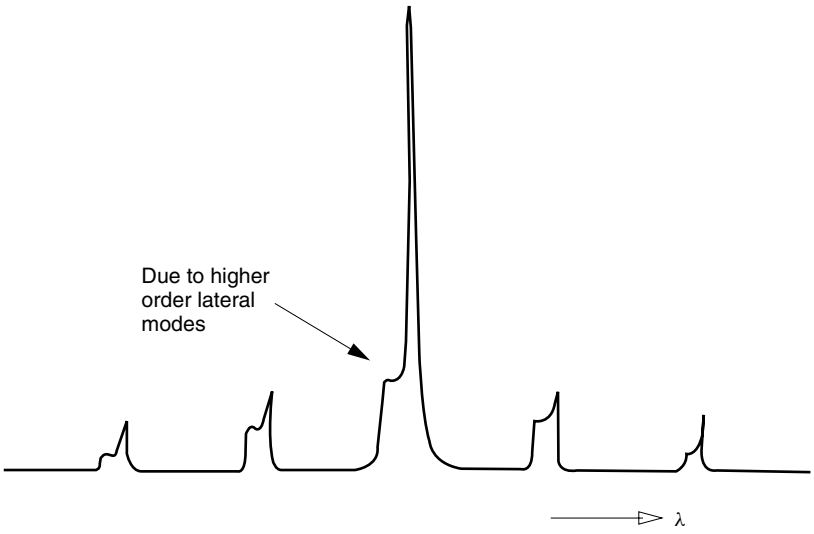


Figure 14.28 Laser spectrum with higher lateral modes.

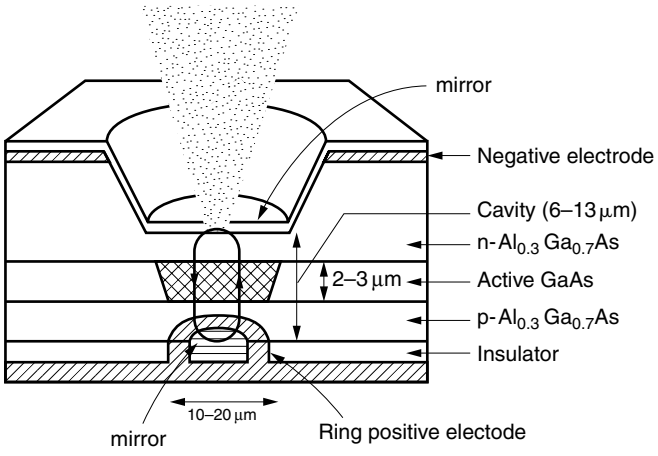


Figure 14.29 Surface emitting (SE) laser. Sometimes this laser is called a vertical cavity surface-emitting laser (VCSEL).

such as shown in Fig. 14.29 has a much shorter cavity length than the edge emitting laser and operates in a single longitudinal mode. The typical cavity length of the SE laser diode is 6–13 μm [20] while that of the normal edge emitting laser diode is 150–250 μm. The wavelength spacing between the longitudinal modes of the SE laser is at least 20 times wider than that of the normal laser diode.

Because the active layer is as short as 2–3 μm, the reflectivity of the mirrors has to be large to meet the threshold condition given by Eq. (14.29). Laser action is obtained by providing high reflectivity mirrors with reflectivities larger than 97%. The SE laser diode is sometimes called a vertical cavity surface-emitting laser (VCSEL).

Other advantages of the VCSEL are the following:

1. Planar technology used to fabricate other electronic or optic components can be used to fabricate the laser as well. Integration of components on a wafer is made easier (monolithic fabrication).
2. Two-dimensional arrays of lasers are possible. These arrays are suitable for the implementation of such devices as optoelectronic interconnects [21].
3. Testing during fabrication is easier.

The disadvantage is the higher threshold current needed to run the laser.

14.9.2 Laser Diodes with Bragg Reflectors

The majority of the end surfaces of the laser cavities so far mentioned have been cleaved facets. The output frequency of the laser diode can be stabilized significantly if the cleaved end facets are replaced by gratings, because the grating selectively reflects a particular wavelength of light while the cleaved surface does not have such selectivity. Figure 14.30 shows a diagram of a distributed Bragg reflection (DBR) [22] laser that employs the sharp frequency dependence of a grating. One or both facets are replaced by corrugated reflectors.

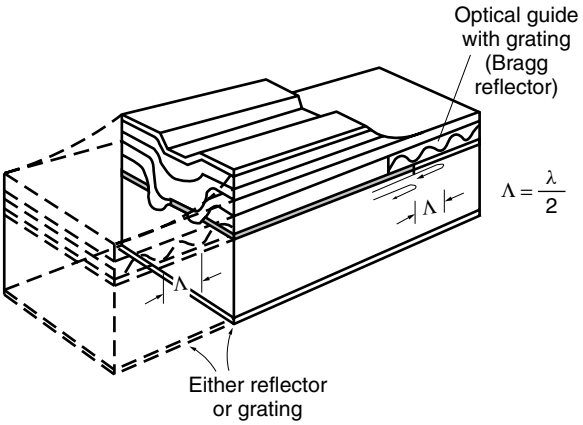


Figure 14.30 Distributed Bragg reflection (DBR) laser.

The corrugated optical guide used as the Bragg reflector does not necessarily have to be located at the end of the cavity. It can be inside the active region or a separate optical guide adjacent to the active layer or the cladding layer. Such a laser is called a distributed feedback (DFB) laser. The DFB laser shown in Fig. 14.31 has the corrugation in a separate guide in order to reduce the loss in the active layer. Both DBR and DFB lasers operate on the same principle.

A heuristic explanation will now be given with a DFB laser whose active region is corrugated along the entire length as shown in Fig. 14.32. Figure 14.32 highlights two possible paths inside the active region. One path is for light that passes straight through the corrugation. The other path is light that is doubly reflected, first at PP' and then at QQ'. For the growth that leads to laser oscillation, it is necessary for all such paths to interfere constructively.

The phase relationship for constructive interference will be found first [23]. Let us assume that the light satisfying the Bragg condition reflects and reverses its direction of propagation when it encounters a peak of the corrugation, but not between the peaks. The reflection centers are discretely spaced by $\lambda/2$. Referring to Fig. 14.32, the

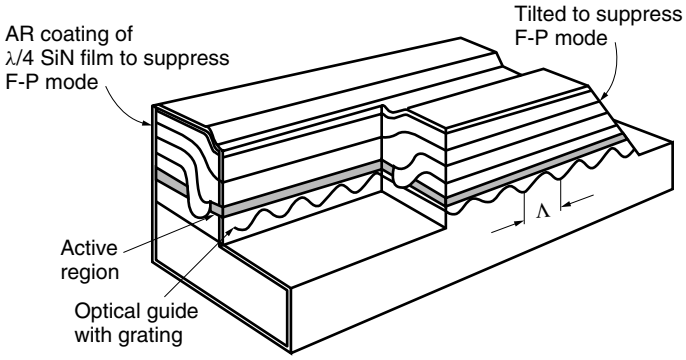


Figure 14.31 Distributed feedback (DFB) laser.

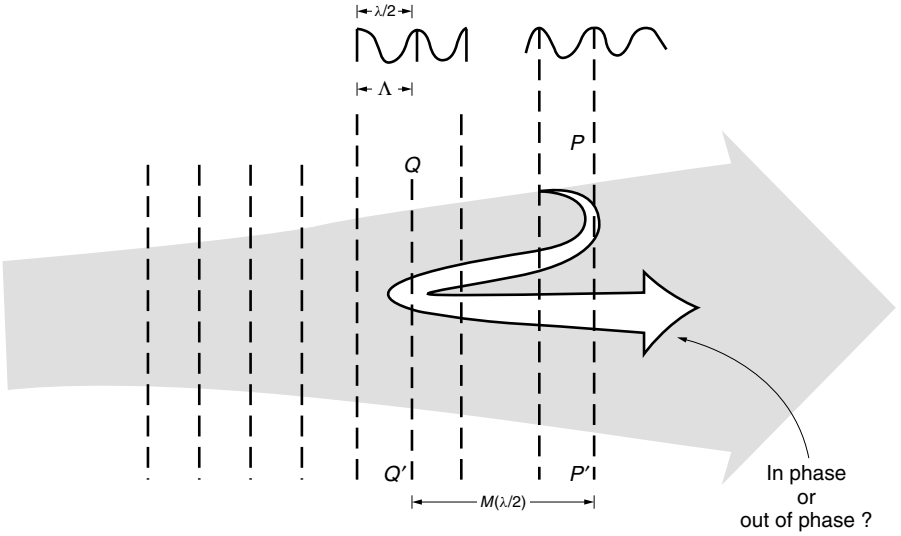


Figure 14.32 Resonance condition of the DFB laser cavity. The difference between the straight-through path and the doubly reflected path is $m\lambda$.

difference between the straight-through path and the doubly reflected path is $m\lambda$, where m is an integer. Does this mean that the lightwaves propagating in the same direction after reflection are always in phase if the frequency satisfies the Bragg condition? The answer is no. The phase shift at each reflection center of the corrugated dielectric grating was left out.

Recall Eq. (5.55) of the increment of the wave joined after each reflection in coupled-mode theory.

$$R(x) = tR_0 + jrS_0 \tag{14.147}$$

The last term on the right-hand side is interpreted as the contribution of the wave reflected from the discontinuity. Note that this term is an imaginary number and has a phase shift equal to $\pi/2$ radians. The reflection centers QQ' and PP' pictured in Fig. 14.32 each impart a $\pi/2$ phase shift to the reflected light for a total phase shift of π radians with respect to the straight-through path. The doubly reflected light destructively interferes with the straight-through light, and the amplitude does not grow.

We are going to stretch our simple approach further and use basic arithmetic to analyze the situation.

When Bragg's condition is satisfied, the reflections from the corrugated reflectors enhance each other. Bragg's condition that the reflected wave is pointed backward in the axial direction is

$$2\Lambda\beta_0 = 2q\pi \tag{14.148}$$

where Λ is the period of the corrugation, β_0 is the propagation constant of the active region, and q is an integer. Equation (14.148) can be rewritten in terms of frequency as

$$\nu_{Bq} = \frac{v}{2\Lambda}q \tag{14.149}$$

The nth cavity resonance frequency including the above-mentioned π-radian phase shift at both ends is obtained as

$$\frac{4\pi}{v}\nu_{cn}L_e + \pi = 2\pi n \tag{14.150}$$

where L_e is the effective length of the cavity and is represented by a large integer number ($\sim$1000) M as

$$L_e = M\Lambda \tag{14.151}$$

From Eqs. (14.150) and (14.151), the nth cavity resonance frequency is

$$\nu_{cn} = \frac{v}{2\Lambda}\frac{1}{M}\left(n - \tfrac{1}{2}\right) \tag{14.152}$$

Figure 14.33 shows the frequencies of the two types of resonances. Visualize two sets of combs of frequency spectra ν_{Bq} and ν_{cn} whose spacing of spectra (teeth of the comb) are vastly different. The spacing of ν_{cn} is about 1000 times more dense than that of ν_{Bq}. The optimum lasing frequency is where the two frequencies overlap; that is,

$$\nu_{cn} = \nu_{Bq} \tag{14.153}$$

From Eqs. (14.149), (14.152), and (14.153), the ideal integer is

$$n = Mq + \tfrac{1}{2} \tag{14.154}$$

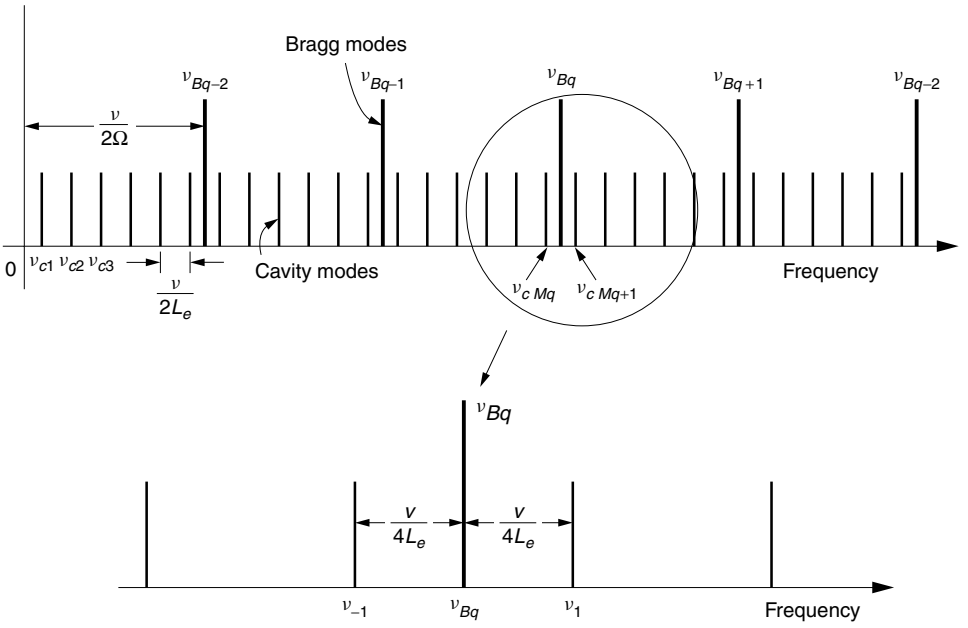


Figure 14.33 Relationship between the Bragg modes and cavity modes of a DFB laser. (After H. Kogelnik and C. V. Shank [23].)

Unfortunately, there is no integer n that can satisfy Eq. (14.154) and the optimum condition does not exist by itself. Let us closely examine the spectrum ν_{cn} near $n = Mq$ using Eqs. (14.149), (14.151), and (14.152). The frequencies ν_{cn} and $\nu_{c(n+1)}$ will be relabeled as ν_{-1} and ν_1, respectively.

$$\begin{aligned} \nu_{-1} &= \nu_{Bq} - \frac{v}{4L_e} \\ \nu_1 &= \nu_{Bq} + \frac{v}{4L_e} \end{aligned} \tag{14.155}$$

The spectrum is shown in the bottom figure of Fig. 14.33. There are two cavity resonance frequencies symmetric with ν_{Bq} but there is no ideal cavity resonance frequency that coincides with the Bragg frequency. It is only after shifting the cavity resonance frequency spectrum to match the Bragg frequency that the optimum operation of the DFB laser can be achieved. Such means are described next.

14.9.3 λ/4 Shift DFB Lasers

The DFB laser [24] oscillates at two frequencies ν_{-1} and ν_1 as shown in Fig. 14.33 because of the π-radian term on the left-hand side of Eq. (14.150). This may be remedied by adding another π radians to the phase of the doubly reflected wave. The additional π-radian phase shift can be created by introducing a $\pi/2$-radian phase shift in the corrugated grating. Referring once again to Fig. 14.32, imagine that the peak at PP' is shifted over by $\lambda/4$. The distance between PP' and QQ' becomes $(m\lambda/2 + \lambda/4)$, and the total path difference for the doubly reflected path becomes $(m\lambda + \lambda/2)$. In terms of phase, this difference is $2m\pi + \pi$, and thus a π-radian phase shift has been introduced to counteract the π radians from the reflections.

Figure 14.34 illustrates methods of operating the DFB laser on a single longitudinal mode. In Fig. 14.34a, the $\lambda/4$ phase shift was not introduced, and the two frequencies ν_1 and ν_{-1} in Eq. (14.155) are present. In Figs. 14.34b–d, a $\lambda/4$ phase shift was introduced by three different methods. In the first method, the $\lambda/4$ shift is introduced by introducing a $\Lambda/2$ shift in corrugation, noting $\lambda = 2\Lambda$. This shift is achieved by removing one-half period of the corrugation, in other words, by reversing the phase of the grating in the right half of the corrugation. This kind of grating was fabricated by projecting an interference pattern made by He–Cd laser beams (0.4416 μm) into photoresist. The phase reversal of the corrugation is achieved by using a combination of negative and positive photoresists or by inserting a phase shift plate in one side of the corrugated grating. The frequency of the $\lambda/4$ shift DFB laser is exactly on the Bragg frequency, as indicated on the left of Fig. 14.34b.

The $\lambda/4$ phase shift can also be created by etching off some of the height of the corrugation at an empirically determined height Δh and length Δl, as shown in Fig. 14.34c.

The next method shown in Fig. 14.34d uses an external reflector in order to create the required π-radian shift in the cavity. The cleaved facet of the laser cavity is first coated with an SiN film and then a gold film reflector. The thickness of the SiN film is designed to create the phase shift needed for the DFB laser to oscillate at the Bragg frequency.

The last method shown in Fig. 14.34e operates in a slightly different manner [25]. Multielectrodes are deposited on a DBR laser as shown in Fig. 14.35. The left section

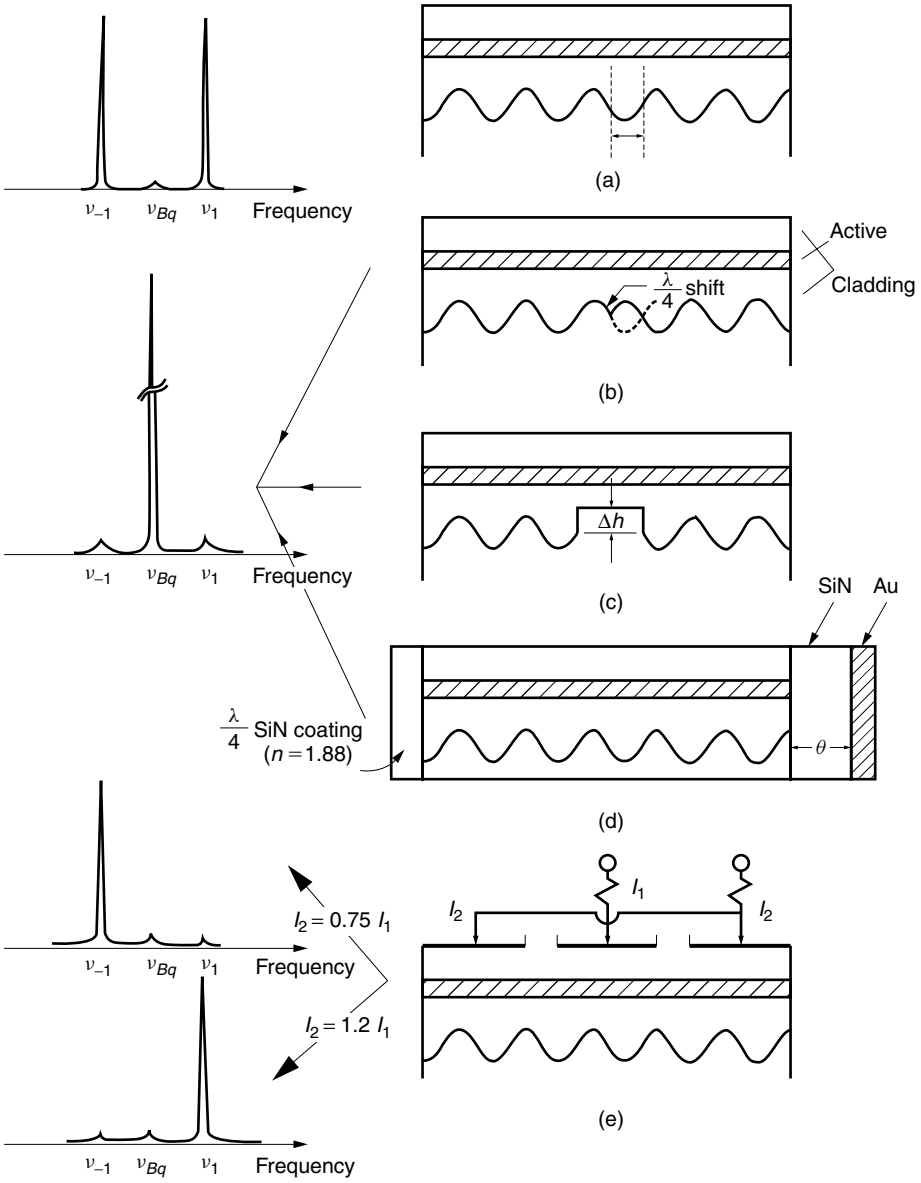


Figure 14.34 Single longitudinal mode (SLM) operation of the DFB lasers. (a) Corrugated grating. (b) $(\Lambda/2 = \lambda/4)$ shift in grating. (c) Phase shift by etching. (d) Phase shift by coating. (e) Phase control by multielectrodes.

plays the role of the active region, the center section plays the role of the phase control region, and the right section acts as a fine tunable Bragg reflector. The tuning resolution is so fine that only one of the two possible modes is differentially favored, so as to oscillate at one frequency. The frequency of the laser can be either ν_{-1} or ν_1 depending on the ratio of the current I_1 of the center electrode and the current I_2 of the right electrode.

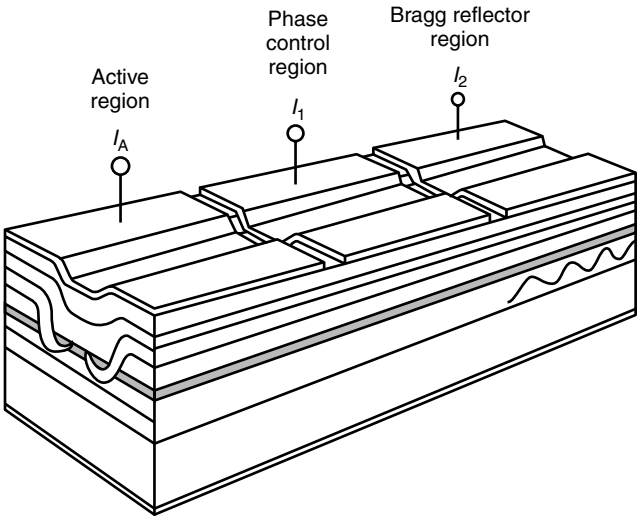


Figure 14.35 Combination of DFB and DBR lasers for wavelength tuning.

An increase in the injection current increases the carrier concentration and that, in turn, decreases the index of refraction because of the plasma effect (see Section 14.4.3.2). The decrease in the index of refraction n increases v and hence increases the Bragg frequency expressed by Eq. (14.149). In short, an increase in the injection current increases the frequency of the Bragg reflection. Thus, an increase in current I_2 shifts the optimum feedback efficiency to a higher frequency and switches the laser frequency to ν_1. Conversely, a decrease in current I_2 switches the frequency to ν_{-1} as indicated in Fig. 14.34e. It should be added that by shifting both I_1 and I_2 in the proper ratios, the position of ν_1 (or ν_{-1}) can be shifted continuously.

Example 14.4 Due to the ease of fabrication, the third-order Bragg frequency is often used. In this case, some light reflects in directions other than the exact backward direction and creates radiation loss as well as stray light. Find the general expression for the direction of radiation from the corrugated grating fabricated in the cladding layer and then find the directions of scattering other than the backward direction.

Solution With the geometry as shown in Fig. 14.36, let us assume that the sources of radiation are the bottom points of the corrugation because they are the points of the largest interaction with the active region underneath. The propagation constant of the evanescent wave is the same as the propagation constant β_0 of the active region. Referring to point O, the phase at P' is

$$\phi_{P'} = \beta_0 \Lambda \tag{14.156}$$

The phase at P of the wave emanating from point O is

$$\phi_P = n_2 k \Lambda \cos\theta \tag{14.157}$$

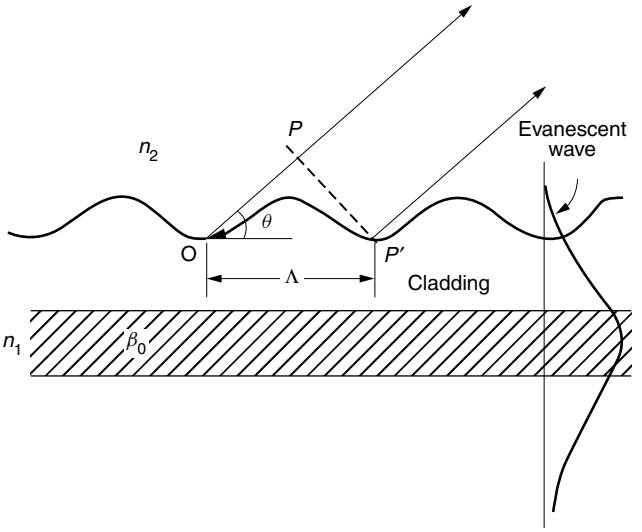


Figure 14.36 Directions of the radiation from the corrugated grating.

where n_2 is the index of refraction of the radiating medium and k is the free-space propagation constant. The angle θ is the angle of radiation. The peaks of the radiation pattern appear in such directions that the waves at P and P' are in phase.

$$\beta_0 \Lambda - n_2 k \Lambda \cos\theta = 2m\pi \tag{14.158}$$

$$\cos\theta = \frac{\beta_0 \Lambda - 2m\pi}{(n_2 k/\beta_0)\beta_0 \Lambda} \tag{14.159}$$

It should be realized that the backscattered wave ($\theta = \pi$) is special in that it propagates exclusively inside the core medium and never goes out of the core and its propagation constant is β_0 rather than $n_2 k$. The value of Λ for this case is obtained from Eq. (14.159), but by replacing $n_2 k$ by β_0 it is obtained directly from Eq. (14.148). The required corrugation period of the qth order Bragg reflection is hence, $\Lambda = 2q\pi/\beta_0$. The directions of radiation other than the backward direction are obtained by inserting this value of Λ into Eq. (14.159) as

$$\cos\theta = \frac{N_0}{n_2}\left(1 - 2\frac{m}{q}\right) \tag{14.160}$$

where $N_0 = \beta_0/k$ is the effective refractive index for the fundamental mode, Eq. (9.44), and is close to n_1. Solutions for the third-order frequency, $q = 3$ with the assumption $N_0 \doteqdot n_2$, are

θ	m
0°	0
70°	1
110°	2
180°	3

Phasefronts drawn to scale are shown in Fig. 14.37. Wavefronts scattered from the adjacent scattering centers are seen in phase. □

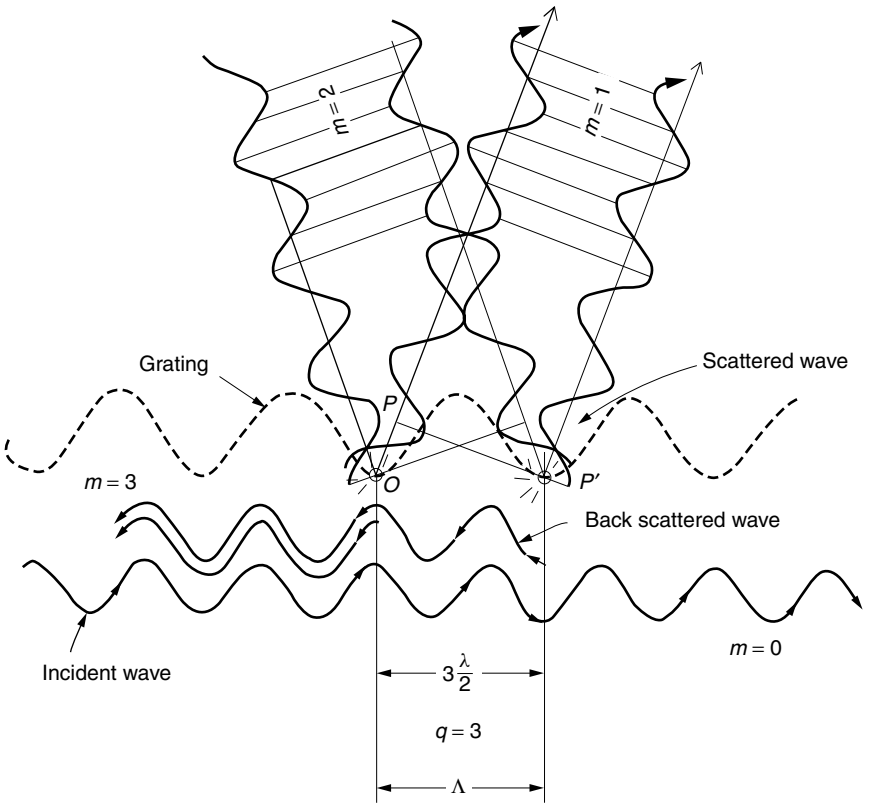


Figure 14.37 Wavefront of the scattered wave from a corrugated grating with $q = 3$, $\Lambda = 3\lambda/2$.

Example 14.5 Taking the example of a metal resonator box, examine the influence of the lateral modes. The resonance mode TM_{nmp} of a metal resonator box is considered. The metal box has the same dimensions as the active layer of the semiconductor laser shown in Fig. 14.38.

Solution The solution

$$E = A \sin k_x x \cdot \sin k_y y \cdot \sin k_z z \tag{14.161}$$

has to satisfy

$$k_x^2 + k_y^2 + k_z^2 = (n_1 k)^2 \tag{14.162}$$

in order to satisfy the wave equation

$$\nabla^2 E + (n_1 k)^2 E = 0 \tag{14.163}$$

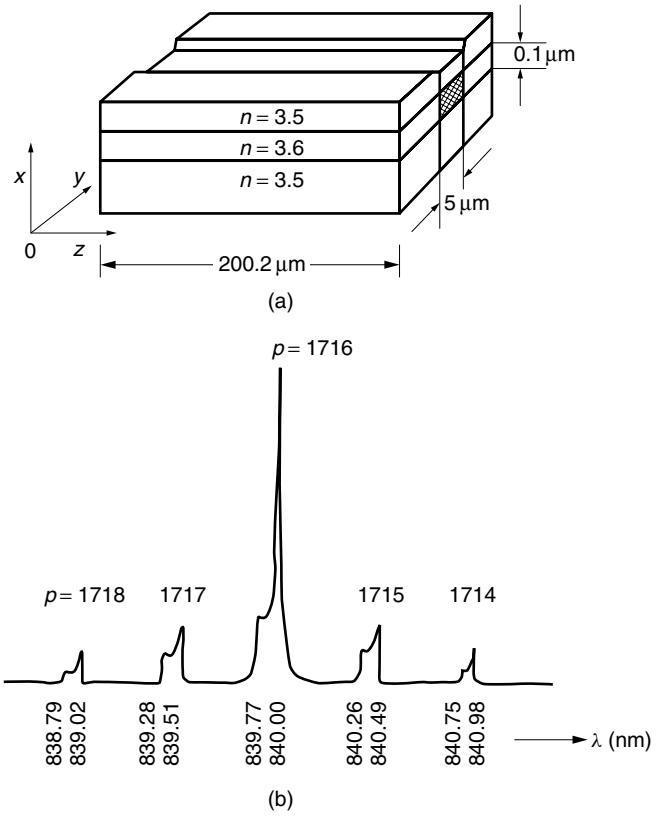


Figure 14.38 Influence of the lateral modes on the resonance wavelength. (a) Dimensions of a laser diode. (b) Spectrum of the modes of a resonator metal box that has the same dimensions as the active layer of the laser diode shown in (a).

k_x, k_y, k_z are determined by the boundary conditions that $E = 0$ on the box walls. Inside the rectangular cavity with dimensions L_x, L_y, L_z,

$$\begin{aligned} k_x L_x &= n\pi \\ k_y L_y &= m\pi \\ k_z L_z &= p\pi \end{aligned} \tag{14.164}$$

Insertion of Eq. (14.164) into (14.162) gives

$$\left(\frac{n\pi}{L_x}\right)^2 + \left(\frac{m\pi}{L_y}\right)^2 + \left(\frac{p\pi}{L_z}\right)^2 = (n_1 k)^2$$

Hence,

$$\lambda = \frac{2n_1}{\sqrt{\left(\dfrac{n}{L_x}\right)^2 + \left(\dfrac{m}{L_y}\right)^2 + \left(\dfrac{p}{L_z}\right)^2}} \tag{14.165}$$

The value of p can be determined by inserting the given condition $n = m = 0$ and $\lambda_{00p} = 0.84$ μm into Eq. (14.165):

$$p = \frac{2n_1}{\lambda_{00p}} L_z = \frac{2(3.6)(200.2)}{0.84} = 1716$$

Then, wavelength λ_{01p} of TM_{01p} is

$$\lambda_{01p} = \frac{2n_1}{\sqrt{\left(\frac{m}{L_y}\right)^2 + \left(\frac{p}{L_z}\right)^2}} = \frac{2(3.6)}{\sqrt{\left(\frac{1}{5}\right)^2 + \left(\frac{1716}{200.2}\right)^2}}$$

$$= 8.3977 \text{ nm}$$

A table of calculated values is given below. Note that λ_{01p} is always shorter than λ_{00p}.

p	λ_{00p} (nm)	λ_{01p} (nm)
1714	840.98	840.75
1715	840.49	840.26
1716	840.00	839.77
1717	839.51	839.28
1718	839.02	838.79

Figure 14.38b shown the spectrum and resembles Fig. 14.28. □

14.9.4 Diode Pumped Solid-State Laser

A neodymium-doped yttrium–aluminum–garnet (Nd:YAG) laser is a laser that can be operated in a single longitudinal mode over a much longer range of cavity lengths than the semiconductor lasers. The neodymium YAG laser is not a *semiconductor* laser but is a *solid-state* laser [1]. Contrary to the semiconductor lasers, the solid-state lasers have discrete energy levels and the frequency bandwidth of the gain curve is about one-thousandth of those of the semiconductor lasers. This makes single longitudinal mode operation easier.

Figure 14.39a shows the geometry and Fig. 14.39b shows the dimensions of a ring resonator type solid-state laser. One immediately realizes that the dimensions of the cavity are about a hundred times bigger than that of a semiconductor laser. Even with these large cavity dimensions, the laser can still maintain single longitudinal mode operation because of the narrow frequency bandwidth of the gain curve of the solid-state laser.

For a 2.5-mm full roundtrip laser cavity, the frequency spacing of the cavity modes $\Delta f = c/2n_1L$ is 33 GHz. The frequency bandwidth of the Nd:YAG gain curve is around 1.2 GHz, and there is no chance that two longitudinal modes are excited simultaneously. With the proper design and alignment of the reflectors, it is fairly easy to suppress the higher transverse modes, so that the Nd:YAG laser can be made to oscillate in a single transverse mode as well as a single longitudinal mode.

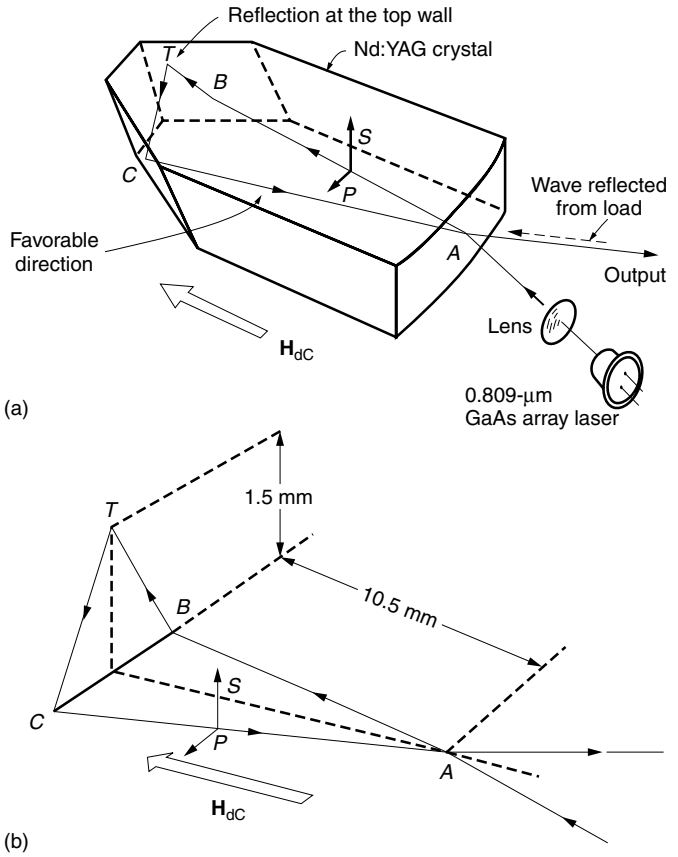


Figure 14.39 Diode pumped solid-state ring laser. (a) Geometry of MISER laser. (b) Light path in the MISER cavity.

The neodymium YAG unidirectional ring laser has such advantages as:

- Single-frequency spectrum.
- Narrow laser linewidth with a short-term (1 ms) frequency bandwidth of 5 kHz and a long-term bandwidth of 100 kHz.
- Immunity to reflections from the load.
- No spatial hole burning in the laser cavity (see Section 14.4).

The disadvantages are:

- Nonplanar structure.
- Difficulty of direct modulation of either frequency or amplitude.
- More complicated structure than semiconductor lasers because the laser has to be optically pumped by a GaAs semiconductor laser of 0.809-μm wavelength of moderately high power.

The monolithic isolated single-mode end-pumped ring (MISER) [26] laser is an example of a laser that took advantage of the large dimensions of the solid-state laser cavity as well as the magnetooptic property of a Nd:YAG crystal.

Next, an explanation will be given on how to accomplish the unidirectional resonance in the MISER laser. As shown in Fig. 14.39b, the full round trip takes the path *ABTCA*, which forms two isoclinal triangles, *ABC* in the horizontal plane and *BTC* in the vertical plane. Total internal reflection is used as mirrors at *B*, *T*, and *C*. Recall that the phase shifts associated with total internal reflection for the $E_{\parallel}$ and $E_{\perp}$ components are different as expressed by Eqs. (2.88) and (2.89). In general, the combination of the mirror surface tilt and the differential phase shift between $E_{\parallel}$ and $E_{\perp}$ creates an elliptically polarized wave. Let us refer to this effect as a geometrically induced change in the state of polarization. If a magnetic field is applied to the crystal as indicated in Fig. 14.39, the Faraday effect (see Section 5.4.1), which is a magnetooptic effect, will also cause a change in the state of polarization. For one direction of travel around the ring, the Faraday effect acts to subtract the geometrically induced polarization change, while for the other direction of travel, the Faraday effect increases the polarization change. At the output coupler, the reflectivity depends on the state of polarization, and the polarization with the highest reflectivity is favored to oscillate, while the oscillation of the other polarization is suppressed. Lasing occurs in one direction of travel only, and hence the ring resonator is termed unidirectional.

What is the advantage of a unidirectional ring resonator over bidirectional resonators? The answer is higher single-mode power. To understand why, the process of spatial hole burning needs to be explained. When the axial mode begins to lase in a bidirectional-type resonator, a standing-wave pattern is created inside the cavity. The locations of the maxima and the nulls are fixed in space. There is no stimulated emission at the null positions. The spatially modulated gain is referred to as spatial hole burning. In the unidirectional resonator, there is a traveling wave excited in one direction but not the other. There is no standing-wave pattern, and therefore no spatial hole burning. The traveling-wave-type laser cavity permits a larger maximum output light power into a single axial mode than the standing-wave-type cavities.

Any light reflected back into the MISER cavity from the load has to go around the cavity in the unfavorable direction, and the laser operation is far less susceptible to externally reflected waves.

Example 14.6 What is the maximum full-round-trip length in the MISER laser cavity that still maintains the single longitudinal mode operation? Assume the frequency bandwidth of the gain curve is 1.2 GHz and the index of refraction of the Nd:YAG crystal is $n = 1.82$. Also assume that the lasing mode coincides with the center of the gain curve.

Solution As shown in Fig. 14.40, a frequency spacing of 0.6 GHz will ensure a single longitudinal mode if the cavity mode is at the gain center. Hence, from Eq. (14.36),

$$L \leq \frac{c}{2\Delta f n} = 13.7 \text{ cm} \tag{14.166}$$

This L is substantially longer than the 6 –13-μm cavity length required by the surface-emitting semiconductor lasers described in Section 14.9.1 for single longitudinal mode operation. □

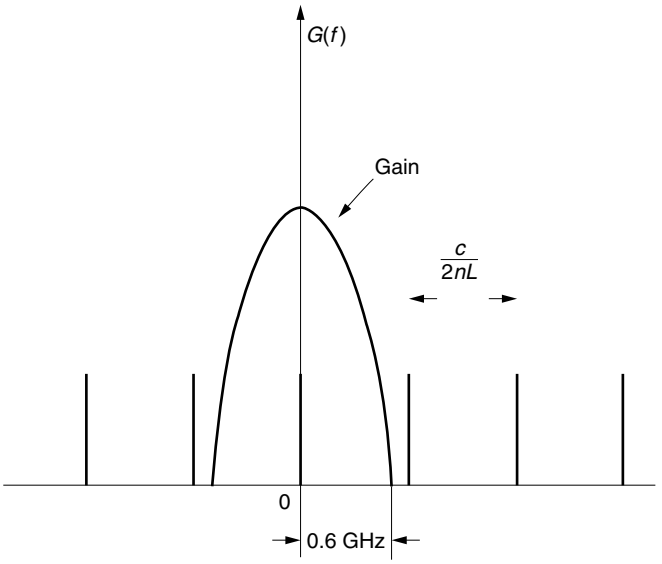


Figure 14.40 Design of the cavity for a Nd:YAG MISER laser.

14.10 WAVELENGTH TUNABLE LASER DIODE

Wavelength tunable lasers are needed as devices for wavelength division multiplexing and as local oscillator lasers for coherent detection. A description of the general principle of wavelength tuning of the DFB laser is given first, followed by that of the superstructure grating (SSG) laser.

14.10.1 Principle of Frequency Tuning of the DFB Laser

With regard to the structure of the DFB laser shown in Fig. 14.41, the lasing conditions will be found [27,28]. The ratio of the fields before and after one trip is

$$rRe^{-j\phi_1 - j\phi_2 - 2(\alpha - g)L} \tag{14.167}$$

where r is the reflection coefficient of the Bragg reflector and R is that of the end mirror, ϕ_1 is the phase shift taking place in the Bragg reflector, and ϕ_2 is the phase delay during one round-trip travel. Let

$$\phi = \phi_1 + \phi_2 \tag{14.168}$$

$$A = rRe^{-2(\alpha - g)L} \tag{14.169}$$

The threshold condition that the wave is enhanced after each trip is

$$Ae^{j\phi} = 1 \tag{14.170}$$

Equation (14.170) reduces to

$$A = 1 \tag{14.171}$$

$$\phi = 2m\pi \tag{14.172}$$

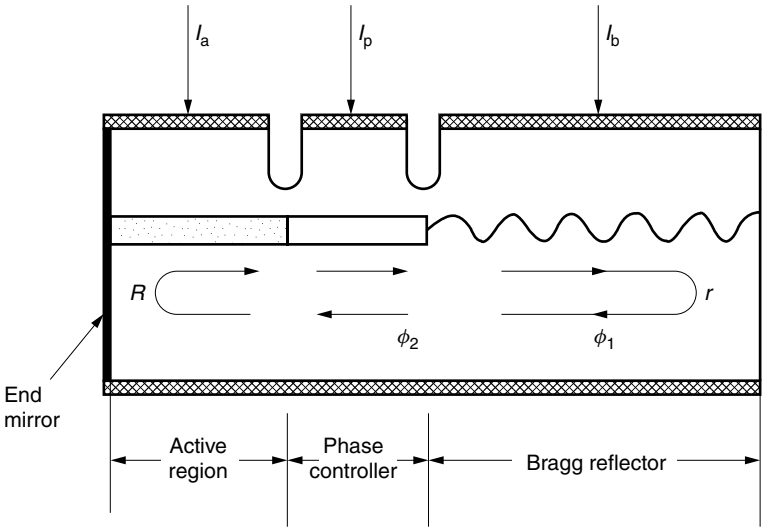


Figure 14.41 Structure of the DFB laser.

Equation (14.171) is the threshold condition that after each round trip the amplitude starts to grow. Equation (14.172) is the condition that after each round trip constructive interference takes place.

The values of r and ϕ_1 of the Bragg reflector are obtained from coupled mode theory [29]:

$$r = \frac{-jK \sinh \gamma L}{\cosh \gamma L + (d_i + j\Delta\beta) \sinh \gamma L} \tag{14.173}$$

where

$$\gamma^2 = K^2 + (d_i + j\Delta\beta)^2 \tag{14.174}$$

$$\Delta\beta = \beta - \beta_0 = \frac{2\pi}{\lambda} n_b - \pi/\Lambda \tag{14.175}$$

and where Λ is the period of the corrugation of the grating, K is the coupling coefficient of the corrugation of the grating, d_i is the loss constant of the material from which the Bragg reflector is made, and n_b is the equivalent refractive index of the Bragg reflector.

Next, the shift of the lasing wavelength due to the injection of the tuning current will be explained graphically. First, the lasing conditions, Eqs. (14.168) and (14.172), are rewritten as

$$\begin{aligned} Y_1 &= -\phi_1 \\ Y_2 &= \phi_2 - 2m\pi \end{aligned} \tag{14.176}$$

and the lasing condition is

$$Y_1 = Y_2 \tag{14.177}$$

Figure 14.42 shows a typical graph of the Bragg reflector loss and phase shift as a function of wavelength. The upper curve shows a quantity related to the mirror loss, that is, the light that does not reflect back from the Bragg reflector. The phase shift Y_1 due to the reflection by the Bragg reflector is shown by the lower curve.

Figure 14.43 shows the plot of ϕ_2 that consists of the phase shift during the round trip and

$$\phi_2 = \frac{2\pi}{\lambda}(n_p l_p + n_a l_a) \tag{14.178}$$

where n_p and n_a are the refractive indices of the phase controller and the active region, respectively, and l_p and l_a are their lengths.

The phase ϕ_2 decreases almost linearly with λ for a small range of λ. The condition of Eq. (14.172) repeats every 2π radians. Y_2 (solid lines) represents the phase ϕ_2 in the region of 0 to 2π radians and Y_2 is obtained from ϕ_2 by subtracting $2m\pi$, where m is an integer called the *longitudinal mode number* of the laser. The shift of ϕ_2 when n_p and n_a are decreased by increasing the tuning current is shown by the dotted lines. The curve of Y_2 shifts toward a shorter wavelength.

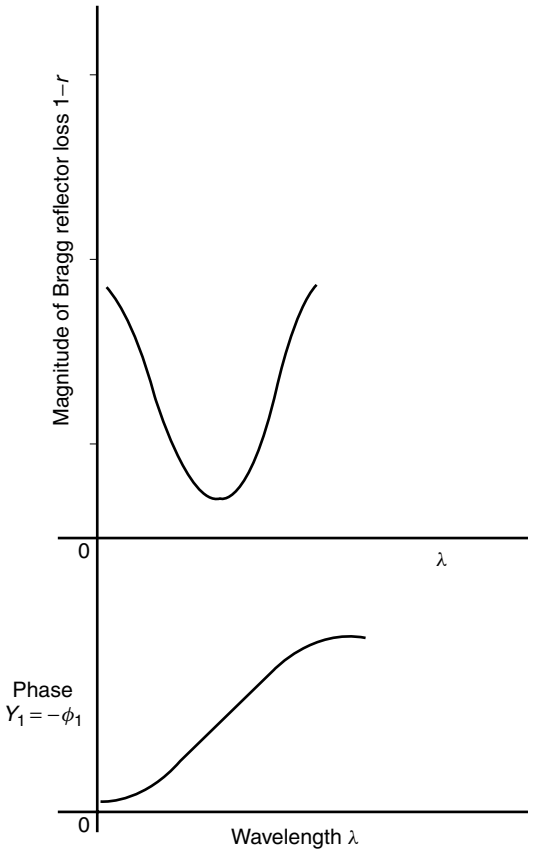


Figure 14.42 Characteristics of the Bragg reflector with respect to wavelength.

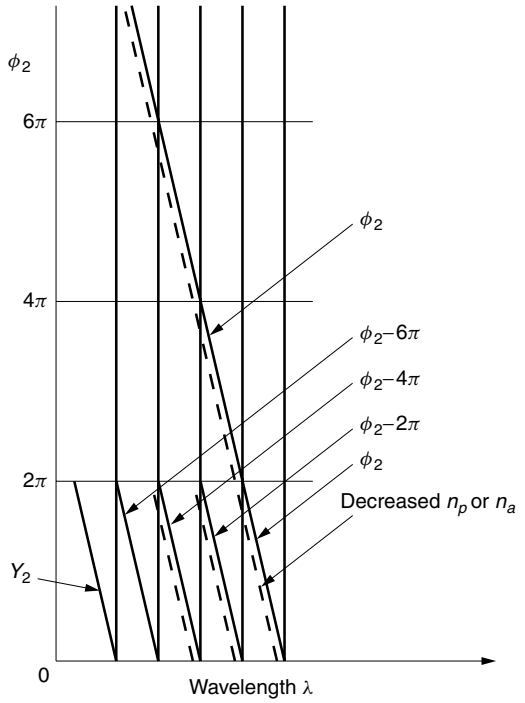


Figure 14.43 Phase delay ϕ_2 associated with one round trip, which continually increases with $1/\lambda$. The phase of the mth longitudinal mode is represented by $Y_2 = \phi_2 - 2\pi m$.

14.10.1.1 Tuning of Wavelength by the Phase Controller Tuning Current I_p Alone

Referring to the solid lines in Fig. 14.44, $Y_1 = Y_2$ is satisfied at four intersections 1, 2, 3, and 4 in the phase diagram. The corresponding mirror losses are indicated by circles in the upper curve.

The threshold point is determined by the smallest mirror loss. Among the four wavelengths, the mirror loss is the smallest at point 2 and lasing starts at this wavelength.

As I_p is increased, the Y_2 curves start to shift toward the left and the intersections move to $1'$, $2'$, $3'$, and $4'$. For instance, the operating point 2 shifts to $2'$.

As I_p is increased, point 2 climbs to a higher point $2'$ on the mirror loss in the upper curve while point 3 slides down to a new point $3'$. As soon as $3'$ comes to a point lower than point $2'$, the operation point hops to $3'$, a longer wavelength. This is the phenomenon of mode hopping. The operating point will keep shifting to point $3''$ as I_p is further increased. The wavelength becomes shorter until the next hopping occurs, at which time the wavelength hops to a longer wavelength. The process repeats and the maximum tuning range is limited to between $2'$ and $3'$, as indicated in Fig. 14.45. This is the case when only the phase controller tuning current I_p is varied.

14.10.1.2 Tuning of Wavelength by the Bragg Reflector Tuning Current I_b Alone

As the Bragg reflector tuning current I_b is increased, both the phase and the Bragg reflector loss shift toward a shorter wavelength, as shown in Fig. 14.46.

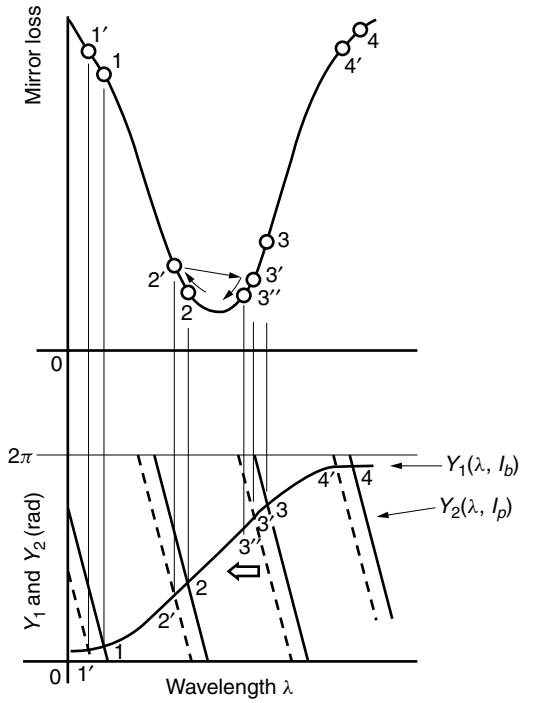


Figure 14.44 Shift of the lasing condition due to the injection of the phase control current. (After K. Kobayashi and I. Mito [27].)

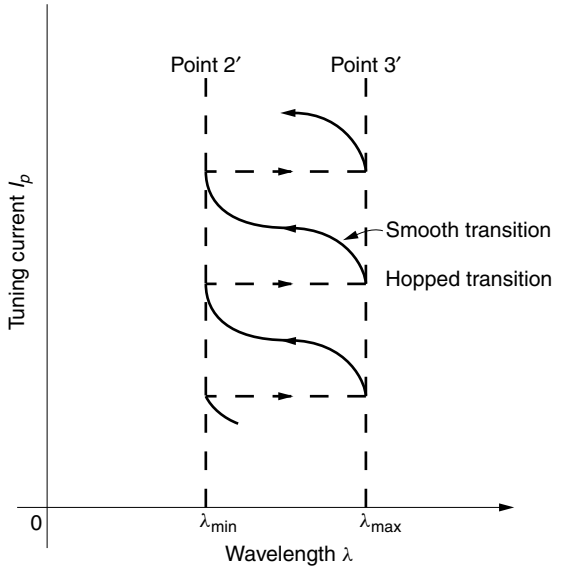


Figure 14.45 Range of wavelength tuning by I_p alone. The path of the operating point repeats and the tuning range is limited. (After K. Kobayashi and I. Mito [27].)

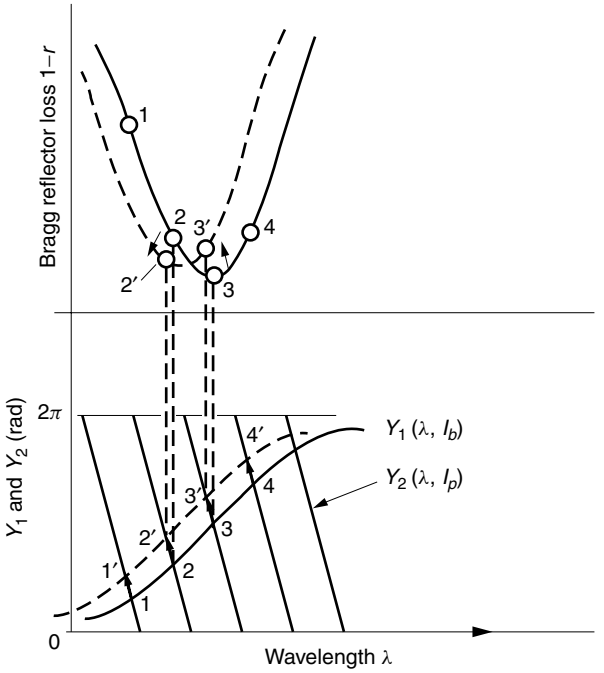


Figure 14.46 Shift of operating point with hopping as I_b is increased. (After K. Kobayashi and I. Mito [27].)

There are four intersections 1, 2, 3, and 4. Among the four intersections, 3 has the lowest Bragg reflector loss and the lasing takes place at 3 when I_b is zero.

As the tuning current I_b is increased, these intersections 1, 2, 3, and 4 move to the upper left to 1′, 2′, 3′ and 4′, as indicated in the bottom of Fig. 14.46, and the wavelength of operation becomes shorter. The corresponding points of the Bragg reflector loss are indicated by the numbers in the top graph.

Among the satisfied phase conditions, it is always with the lowest Bragg reflector loss that lasing takes place. With the top graph of Fig. 14.46, the operating point is at 3 when I_b is zero, but as I_b is increased, the Bragg reflector loss moves up to a new point 3′. At the same time, operating point 2 moves downward to 2′. As soon as 2′ becomes lower than 3, the wavelength of operation hops from point 3 to 2′ and there are no in-between operating points. If I_b is further increased, the operating point will hop to 1′.

Thus, the wavelength of operation decreases with increasing I_b but with a series of wavelength hopping as shown in Fig. 14.47, and no continuous tuning is achievable.

14.10.1.3 Continuous Wavelength Tuning by Combining I_p and I_b

By combining the phase controller and Bragg reflector tuning currents alternately, continuous wavelength tuning is possible. The phase controller Y_2 is shifted first, and then the Bragg reflector phase Y_1 is shifted to follow the shift of the Y_2.

Figure 14.48 explains this mode of operation. Let's start from the operating point 1. First, Y_2 is moved to the left by injecting I_p and the operating point moves toward point 2. Then Y_1 is moved toward the left by injecting I_b and the operating point

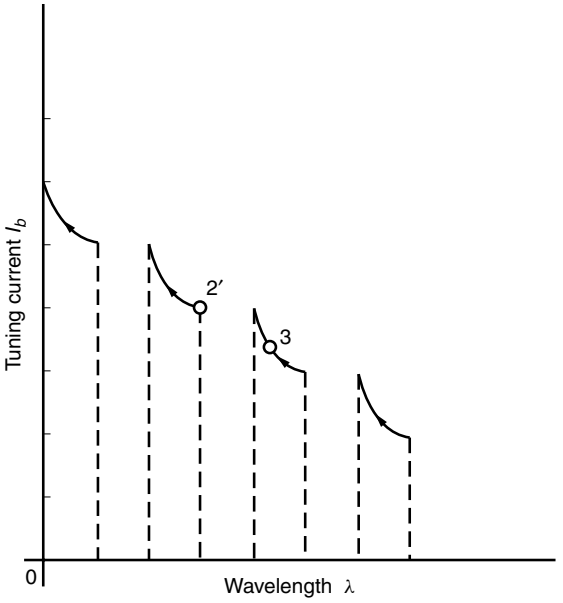


Figure 14.47 Decrease in wavelength of operation when I_b alone was increased. Note existence of hopping. (After K. Kobayashi and I. Mito [27].)

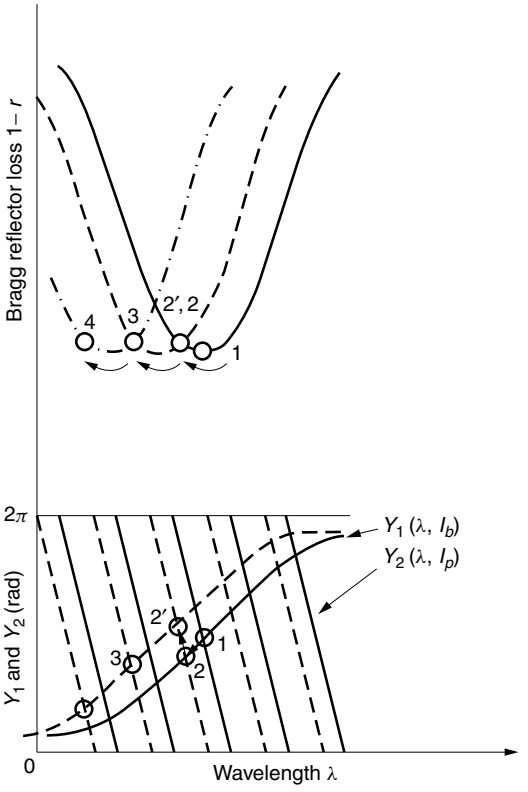


Figure 14.48 Continuously tunable laser diode. (After K. Kobayashi and I. Mito [27].)

moves by a small amount to point 2′. Next, I_p is increased and the operating point is moved further to the left to point 3. Thus, by alternately changing I_p and I_b, the operating wavelength can be shifted continuously as shown in Fig. 14.49.

The tunable range of a tunable DFB laser, however, is at most 1% of the center wavelength because it is limited by the amount of change Δn in the refractive index of the Bragg reflector. The maximum amount of change Δn in the refractive index by the plasma effect of the injection current is around 1%. The operation of the SSG laser diode, which has a much wider range than the tunable DFB laser, will be presented in the next section.

14.10.2 Superstructure Grating Laser Diode (SSG-LD)

The SSG-LD [30,31] made it possible to expand the tuning range by a factor of 10 over the tunable DFB lasers using the vernier effect between the reflection peaks of two superstructure gratings whose reflection peaks are spaced slightly differently.

The distribution of the refraction index of the superstructure grating is shown in Fig. 14.50a. It is a superstructure in the sense that substructures are repeated to form an overall structure. The substructures are repeated with a period of Λ_s. Each substructure is a grating and the period is chirped so that the reflectivity is uniform for a wide band of frequencies. The reflection center for a lower frequency is at the leftmost of each substructure $a_1, a_2, a_3, \ldots, a_n$ and that of a higher frequency is at the rightmost of each substructure $b_1, b_2, b_3, \ldots, b_n$. It should be noted that the spacing $a_{n-1} - a_n$ for the lower frequency and $b_{n-1} - b_n$ for the higher frequency

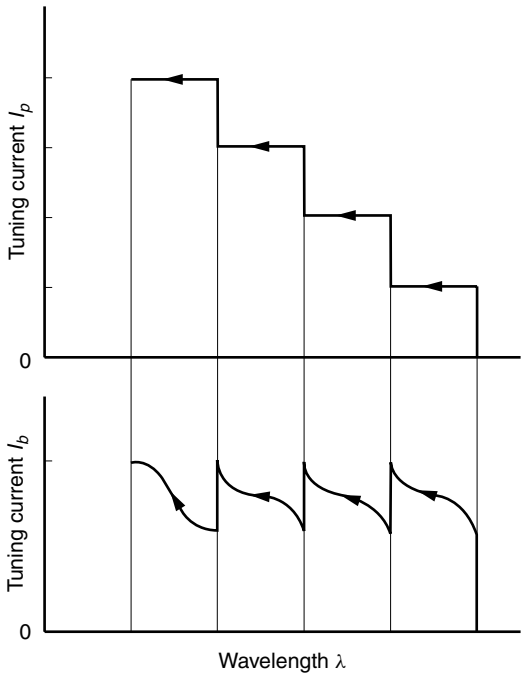


Figure 14.49 Tuning currents of I_p and I_b for continuous tuning operation. (After K. Kobayashi and I. Mito [27].)

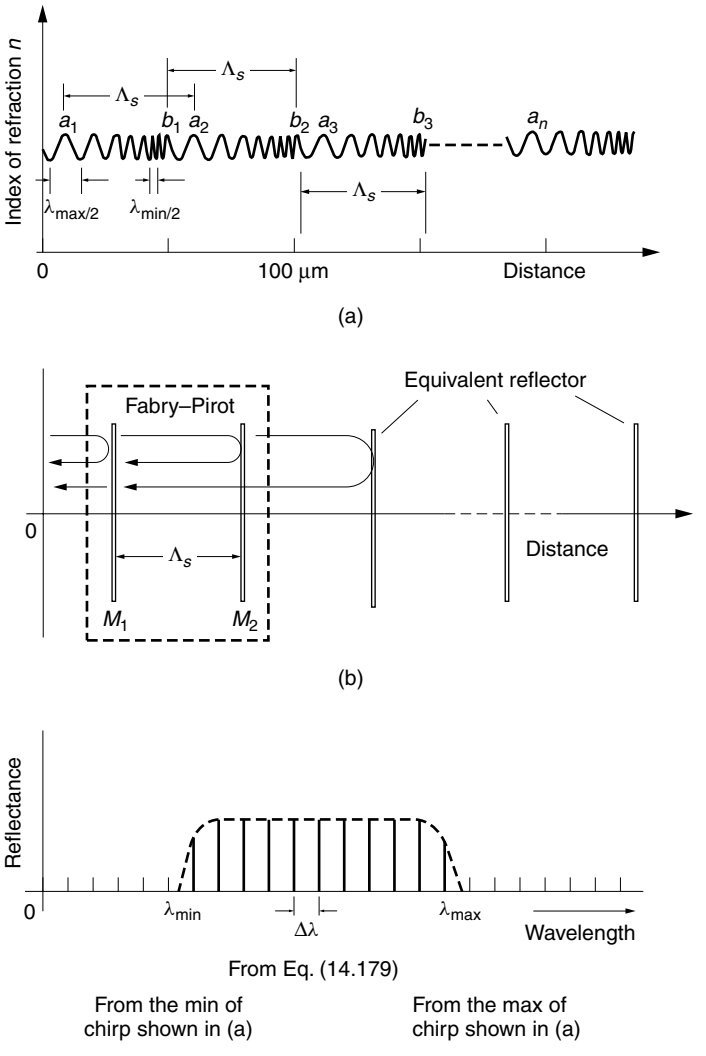


Figure 14.50 Superstructure grating laser diode. (a) Refractive index distribution of the superstructure grating. (b) Equivalence of superstructure grating to a Fabry–Pérot resonator. (c) Spectrum of reflection peaks of the superstructure grating. (After Y. Tohmori et al. [30].)

are both Λ_s. Thus, the superstructure grating can approximately be regarded as an array of nonfrequency-dependent scattering centers spaced at Λ_s, as shown in Fig. 14.50b. The reflection from such an array structure can be regarded as multiple reflections taking place in a Fabry–Pérot resonator with its reflector spacing Λ_s (see Section 3.1.1). The reflection characteristic from the superstructure is similar to that of the Fabry–Pérot resonator. The spacing $\Delta\lambda$ of the reflection peaks of such a Fabry–Pérot resonator is

$$\Delta\lambda = \frac{\lambda^2}{2n\Lambda_s} \tag{14.179}$$

where n is the refractive index of the grating region. The injection current can change the value of the refractive index, and thus the spacing $\Delta\lambda$ of the reflection peaks is controlled by the injection current. Figure 14.50c shows the pattern of the reflection peaks with respect to the wavelength. The configuration of the substructure determines the shape of the envelope of the reflection peaks (see Section 1.4.9). The minimum light wavelength is determined by $\lambda_{\min}/2$, which is the shortest period of the chirp structure shown in Fig. 14.50a, and the maximum wavelength $\lambda_{\max}$ is determined by $\lambda_{\max}/2$, which is the longest period of the chirp structure.

Next, the vernier effect will be explained. The SSG laser diode (SSG-LD) has two superstructure gratings at the front and rear ends of the diode as shown in the cutout view of the SSG-LD in Fig. 14.51. These two superstructure gratings are similar except for their period Λ_s. One of them is slightly shorter than the other. Let us say for now the period Λ_s of the front grating is slightly shorter, and hence the period $\Delta\lambda$ given by Eq. (14.179) is slightly wider than that of the rear grating. Figure 14.52 shows the reflection patterns back to back.

In order to build up the amplitude, the lightwave has to pass not just once but repeatedly through the active region. When the wavelength of the reflection peaks of the front and rear superstructure gratings coincide, the light of this wavelength is reflected at both ends. Repeated passage of the lightwave through the active region becomes possible, and lasing takes place. For instance, Fig. 14.52 shows the configuration of three wavelengths at λ_n, λ_n', and λ_n''. These are three possible matches but λ_n' and λ_n'' are outside the envelope of the reflection peaks, and their reflectivities are too small to lase. The shape of the envelope has to be designed so that λ_n' and λ_n'' are outside the envelope. As mentioned earlier, the shape of the envelope is manipulated by the shape of the chirp in the substructure shown in Fig. 14.50a.

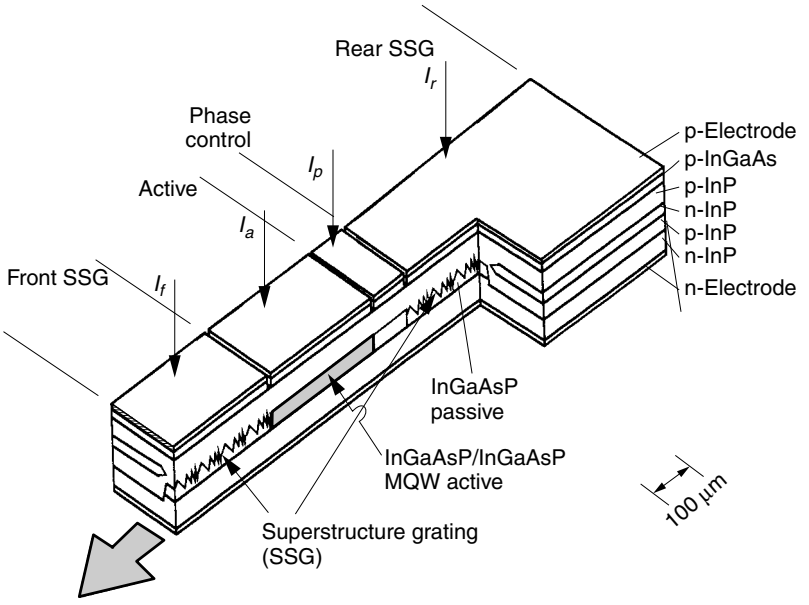


Figure 14.51 Schematic drawing of the multiple phase shift SSG-DBR laser. (After Y. Tohmori et al. [30].)

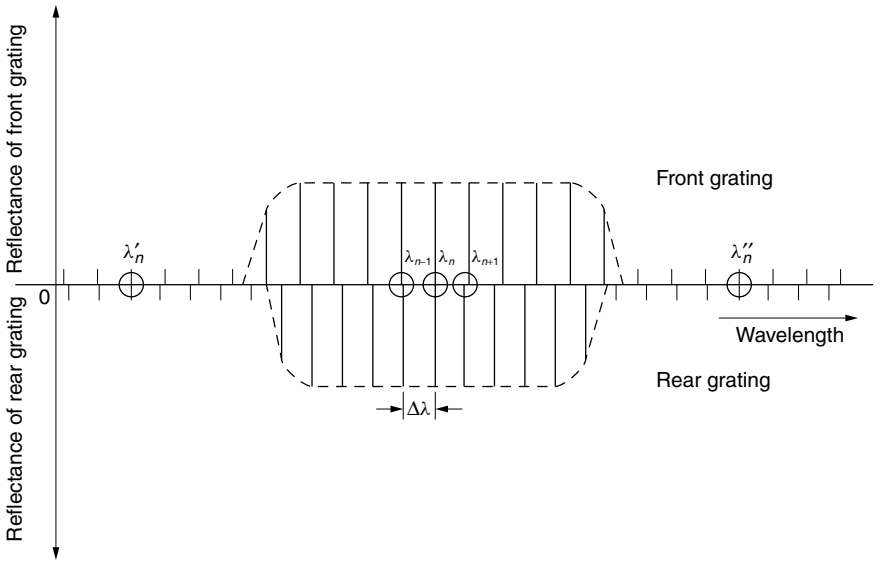


Figure 14.52 Principle of stepping the wavelength of the SSG-LD.

Next, the manner of the wavelength tuning will be explained. If an electric current is injected into the grating regions, the refractive index n is decreased, the spacing $\Delta\lambda$ of the reflection peaks expands, and each peak shifts toward a longer wavelength (toward the right). Now, if the tuning current is applied only to the rear grating, the lower spectrum in Fig. 14.52 shifts to the right while the upper spectrum is unchanged, and the lasing wavelength will shift to λ_{n+1}, where the two reflection peaks will meet. On the other hand, when the tuning current is applied only to the front grating, the lasing wavelength will hop to λ_{n-1}. If both tuning currents are applied, a finer tuning of the wavelength of oscillation is possible.

Only a small amount of shift of the reflection peaks shown in the circles in Fig. 14.52 can create a substantial wavelength shift. This resembles the action of the vernier scale where with a slight movement of the vernier scale, the position of the matched graduation of the regular and vernier scale moves over a large number of graduations.

In addition to the front and rear grating, the laser is equipped with a phase control region. This region is used for adjusting the phase so that the phase of the wave after one round trip matches that of the previous round trip to ensure that the wave constructively adds up after each round trip and the amplitude builds up.

Figure 14.53 shows how the lasing wavelength shifts by manipulating the phase control currents. The horizontal axis is the wavelength and the vertical axis is the light output with representative combinations of control currents. The wavelength of the output light shifts over 50 nm by various choices of front, rear, and phase tuning currents.

14.11 LASER DIODE ARRAY

Laser diode arrays are employed in situations where a multiplicity of independent sources are required, such as in driving a multifiber cable. In this case, the laser diodes

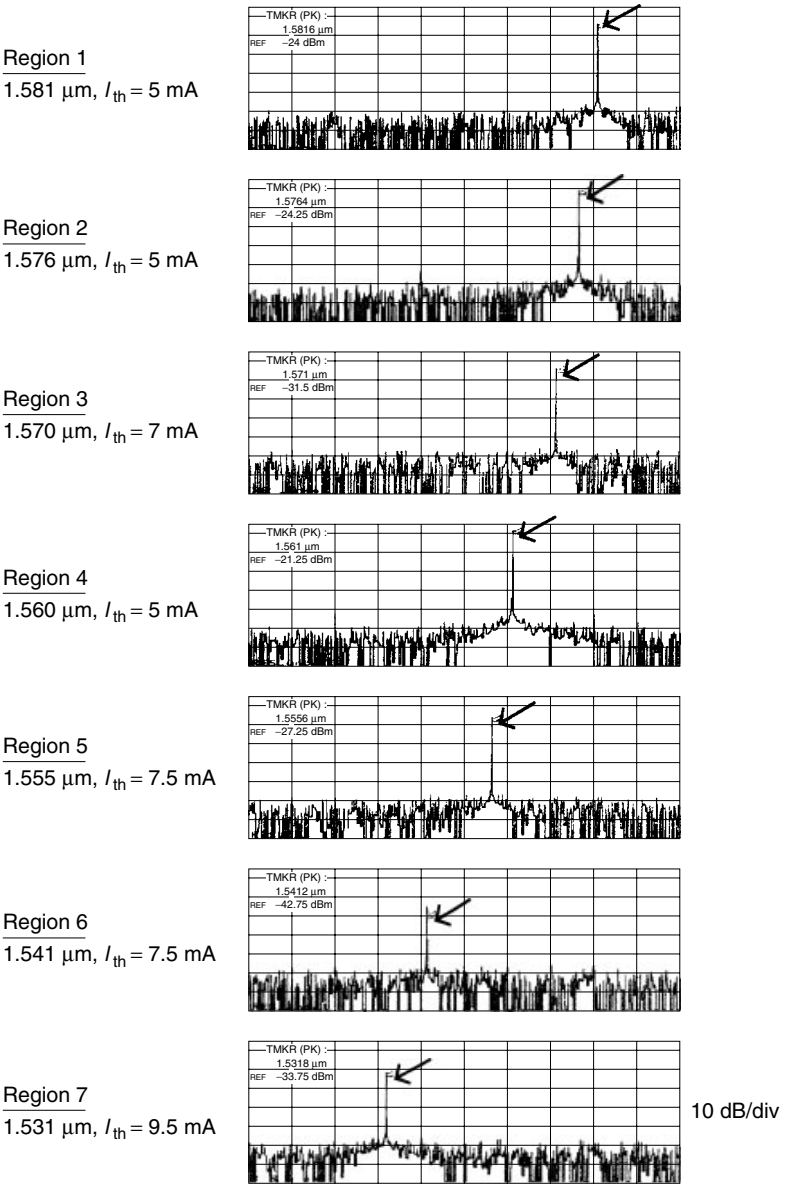


Figure 14.53 Lasing spectra of seven wavelength regions in a 50-nm tuning range. (After Y. Tohmori et al. [30].)

are connected in one-to-one correspondence, that is, one light source to one optical fiber in the cable. The surface-emitting laser diode array belongs to this category. The individual lasers are optically isolated.

Figure 14.54 shows a vertical cavity surface-emitting laser (VCSEL) array [32] used for a free-space optical interconnection. The array not only connects optical channels at an ultrafast switching speed but also overcomes the intrinsic limits of physical wiring [33]. Figure 14.54a is a schematic of a basic VCSEL array. Figure 14.54b

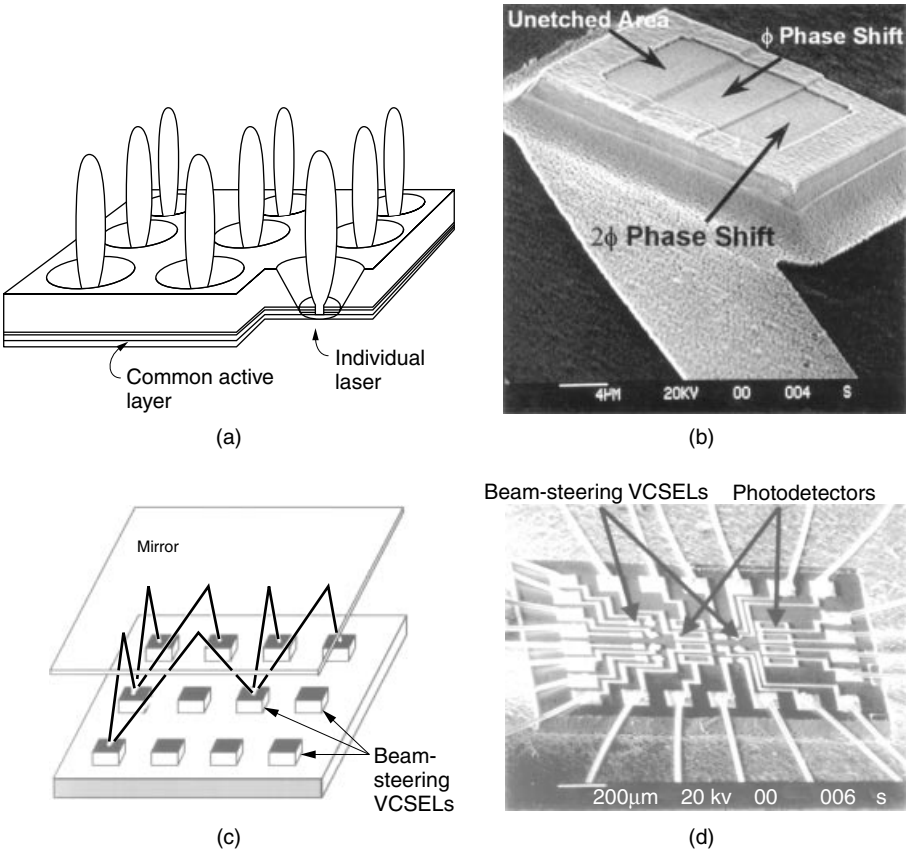


Figure 14.54 Vertical cavity surface-emitting laser (VCSEL) diode array. (a) VCSEL array. (b) SEM photograph of VCSEL array with beam-steering router. (c) Schematic diagram of the VCSEL array optical interconnect. (d) SEM photograph of a packaged 4×4 photonic network. (After L. Fan et al. [32].)

shows a beam-steering VCSEL. A three-element phase shifter film is deposited on the outer surface of the VCSEL and the direction of the light beam is electronically steered. (The principle is explained in Fig. 14.24.) Such beam-steering VCSELs are used for implementing multistage space division switching networks, as illustrated in Fig. 14.54c. Figure 14.54d is a scanning electron microscope photograph of a packaged 4×4 optical interconnect.

An alternate use of laser diode arrays is in a many-to-one correspondence, whereby many laser diode sources are fed into a single fiber or single optical component. This arrangement is used to increase the light power into the optical component. In order to effectively increase the total output from the array, the element lasers have to be excited coherently with each other. The rest of this section will be devoted to this category of coherent laser diode arrays.

Normally, the elements of the coherent laser array are put close together. If all array elements have the same amplitude and phase and are closely spaced, the outputs of the array elements approximate the distribution shown in Fig. 14.24b, and a narrow symmetric, high intensity beam can be obtained. It is important that all the array

elements have the same phase. If for instance, the left half of the array elements are in opposite phase to the right half of the elements, they form an antisymmetric distribution, and a null appears in the middle of the radiation pattern as shown in Fig. 14.24d. The outputs of the array elements should discretely approximate either the pattern in Fig. 14.24b or c.

Figure 14.55 shows a Y-branch laser diode array [34] that forces the output from each array element to be identical with the adjacent elements. The geometry of the Y-branch laser array is shown in Fig. 14.55a. For the purpose of illustrating the principle of operation of the array, only one period of the array is shown separately in Fig. 14.55b. Let us assume first that the amplitude E_1 of the bottom left arm is different from E_3 of the bottom right arm. Amplitudes are arbitrary, but phases are assumed either in phase or 180° out of phase. The general case is left for Problem 14.4. E_1 and E_3 can be decomposed into a symmetric (in-phase) component E_s, and an antisymmetric (180° out-of-phase) component E_a, as shown in Fig. 14.55b. The E_s components in both arms of the same amplitude oscillate in phase, whereas the E_a components in both arms of the same amplitude oscillate 180° out of phase. Using the law of superposition, E_1 and E_3 are expressed in terms of symmetric and antisymmetric components as

$$\begin{aligned} E_1 &= E_s + E_a \\ E_2 &= E_s - E_a \end{aligned} \tag{14.180}$$

Then, E_s and E_a are

$$\begin{aligned} E_s &= \frac{E_1 + E_2}{2} \\ E_a &= \frac{E_1 - E_2}{2} \end{aligned} \tag{14.181}$$

Now consider the inputs into the upper arms. The E_s components add constructively and the input to the upper arms is $2E_s$ whereas the E_a components add destructively and the input to the upper arm is zero, as indicated in Fig. 14.55b. Picture Fig. 14.55b with laser cavity mirrors located at the top of branch E_2 and at the very bottom of branches E_1 and E_3. The input into the upper arm is amplified and then reflected downward into the lower arms, and, depending on the reflectivity of the mirrors, this up-and-down pattern may be repeated several times before the light radiates out from either the top

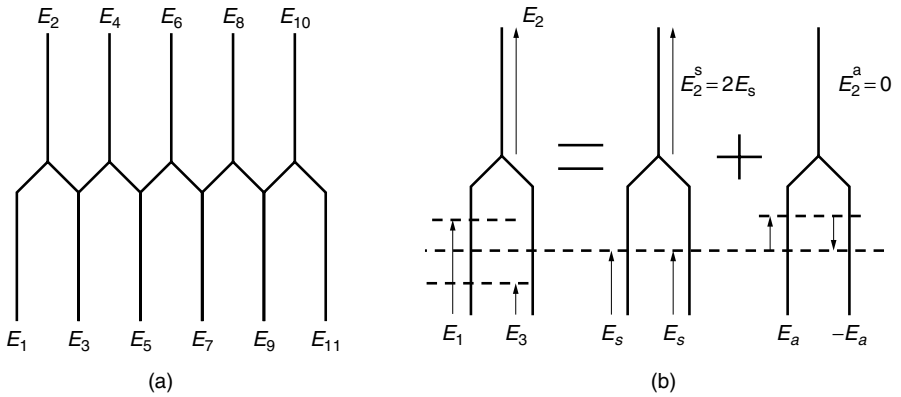


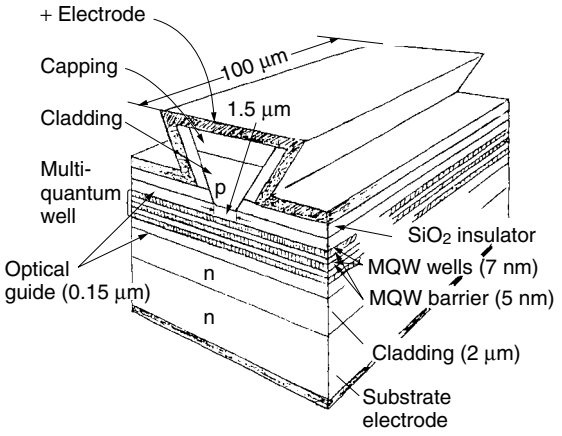
Figure 14.55 Y-branch laser array. (a) Branch laser array. (b) Explanation of the operation.

or bottom apertures. The symmetric components $2E_s$ are much more amplified than the antisymmetric components whose input to the upper arm is null. This is true with all neighboring arms in the Y-branch array shown in Fig. 14.55a. Thus, $E_2, E_4, \ldots, E_{2n}$ are excited in phase, and so are $E_1, E_3, E_5, \ldots, E_{2n-1}$. The aperture pattern of the array discretely resembles that in Fig. 14.24b if the spacing between the elements is small. Such a laser array emits a narrow high-intensity beam.

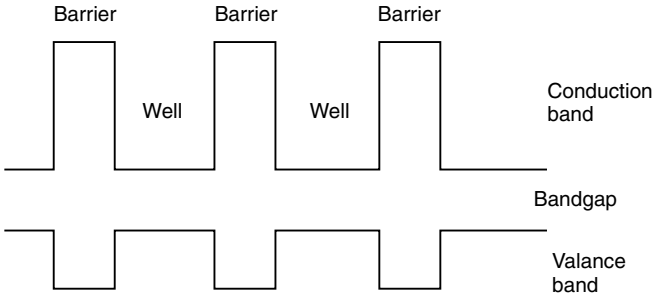
14.12 MULTI-QUANTUM-WELL LASERS

By alternately stacking thin layers of wider and narrower bandgaps as shown in Fig. 14.56a, a series of potential wells such as shown in Fig. 14.56b is formed that is called a *multi-quantum well* (MQW) [35]. As the dimensions of the potential wells are reduced to the order of 10 nm, the movement of the electrons is restricted inside the potential well, and quantum effects become prominent.

The laser gain curves made up of Eqs. (14.2), (14.5), and (14.19) are strongly governed by the shapes of the density of the states $g_c(E_2)$ and $g_v(E_1)$. The quantum well effect makes it possible to tailor the shapes of the functions of $g_c(E_2)$ and $g_v(E_1)$



(a)



(b)

Figure 14.56 Multi-quantum well. (a) Structure of multi-quantum-well laser ($\lambda = 0.835$ μm). (b) Potential well of MQW.

so that the gain of the laser is peaked at a specified frequency. A semiconductor laser using the quantum well effect has such good features as:

- Narrow frequency band gain curve
- Lower threshold current
- Less temperature dependence
- Frequency of emission that can be designed by the dimension of the wells

Let's start with an analysis of the energy states in a bulk semiconductor before going into the quantum well laser.

14.12.1 Energy States in a Bulk Semiconductor

The kinetic energy of an electron with total energy E and potential energy V is

$$\tfrac{1}{2}m_e^* v^2 = E - V \tag{14.182}$$

where v is the velocity of the particle and m_e^* is the effective mass of an electron inside the semiconductor crystal. The momentum p of such a particle is

$$p = m_e^* v \tag{14.183}$$

Combining Eqs. (14.182) and (14.183),

$$p = \sqrt{2m_e^*(E - V)} \tag{14.184}$$

It was de Broglie who postulated that any particle of matter that has momentum should also possess a wave nature, called a *matter wave*. The wavelength λ of the matter wave is given by

$$p = \frac{h}{\lambda} \tag{14.185}$$

From Eqs. (14.184) and (14.185), the wavelength is expressed as

$$\lambda = \frac{h}{\sqrt{2m_e^*(E - V)}} \tag{14.186}$$

If Eq. (14.186) is inserted into the wave equation,

$$\nabla^2\psi + \left(\frac{2\pi}{\lambda}\right)^2 \psi = 0 \tag{14.187}$$

then the wave equation for the matter wave is obtained as

$$\nabla^2\psi + \frac{2m_e^*(E - V)}{\hbar^2}\psi = 0 \tag{14.188}$$

where

$$\hbar = \frac{h}{2\pi} \tag{14.189}$$

Equation (14.188) is called *Schrödinger's equation*.

The wavefunction that satisfies Schrödinger's equation will be found [6]. In the bulk case, the analysis is made for the potential well constant throughout the well as shown

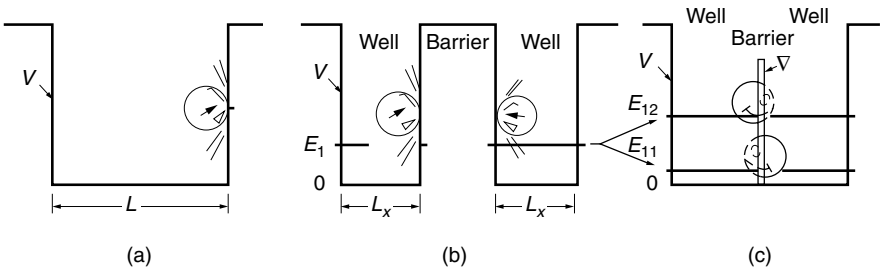


Figure 14.57 Comparison of the quantum-well structure among (a) bulk well, (b) weakly coupled double quantum well, and (c) strongly coupled double quantum well.

in Fig. 14.57a. V being constant, Eq. (14.188) becomes a linear, homogeneous, second order, partial differential equation and its general solution is

$$\psi = A \sin(k_x x + \phi_x) \sin(k_y y + \phi_y) \sin(k_z z + \phi_z) \tag{14.190}$$

Inserting Eq. (14.190) into (14.188), the values of k_x, k_y, and k_z are found as

$$\frac{2m_e^* E}{\hbar^2} = k_x^2 + k_y^2 + k_z^2 \tag{14.191}$$

where

$$V = 0 \tag{14.192}$$

is assumed inside the well.

For simplicity, let us say the electron is confined inside cubic potential walls with size L^3 as shown in Fig. 14.58. It is further assumed that the potential walls are so high that $\psi = 0$ on and outside the walls. First, the boundary condition in only the x direction is considered. In order that $\psi = 0$ at both $x = 0$ and $x = L$,

$$\begin{aligned} \phi_x &= 0 \\ k_x L &= n_x \pi \end{aligned} \tag{14.193}$$

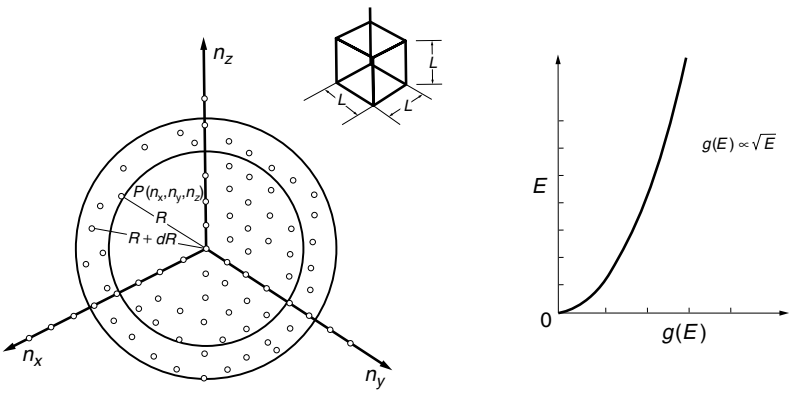


Figure 14.58 Modes in a bulk semiconductor.

where n_x is a positive integer called the mode number. Only positive integers are used for designating the mode number. Similarly,

$$\begin{aligned} k_x &= \frac{\pi}{L} n_x \\ k_y &= \frac{\pi}{L} n_y \\ k_z &= \frac{\pi}{L} n_z \end{aligned} \tag{14.194}$$

Thus, the solution ψ of Schrödinger's equation is

$$\psi = \sin k_x x \cdot \sin k_y y \cdot \sin k_z z \tag{14.195}$$

Let us consider (n_x, n_y, n_z) coordinates such as shown in Fig. 14.58, where n_x, n_y, n_z are positive integer points. Any one dot that is made up of a combination of n_x, n_y, n_z represents a mode of the electron motion inside the cubic potential well.

Inserting Eq. (14.194) into (14.191) gives

$$\frac{2m_e^* E}{\hbar^2} \left(\frac{L}{\pi}\right)^2 = n_x^2 + n_y^2 + n_z^2 \tag{14.196}$$

The energy level E that belongs to each mode can be calculated from Eq. (14.196).

The right hand side of Eq. (14.196) is the square of the distance R to the origin. The states having energy E lie on the sphere with radius R such that

$$R = \frac{\sqrt{2m_e^* E}}{\hbar} \frac{L}{\pi} \tag{14.197}$$

or

$$E = \frac{\hbar^2}{2m_e^*} \left(\frac{\pi}{L}\right)^2 R^2 \tag{14.198}$$

Because n_x, n_y, n_z are restricted to be integers, the associated energy levels E are also discrete, but the spacing between the adjacent energy levels is very minute (compared to the 0.8 eV energy bandgap of a $\lambda = 1.55$ μm semiconductor laser.) because when $L = 1$ mm is inserted into Eq. (14.196)

$$E = 9.2 \times 10^{-12} (n_x^2 + n_y^2 + n_z^2) \text{ eV} \tag{14.199}$$

This is the reason why the energy levels in a bulk semiconductor are practically continuous.

We are finally ready to calculate the density of states $g(E)$ of an electron in a unit volume of the bulk semiconductor. The density of states is defined such that the number of states between E and $E + dE$ is represented by $g(E)\,dE$. $g(E)\,dE$ will be found by calculating the difference between the total number of states whose associated energy is smaller than E and the total number with energy smaller than $E + dE$. Referring to Fig. 14.58, n_x, n_y, n_z are all positive integers and only the lattice points in one-eighth

of the sphere are counted. Including the orientation of the spin of the electron, the total number M of states whose associated energy is smaller than E is

$$M = 2\left(\frac{1}{8}\right)\left(\frac{4}{3}\right)\pi R^3 \tag{14.200}$$

The number of states between E and $E + dE$ is

$$g(E)\,dE = \frac{1}{L^3}\frac{dM}{dE}dE \tag{14.201}$$

L^{-3} was included to define the quantity per unit volume. Using Eq. (14.198) in (14.200) and inserting into Eq. (14.201) finally gives

$$g(E) = \frac{\sqrt{2}m_e^{*3/2}}{\hbar^3\pi^2}\sqrt{E} \tag{14.202}$$

Equation (14.202) is plotted as a function of E in Fig. 14.58. The function $g(E)$ is smoothly increasing starting with $E = 0$. There are no special peaks as are found with the quantum well effect, as will be shown next.

The result can easily be extended to the proof of Eq. (14.3). If an electron inside the semiconductor has an effective mass and total energy E_2, and the potential energy is taken as $V = E_c$, then instead of Eq. (14.192), the density $g_c(E_2)$ of the states per unit volume in the conduction band is

$$g_c(E_2) = \frac{\sqrt{2}m_e^{*3/2}}{\hbar^3\pi^2}\sqrt{E_2 - E_c} \tag{14.203}$$

which is Eq. (14.3).

14.12.2 Energy States in a Quantum Well

Figures 14.57b and 14.57c show the MQW of the simplest kind. Two wells are partitioned by a potential barrier. When the potential barrier is set too high and wide, as shown in Fig. 14.57b, an electron in one well cannot tunnel across the potential barrier to the adjacent well, and the electron remains in the same well. The motion of such an electron can be treated as if the electron is inside a single well. If there are N wells, the problem can be treated as if there are N separate wells. This case is classified as the weakly coupled case. In the weakly coupled case, the analysis for the single electron in one potential well applies to every well and the analysis is simple.

However, when the barrier becomes thin and narrow, as in Fig. 14.57c, the tunneling effect becomes significant and the electrons begin to sneak into each other's wells. This is similar to what happens to photons in an optical directional coupler, where photons in one of the optical guides sneak into the other guide by means of the evanescent wave if the spacing is narrow. This case is classified as the strongly coupled case. The analysis of strong coupling is much more complicated than that of weak coupling, and for this reason only the weakly coupled case will be presented here.

The case when an electron is confined in one of the two potential wells shown in Fig. 14.57b will be treated. The barrier wall is assumed to be so high and thick that the electron inside the well cannot penetrate into the wall and the wavefunction is zero on and inside the potential barrier. The width L_x of the well in the x direction

is significantly narrower than the other two dimensions, which are assumed L. The boundary conditions in Eq. (14.194) have to be modified. The wavefunction in the x direction is

$$X(x) = A \sin(k_x x + \phi_x) \tag{14.204}$$

In the quantum well, $X(x)$ has to vanish at both walls at $x = 0$ and $x = L_x$. This means that $X(0) = 0$, $\phi_x = 0$, and

$$X(x) = A \sin k_x x \tag{14.205}$$

Equation (14.205) has to satisfy yet another boundary condition:

$$X(L_x) = 0 \tag{14.206}$$

hence,

$$k_x L_x = n_x \pi \tag{14.207}$$

where n_x is a positive integer

$$n_x = 1, 2, 3, 4, \ldots \tag{14.208}$$

Thus, Eqs. (14.194) and (14.196) have to be modified as

$$\begin{aligned} k_x &= \frac{\pi}{L_x} n_x \\ k_y &= \frac{\pi}{L} n_y \\ k_z &= \frac{\pi}{L} n_z \end{aligned} \tag{14.209}$$

$$\frac{2m_e^* E}{\hbar^2} \left(\frac{L}{\pi}\right)^2 = n_x'^2 + n_y^2 + n_z^2 \tag{14.210}$$

with

$$n_x' = \frac{L}{L_x} n_x \tag{14.211}$$

Figure 14.59 represents the lattice points in n_x', n_y, n_z coordinates. From Eq. (14.211), as n_x changes with unit increment, n_x' changes with a large increment because L/L_x is a large number like 10^5. As a result, the lattice points look more like sheets. There are no lattice points in the region

$$0 < n_x' = \frac{L}{L_x} \tag{14.212}$$

and the first sheet appears in the plane of

$$n_x' = \frac{L}{L_x} \tag{14.213}$$

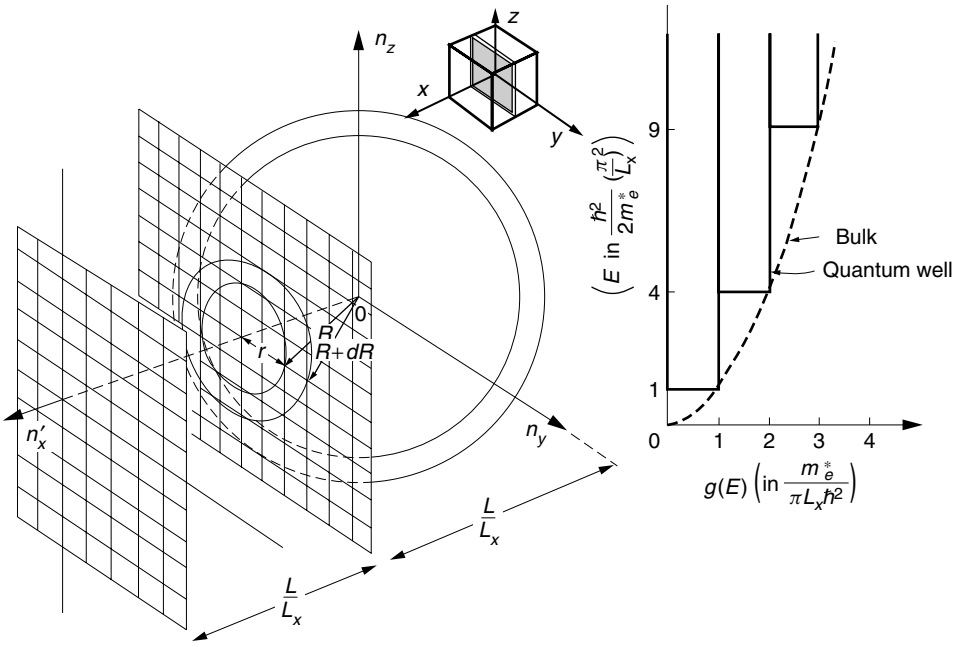


Figure 14.59 Modes in a quantum well.

The associated energy level of the mode where the sphere is tangent to the sheet is obtained from Eq. (14.210) with Eq. (14.213) and $n_y = n_z = 0$ as

$$E_1 = \frac{\hbar^2}{2m_e^*}\left(\frac{\pi}{L_x}\right)^2 \tag{14.214}$$

which is indeed the first quantized energy level of the electron in the potential well. The next sheet appears at twice the distance, and the subsequent n_xth sheet appears at

$$n'_x = \frac{L}{L_x}n_x \tag{14.215}$$

The points for a given energy E are located on the surface of the sphere with radius R given by Eq. (14.198). With the configuration shown in Fig. 14.59, first consider the case when the sphere is cut by only one sheet and R is larger than L/L_x but smaller than $2L/L_x$, or from Eq. (14.198),

$$\frac{\hbar^2}{2m_e^*}\left(\frac{\pi}{L_x}\right)^2 < E < \frac{\hbar^2}{2m_e^*}\left(2\frac{\pi}{L_x}\right)^2 \tag{14.216}$$

The energy represented by the points in the first sheet with $n_x = 1$ is

$$E = \frac{\hbar^2}{2m_e^*}\left(\frac{\pi}{L}\right)^2\left[\left(\frac{L}{L_x}\right)^2 + r^2\right] \tag{14.217}$$

where

$$r^2 = n_y^2 + n_z^2$$

In order to find the density of states, the number of modes M with energy less than E has to be found. Recall that n_x and n_y are positive integers and only the lattice points in one-quarter of the circle are counted, and there are two spin states for each point.

$$M = \left(\frac{1}{4}\right)(2)\pi r^2 \tag{14.218}$$

The rate of increase in M with respect to r is

$$\frac{dM}{dr} = \pi r \tag{14.219}$$

On the other hand, the rate of increase in E with respect to r is

$$\frac{dE}{dr} = \frac{\hbar^2}{m_e^*}\left(\frac{\pi}{L}\right)^2 r \tag{14.220}$$

Thus, the rate of increase in M with respect to E is

$$\frac{dM}{dE} = \frac{dM}{dr}\frac{dr}{dE} = \frac{m_e^* L^2}{\pi\hbar^2} \tag{14.221}$$

Equation (14.221) has to be divided by the volume $L_x L^2$ of the potential well in order to obtain the density of states per unit volume

$$g_1(E) = \frac{m_e^*}{\pi L_x \hbar^2} \tag{14.222}$$

Note that $g_1(E)$ is independent of r, R, and E.

Next, the case when the sphere intercepts two sheets is considered. The rate of increase in M in the second sheet is obtained with $n_x = 2$. Only the first term in Eq. (14.210) has the factor n_x, but this n_x soon disappears when dM/dr and dE/dr are calculated. The density of states for $n_x = 2$ is the same as before, and is Eq. (14.222), but now the sphere is cut by two sheets, and the rate of increase in the number of states for each sheet is independent of R. The total increase in states is simply the addition of the two contributions, and

$$g_2(E) = 2\frac{m_e^*}{\pi L_x \hbar^2} \tag{14.223}$$

Finally, the density of states at an arbitrary energy level is considered. Every time a new sheet is cut by the sphere, the density of the states increases by $m_e^*/\pi L_x\hbar^2$. The density of states per unit volume inside the potential well is therefore

$$g(E) = \frac{m_e^*}{\pi L_x \hbar^2}\sum_{n=1}^{N} H(E - E_{x_n}) \tag{14.224}$$

where

$$E_{x_n} = \frac{\hbar^2}{2m_e^*}\left(n_{x_n}\frac{\pi}{L_x}\right)^2 \tag{14.225}$$

and $H(x)$ is *Heavyside's step function*, defined as

$$H(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & x = 0 \\ 1 & x > 0 \end{cases} \tag{14.226}$$

The curve of $g(E_x)$ looks like a staircase function. As compared to the smooth curve of the bulk semiconductor laser, it steps up by $m_e^*/\pi L_x \hbar^2$ every time E_x exceeds E_{n_x}, as shown in Fig. 14.59.

Next, the influence of this staircase like function on the gain curve is explained.

14.12.3 Gain Curves of the MQW Laser

The gain curve has already been calculated for a bulk p-n junction laser as described in Section 14.2.1. A similar approach will be used to calculate the gain curve of the active layer of the MQW laser. Using the Fermi–Dirac distribution function, the number of states that are occupied in the energy levels between E_2 and $E_2 + dE_2$ in the conduction band is given by

$$N = \int g_c(E_2) f_c(E_2)\, dE_2 \tag{14.227}$$

and that of holes in the valence band is

$$P = \int g_v(E_1)[1 - f_v(E_1)]\, dE_1 \tag{14.228}$$

Figures 14.60a–c show the operation $g_c(E_2)f_c(E_2)$ and $g_v(E_1)[1 - f_v(E_1)]$. The width of $g_c(E_2)f_c(E_2)$ for a MQW laser is narrower in the E axis dimension than that shown by the dashed line of a bulk laser. This is the very reason why the MQW laser has a more pointed and higher peak gain in a narrower frequency band than a bulk laser does.

A narrower curve means more efficient use of the injection current and hence the reduction of N_α. The reduction of N_α leads to a variety of advantages such as a higher gain in Eq. (14.26), a lower J_{th} in Eq. (14.49), a larger power in Eq. (14.58), a shorter turn-on time in Eq. (14.61), and a higher modulation frequency ω_r in Eq. (14.89).

In the following, an outline for calculating the gain will be presented. The gain curve of the MQW laser will be obtained by calculating the quasi-Fermi level for a given carrier density. Then the calculated quasi-Fermi levels are used to calculate the gain in Eq. (14.25). The carrier densities need to be calculated first before calculating the Fermi level. From Eq. (14.2), the number N of electrons in the entire conduction band is

$$\begin{aligned} N &= \int g_c(E_2) f_c(E_2)\, dE_2 \\ &= \frac{m_e^*}{\pi L_x \hbar^2} \sum_{n=1}^{\infty} n \int_{E_{x_n}}^{E_{x_{(n+1)}}} \frac{dE_2}{e^{(E_2 - F_c)/kT} + 1} \end{aligned} \tag{14.229}$$

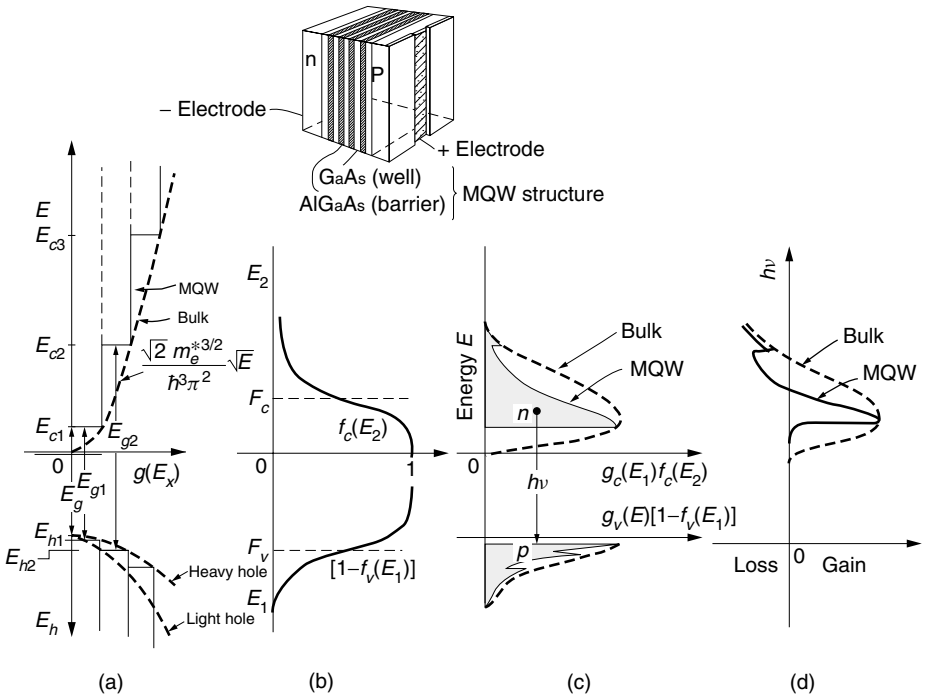


Figure 14.60 Composite diagram to explain the operation of a MQW laser. (a) Density $g(E)$ of energy states. (b) Fermi levels $f(E_2)$ and $[1 - f_v(E_1)]$. (c) Population density $g_c(E_2)f_c(E_2)$ and $g_v(E_1)[1 - f_v(E_1)]$. (d) Gain of the MQW laser.

Using the relationship

$$\int \frac{dx}{e^x + 1} = -\ln(e^{-x} + 1) \tag{14.230}$$

the integral of Eq. (14.229) is calculated as

$$\begin{aligned} N = \frac{m_e^* kT}{\pi L_x \hbar^2} \Big\{ & 1 \left[\ln\left(e^{-(E_{c1}-F_c)/kT} + 1\right) - \ln\left(e^{-(E_{c2}-F_c)/kT} + 1\right) \right] \\ & + 2 \left[\ln\left(e^{-(E_{c2}-F_c)/kT} + 1\right) - \ln\left(e^{-(E_{c3}-F_c)/kT} + 1\right) \right] \\ & + 3 \left[\ln\left(e^{-(E_{c3}-F_c)/kT} + 1\right) - \ln\left(e^{-(E_{c4}-F_c)/kT} + 1\right) \right] \\ & + \cdots \Big\} \end{aligned} \tag{14.231}$$

Aside from the very first square bracket in Eq. (14.231), the first terms in all the subsequent square brackets are reduced by an amount that corresponds to the second term in the previous square brackets, and Eq. (14.231) becomes

$$N = \frac{m_e^*}{\pi L_x \hbar^2} kT \sum_{n=1}^{\infty} \ln\left(1 + e^{(F_c - E_{cn})/kT}\right) \tag{14.232}$$

Similarly, the density of holes in the valence band is

$$P = \frac{m_h^*}{\pi L_x \hbar^2} kT \sum_{n=1}^{\infty} \ln\left(1 + e^{(E_{hn} - F_v)/kT}\right) \tag{14.233}$$

If the assumption is made that only the $n = 1$ level is filled with carriers, the quasi-Fermi levels of the MQW as a function of the carrier concentration are

$$F_c = E_{c1} + kT \ln\left(e^{N/N_c^Q} - 1\right) \tag{14.234}$$

$$F_v = E_{h1} - kT \ln\left(e^{P/N_v^Q} - 1\right) \tag{14.235}$$

where

$$\begin{aligned} N_c^Q &= \frac{m_e^* kT}{\pi L_x \hbar^2} \\ N_v^Q &= \frac{m_h^* kT}{\pi L_x \hbar^2} \end{aligned} \tag{14.236}$$

The quasi-Fermi levels together with Eqs. (14.2)–(14.4) and (14.5)–(14.7) are inserted into Eq. (14.19) to calculate the gain curve. The calculated gain curves for a $Ga_{0.47}In_{0.53}As/InP$ MQW laser are shown in Fig. 14.61 [36].

General observations concerning the calculated results show that, compared to the bulk p-n junction laser of Fig. 14.5 represented by the dashed line in Fig. 14.61, the MQW laser has the following characteristics:

- A narrower, more pointed gain curve
- A higher peak gain
- A shift of the center toward higher frequency from that of the bulk

The narrower pointed curve contributes to the stability of the oscillation. As mentioned earlier, the higher peak gain means a lowering of the MQW laser threshold current. The dependence of the shift of the center frequency on L_x means that the frequency of the MQW laser can be designed by adjusting L_x, as shown in Example 14.7.

14.12.4 Structure and Characteristics of a MQW Laser

The structure and characteristics of the MQW laser will be explained by taking a specific example of a MQW laser of the ridge type shown in Fig. 14.56a. There are three quantum wells 7-nm thick with 5-nm barriers. The quantum wells are sandwiched by 0.15-μm-thick optical guides made of $Al_{0.28}Ga_{0.72}As$ whose refractive index is $n = 3.50$. The optical guide/quantum wells/optical guide structure is further sandwiched by 2-μm-thick cladding layers of $Al_{0.8}Ga_{0.2}As$ whose refractive index is $n = 3.27$, which is lower than that of the guiding layer. The light is confined in the optical guide centered by the multi-quantum wells.

The quantum wells being thin, the confinement factor Γ (see Section 13.7) is 3.5%. Let n_w denote the number of quantum wells. If n_w is increased in order to improve the overall confinement factor, the injection carrier density $N = \tau J/2etn_w$ decreases with an increase in the total thickness tn_w, where t is the thickness of the potential well and J is limited by the heat dissipation. There is an optimum number of quantum wells.

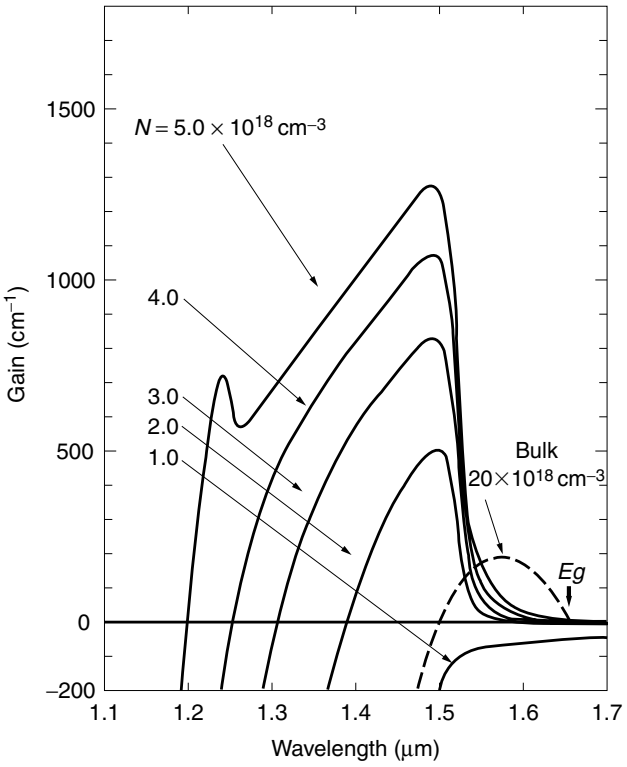


Figure 14.61 Gain spectrum of a $Ga_{0.47}In_{0.53}As/InP$ MQW laser with $L_x = 100$ Å. (After M. Asada, A. Kameyama, and Y. Suematsu [36].)

A capping layer that adheres to both metal and semiconductor materials is deposited on top of the upper cladding layer. These two layers are then etched to form a mesa structure (steep hill structure). A SiO_2 insulating layer is deposited except in the region of the capping layer. Then the positive electrode is deposited. The injection current is constricted in the narrow neck region at the bottom of the ridge.

Finally, both front and rear facets of the laser are coated with high reflection coatings with $R_f = 70\%$ and $R_r = 86\%$, respectively. The total length of the laser cavity is 100 μm.

The operational characteristics of this specific MQW laser are summarized as follows. The threshold current is 3.98 mA and is less than one-tenth the threshold current of a bulk double heterojunction laser. As mentioned earlier, this low threshold current means:

1. Biasing is practically unnecessary.
2. The relaxation oscillation decays fast and a high bit rate is possible.
3. The turn-on delay time is short.
4. The temperature dependence is less.

Expanding on the last point, the characteristic temperatures are large for this MQW laser ($T_0 = 154$ K and $T_0' = 352$ K). The output power change is ±0.5 dB for the temperature range $25 \pm 25\,^{\circ}C$.

Example 14.7 $Ga_{0.47}In_{0.53}$ As grown on an InP substrate was used as a material in the potential well of a MQW laser. The energy gap of the semiconductor is $E_g = 0.75$ eV. The potential barrier was made thick enough to assume the weakly coupled case. The width of the well is $L_x = 10$ nm.

The effective masses of $Ga_xIn_{1-x}As_yP_{1-y}$ grown on an InP substrate ($y = 2.2x$) are given by

$$\begin{aligned} m_e^* &= (0.080 - 0.039y)m_0 \\ m_h^* &= (0.4 + 0.05y)m_0 \\ m_0 &= \text{ electron mass } = 0.9109 \times 10^{-30} \text{ kg} \end{aligned} \tag{14.237}$$

(a) Calculate the values and fill in the blanks in the graph in Fig. 14.62.

(b) Design the well width L_x of the MQW laser of the same material so that the wavelength of the emitted light becomes 1.55 μm.

Solution

(a) The mole fractions of the given semiconductor are

$$x = 0.47 \quad \text{and} \quad y = 1.0$$

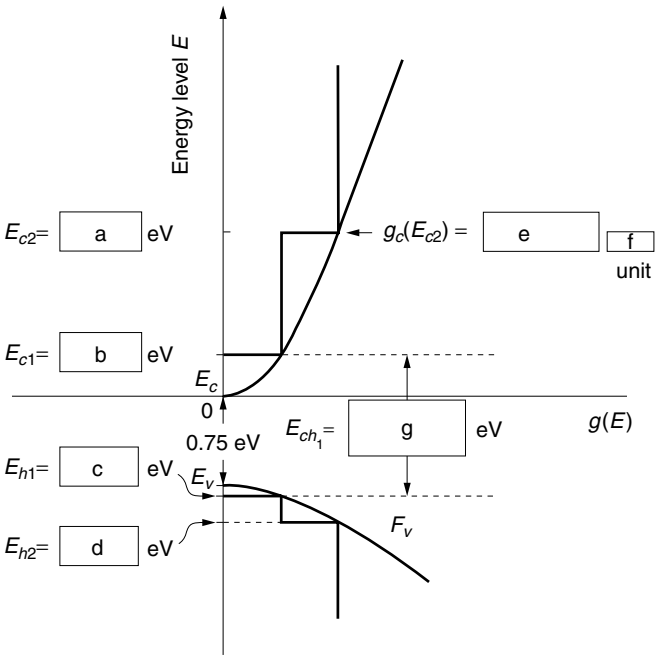


Figure 14.62 Density $g(E)$ of states versus energy level for a $Ga_{0.47}In_{0.53}As$ MQW laser. The quantities (a)–(g) are to be calculated.

Thus, the effective masses are

$$m_e^* = 0.041 m_0$$
$$m_h^* = 0.45 m_0$$

The lowest quantized energy state in the conduction band E_{c1} is obtained using Eq. (14.214):

$$E_{c1} = \frac{\hbar^2}{2m_e^*}\left(\frac{\pi}{L_x}\right)^2 \tag{14.238}$$
$$E_{c1} = \frac{(1.055 \times 10^{-34})^2}{2(0.041)(0.9109 \times 10^{-30})}\left(\frac{\pi}{10^{-8}}\right)^2 \frac{1}{1.6 \times 10^{-19}}$$

The last factor is to convert joules into electron volts:

$$E_{c1} = 0.0919 \text{ eV}$$

The second lowest quantized energy state is, from Eq. (14.225),

$$E_{c2} = 4E_{c1} = 4(0.0919) = 0.368 \text{ eV}$$

The quantized energy states E_{h1} in the valence band can be found from the ratio of the effective masses of m_h^* and m_e^*

$$E_{h1} = 0.0919\frac{0.041}{0.45} = 0.00837 \text{ eV}$$
$$E_{h2} = 0.368\frac{0.041}{0.45} = 0.0335 \text{ eV}$$

Note that E_{h1} and E_{h2} are measured downward from the band edge of the valence band.

Next, the density of states $g_1(E)$ of the electrons per unit energy per unit volume is, from Eq. (14.222),

$$g_c(E_{c2}) = 2g_1(E) = \frac{2m_e^*}{\pi L_x \hbar^2}$$
$$= \frac{(2)(0.041)(0.9109 \times 10^{-30})}{\pi(10^{-8})(1.055 \times 10^{-34})^2}$$
$$= 2.14 \times 10^{44}(\text{J} \cdot \text{m}^3)^{-1}$$

This value is converted into the unit of $(\text{eV} \cdot \text{cm}^3)^{-1}$

$$[2.14 \times 10^{44}(\text{J} \cdot \text{m}^3)^{-1}](1.6 \times 10^{-19})(10^{-6}) = 3.42 \times 10^{19}(\text{eV} \cdot \text{cm}^3)^{-1} \tag{14.239}$$

The units of $g_c(E_{c2})$ are per unit energy per unit volume, which is verified by

$$\frac{m_e^*}{\pi L_x \hbar^2} = \frac{\text{kg}}{\text{m}(\text{J} \cdot \text{s})^2} = \frac{1}{\text{J} \cdot \text{m}^3} \tag{14.240}$$

where use was made of

$$\text{J} = \text{force} \times \text{distance} = \text{kg}\frac{\text{m}}{\text{s}^2} \cdot \text{m}$$

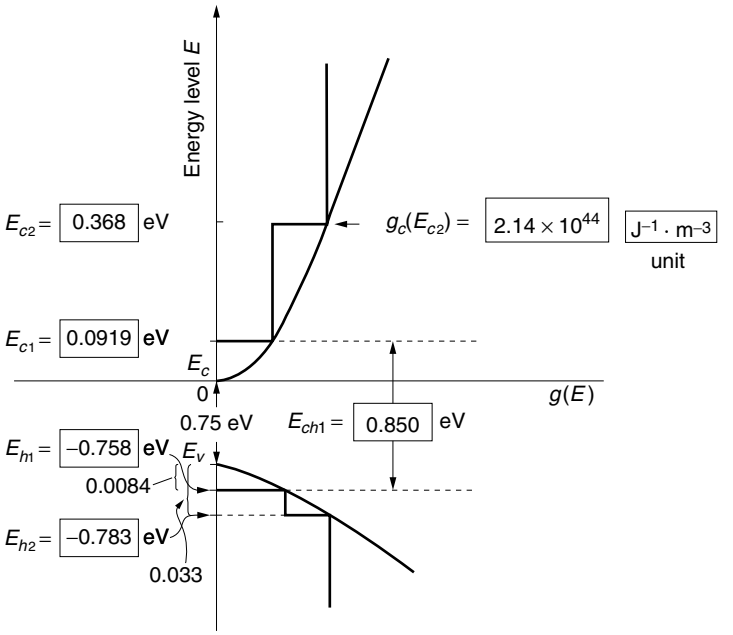


Figure 14.63 The density $g(E)$ of states versus energy level for a $Ga_{0.47}In_{0.53}As$ MQW laser. The results of the calculations in Fig. 14.62 are given.

Thus, $\int g(E_c)\,dE_c$ represents the number of states per unit volume. The results are shown in Fig. 14.63.

The energy bandgap between the bottom of the quantized level in the conduction band and top of the quantized level in the valence band is

$$E_{c1} + E_{h1} + 0.75 = 0.850 \text{ eV}$$

(b) The quantum of the $\lambda = 1.55$ μm photon is

$$h\nu = (4.14 \times 10^{-15}) \left(\frac{3 \times 10^8}{1.55 \times 10^{-6}} \right)$$
$$= 0.801 \text{ eV}$$

Alternatively, this result could have been obtained directly from the formula

$$\lambda = \frac{1.24}{E \text{ (eV)}} \mu\text{m} \tag{14.241}$$

From the energy diagram, we have

$$E_{c1} + E_{h1} + E_g = h\nu \tag{14.242}$$

$$\left[\frac{\hbar^2}{2m_e^*} \left(\frac{\pi}{L_x} \right)^2 + \frac{\hbar^2}{2m_h^*} \left(\frac{\pi}{L_x} \right)^2 \right] \frac{1}{1.6 \times 10^{-19}} + 0.75 = 0.801$$

Solving for L_x gives

$$L_x = 14 \text{ nm} \qquad \square$$

By expanding the quantum size effect to two dimensions, and then to all three dimensions, a quantum wire structure and a quantum dot (box) structure can be formed, respectively. Lasers using the quantum effect in more than one dimension further accentuate the merits of the MQW lasers [37,38]. In the next section, the density of states for the quantum wire and quantum dot will be obtained for comparison with the quantum well.

14.12.5 Density of States of a Quantum Wire and Quantum Dot

Such a potential well as shown in Fig. 14.64 confined in both x and y directions is called a quantum wire. With the quantum wire, Eq. (14.196) is modified as

$$\frac{2m_e^* E}{\hbar^2}\left(\frac{L}{\pi}\right)^2 = n_x'^2 + n_y'^2 + n_z^2 \tag{14.243}$$

with

$$n_x' = \frac{L}{L_x} n_x \tag{14.244}$$

$$n_y' = \frac{L}{L_y} n_y \tag{14.245}$$

Let us assume

$$L_x = L_y = L_d \tag{14.246}$$

L_x and L_y are significantly narrower than L and the lattice spacings are now wide in both x and y directions and narrow only in the z direction. The lattice points corresponding to the lowest energy look like a post located at $n_x' = n_y' = L/L_d$, as shown in Fig. 14.64a. Since neither n_x' nor n_y' is allowed to be negative, posts exist only in the first quadrant of the $n_x' - n_y'$ plane. The radius R_1 of the sphere that just touches the post nearest the origin is $\sqrt{2}L/L_d$ and its associated energy E_1 is, from Eq. (14.198),

$$E_1 = \frac{\hbar^2}{m_e^*}\left(\frac{\pi}{L_d}\right)^2 \tag{14.247}$$

Toward the objective of finding the density of states, the number of modes M with energy smaller than E will be calculated by counting the lattice points that are enclosed by the sphere. From Figs. 14.64b and 14.64c, the number of modes M is equal to r and

$$M = \sqrt{R^2 - R_1^2} \tag{14.248}$$

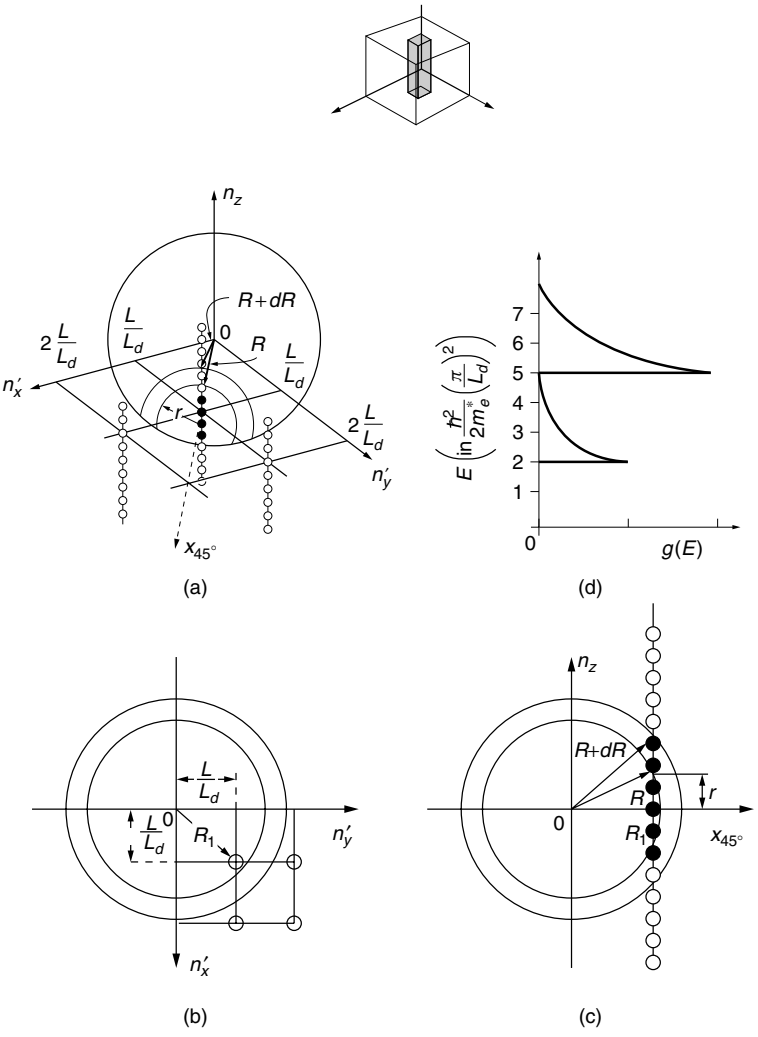


Figure 14.64 Calculation of the density of states of a quantum wire. (a) Lattice points. (b) Plan view. (c) Side view. (d) Density of states.

and

$$\frac{dM}{dR} = \frac{R}{\sqrt{R^2 - R_1^2}} \tag{14.249}$$

From Eq. (14.198), the derivative of E with respect to R is

$$\frac{dE}{dR} = \frac{\hbar^2}{2m_e^*}\left(\frac{\pi}{L}\right)^2 \cdot 2R \tag{14.250}$$

The number of modes dM between E and $E + dE$ is

$$dM = \frac{dM}{dR}\frac{dR}{dE}\,dE \tag{14.251}$$

Multiplying by 2 for spin states and dividing by LL_d^2 for the number per unit volume, the number of states $g_1(E)\,dE$ between E and $E + dE$ becomes

$$g_1(E)\,dE = \frac{2}{LL_d^2}\frac{dM}{dR}\frac{dR}{dE}\,dE \tag{14.252}$$

Inserting Eqs. (14.249) and (14.250) into Eq. (14.252) and summarizing the results gives

$$g_1(E) = \begin{cases} 0 & E < E_1 \\ \dfrac{\sqrt{2m_e^*}}{\hbar}\dfrac{1}{\pi L_d^2}\dfrac{1}{\sqrt{E - E_1}} & E > E_1 \end{cases} \tag{14.253}$$

with

$$E_1 = 2\frac{\hbar^2}{2m_e^*}\left(\frac{\pi}{L_d}\right)^2 \tag{14.254}$$

$g_1(E)$ is zero until the sphere reaches the first post and then becomes $g_1(E)$ after the sphere encloses the first post. $g_1(E)$ decreases with $\sqrt{E - E_1}$.

With a further increase in the radius of the sphere, the sphere simultaneously reaches the next two posts located at $n_x' = L/L_d$, $n_y' = 2L/L_d$ and at $n_x' = 2L/L_d$, $n_y' = L/L_d$.

Both posts are located at a distance $\sqrt{5}L/L_d$ from the origin. The energy associated with the radius is, from Eq. (14.198),

$$E_2 = 5\frac{\hbar^2}{2m_e^*}\left(\frac{\pi}{L_d}\right)^2 \tag{14.255}$$

The density of states $g_2(E)$ associated with these two posts is obtained in a similar manner as Eq. (14.253) and

$$g_2(E) = \begin{cases} 0 & E < E_2 \\ \dfrac{\sqrt{2m_e^*}}{\hbar}\dfrac{2}{\pi L_d^2}\dfrac{1}{\sqrt{E - E_2}} & E > E_2 \end{cases} \tag{14.256}$$

with

$$E_2 = 5\frac{\hbar^2}{2m_e^*}\left(\frac{\pi}{L_d}\right)^2 \tag{14.257}$$

With the sphere totally enclosing the first three posts, the density of states becomes

$$g(E) = \begin{cases} 0 & E < E_1 \\ g_1(E) & E_1 < E < E_2 \\ g_1(E) + g_2(E) & E_2 < E < E_3 \end{cases} \tag{14.258}$$

where

$$E_3 = 9\frac{\hbar^2}{2m_e^*}\left(\frac{\pi}{L_d}\right)^2 \tag{14.259}$$

For the region $E > E_3$ the density of states can be obtained in a similar manner.

The overall density of states of the quantum wire is plotted in Fig. 14.64d. The regions of high density of states are more concentrated than those of the quantum well.

Next, the density of states of the quantum dot will be calculated. The quantum dot has walls in all three dimensions, as shown in the inset of Fig. 14.65. Equation (14.196) has to be modified as

$$\frac{2m_e^* E}{\hbar^2}\left(\frac{L}{2\pi}\right)^2 = n_x'^2 + n_y'^2 + n_z'^2 \tag{14.260}$$

Let us assume that

$$L_x = L_y = L_z = L_d \tag{14.261}$$

The lattice points are dispersed in the (n_x', n_y', n_z') space. The density of states is zero until the sphere reaches the lattice point nearest the origin, which is located at

$$(n_x', n_y', n_z') = \frac{L}{L_d}(1, 1, 1) \tag{14.262}$$

The radius of the sphere that reaches the nearest lattice point is

$$R = \sqrt{3}\left(\frac{L}{L_d}\right) \tag{14.263}$$

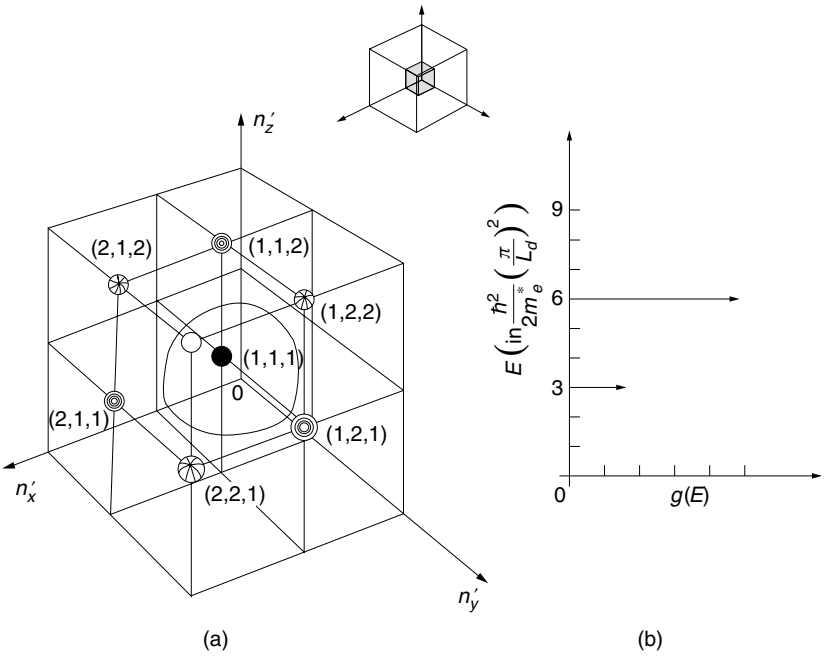


Figure 14.65 Calculation of the density of states of a quantum dot. (a) Lattice points. (b) Density of states.

The energy E associated with this sphere is, from Eq. (14.198),

$$E_1 = 3\frac{\hbar^2}{2m_e^*}\left(\frac{\pi}{L_d}\right)^2 \tag{14.264}$$

Thus, after multiplying by 2 for spin states, the density of states $g_1(E)$ before reaching the second nearest lattice points is

$$g_1(E) = 2\delta(E - E_1) \quad \text{for } E < E_1 \tag{14.265}$$

The next lattice points are located at

$$(n_x', n_y', n_z) = \begin{cases} \dfrac{L}{L_d}(2, 1, 1) \\ \dfrac{L}{L_d}(1, 2, 1) \\ \dfrac{L}{L_d}(1, 1, 2) \end{cases} \tag{14.266}$$

These are all in the radius

$$R = \sqrt{6}\left(\frac{L}{L_d}\right) \tag{14.267}$$

The energy E associated with the sphere is, from Eq. (14.198),

$$E_2 = 6\frac{\hbar^2}{2m_e^*}\left(\frac{\pi}{L_d}\right)^2 \tag{14.268}$$

Thus, the density of states $g_2(E)$ before reaching E_3 is

$$g_2(E) = 6\delta(E - E_2) \quad \text{for } E < E_2 \tag{14.269}$$

The next lattice points are located at

$$(n_x', n_y', n_z') = \begin{cases} \dfrac{L}{L_d}(2, 2, 1) \\ \dfrac{L}{L_d}(2, 1, 2) \\ \dfrac{L}{L_d}(1, 2, 2) \end{cases} \tag{14.270}$$

These are all at the radius

$$R = \sqrt{9}\left(\frac{L}{L_d}\right) \tag{14.271}$$

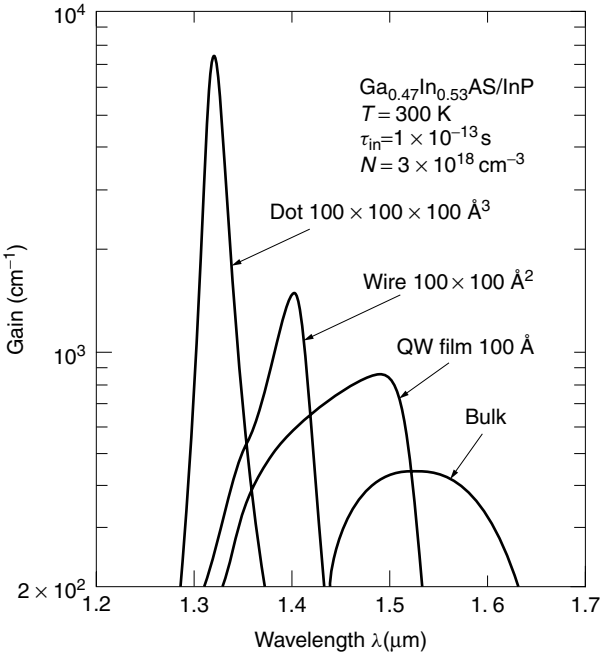


Figure 14.66 Gain spectra calculated for $Ga_{0.47}In_{0.53}As/InP$: $100 \times 100 \times 100$ Å^3 cubic quantum dot, 100×100 Å^2 quantum wire, 100 Å quantum film, and bulk crystal at $T = 300$ K. (After M. Asada, Y. Miyamoto, and Y. Suematsu [37].)

The energy E associated with the sphere is, from Eq. (14.198),

$$E_3 = 9\frac{\hbar^2}{2m_e^*}\left(\frac{\pi}{L_d}\right)^2 \tag{14.272}$$

The results are plotted in Fig. 14.65b. They consist of a series of delta functions and the curve of the density of states of the quantum dot is the sharpest of all three kinds of quantum effects. Also, by repeating the same procedure as demonstrated in Fig. 14.4, the order of the gain curves from sharpest to least sharp is as follows: quantum dot, quantum wire, quantum well, and then bulk. Calculated results are shown in Fig. 14.66.

14.13 ERBIUM-DOPED FIBER LASER

An erbium-doped fiber amplifier can immediately be converted into a fiber laser if its output light is fed back to the input port. The feedback is achieved either by connecting the input port to the output port, forming a ring-type fiber laser, or by installing reflectors at the input and output ports, thereby creating a Fabry–Pérot-type fiber laser. The hybrid of the above two types is known as a *sigma laser*.

Figure 14.67a shows a schematic of the ring-type fiber laser. The traveling-wave [39] operation eliminates the problem of spatial hole burning (see Section 14.4.3.3). To

ensure the traveling-wave excitation, an optical isolator is installed. The optical isolator allows propagation in only one direction, as indicated by the arrow in Fig. 14.67a.

Since the length of the fiber ring is long, the frequency spacing between the longitudinal modes is close. For instance, from Eq. (14.36), $\Delta\nu$ is of the order of 10 MHz for a ring length of 10 meters. Tight control of the fiber length is important in order for the ring laser to be free from mode instabilities due to either temperature or vibration. A device for stretching the fiber is installed to fine tune the length of the fiber ring. The fiber is stretched by means of a piezoelectric ceramic cylinder onto which the fiber is wound. The diameter of the cylinder is controlled by the servolooped applied electric field [40] to stabilize the wavelength of oscillation.

Figure 14.67b is a schematic of the Fabry–Pérot-type fiber laser [41]. Referring to the diagram, the components and their symbols are explained as follows. An erbium-doped fiber amplifier A serves as the active medium. The reflector HM on the left is a half-mirror, and the reflector FRM on the right is a combination of Faraday rotator and mirror. The direction of polarization of the wave reflected from the FRM is rotated by 90°, which means that horizontally polarized light after reflection is converted into a vertically polarized wave, or vice versa. The polarizing beamsplitter (PBS) reflects the vertically polarized wave and transmits the horizontally polarized wave (see Section 6.5.2). The external grating G controls the wavelength of oscillation.

The path of light through the Fabry–Pérot fiber laser will now be traced. The directions of polarization are indicated by H and V for horizontal and vertical polarization, respectively. The starting point is taken as the light just reflected from mirror HM.

$$\text{HM} \xrightarrow{H} \text{PBS} \xrightarrow{H} \text{A} \xrightarrow{H} \text{FRM} \xrightarrow{V}$$

$$\underbrace{\qquad\qquad}_{\text{to the right}}$$

$$\xrightarrow{V} \text{A} \xrightarrow{V} \text{PBS} \xrightarrow{V} \text{G} \xrightarrow{V} \text{PBS} \xrightarrow{V}$$

$$\underbrace{\qquad\qquad}_{\text{to the left}}$$

$$\xrightarrow{V} \text{A} \xrightarrow{V} \text{FRM} \xrightarrow{H} \text{A} \xrightarrow{H} \text{PBS} \xrightarrow{H}$$

$$\underbrace{\qquad\qquad}_{\text{to the right}} \qquad \underbrace{\qquad\qquad}_{\text{to the left}}$$

$$\xrightarrow{H} \text{HM} \xrightarrow{H} \text{PBS} \cdots$$

As seen above, the directions of polarization of the two counterpropagating waves in A are always perpendicular to each other so that no standing wave is formed and the problem of spatial hole burning is eliminated.

Figure 14.67c is a hybrid of the above two systems and is called a sigma laser [41]. Instead of being reflected back by a mirror at the left end, the light beam makes a U-turn. The U-turn contains a kind of optical isolator that is accompanied by a rotation of the direction of polarization by 90° (the optical isolator can be replaced by a polarization-maintaining fiber twisted by 90°).

Let us say a horizontally polarized beam is incident onto the PBS from A. The beam will pass straight through the PBS and reach the optical isolator. The output from the

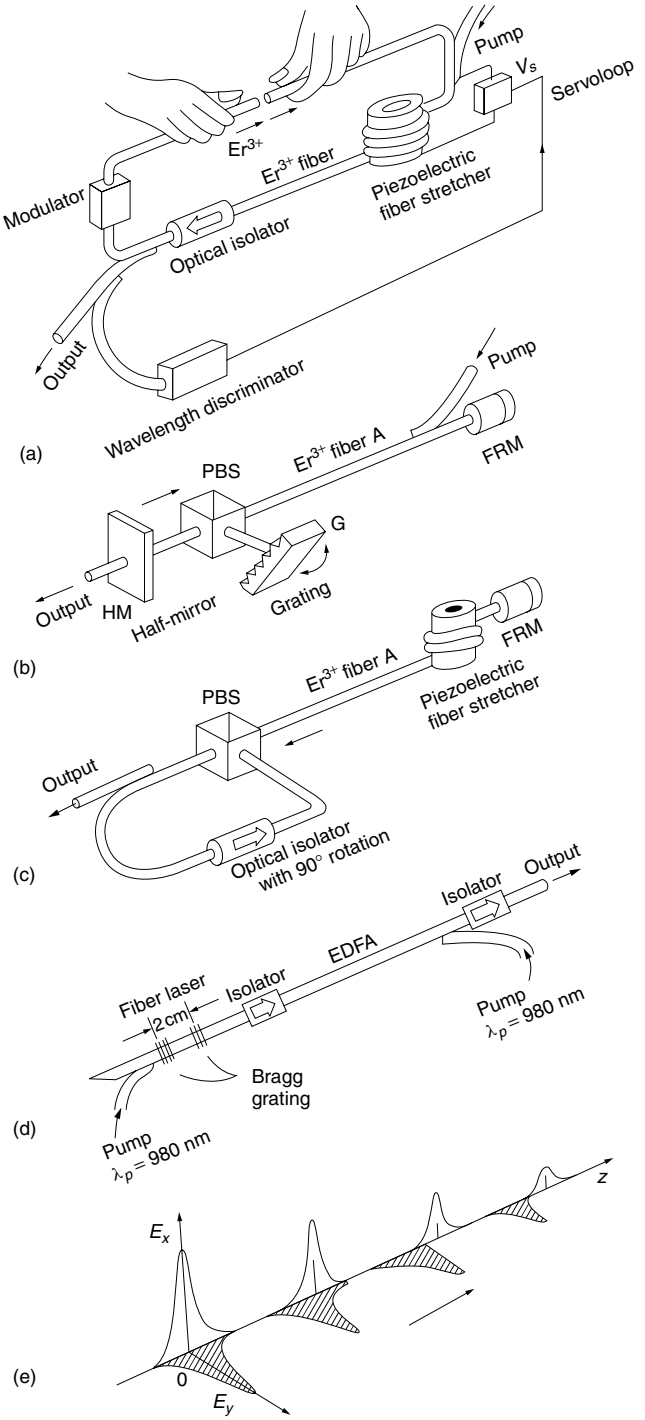


Figure 14.67 Erbium-doped fiber laser. (a) Ring type. (b) Fabry–Pérot type. (c) Sigma type. (d) Erbium-doped fiber grating laser (EDFGL). (e) Pulse spread due to the birefringence of the Er^{3+}-doped fiber.

optical isolator is vertically polarized. The vertically polarized wave is deflected by the PBS toward A. Upon reflection from the FRM, the vertically polarized wave changes back to horizontal polarization. The horizontally polarized wave is incident onto the PBS, and we are back to our starting point. The wave traveling to the right in A is always vertically polarized, and the wave traveling to the left is always horizontally polarized. This eliminates the problem of spatial hole burning.

Figure 14.67d is another type of fiber laser [42]. Bragg reflectors that have been written by photobleaching the glass are used at both ends of the fiber laser. The structure is quite similar to the solid-state DBR laser shown in Fig. 14.30. This is a fiber version of it. In the laser itself, the fiber length is only a few centimeters long, and since gain is proportional to length, a short length means less output power. The lack of output power is supplemented by an EDFA, which accompanies the laser. One advantage of a short laser section is the minimization of the output power fluctuation due to the relaxation oscillation. Another advantage is the assurance of single longitudinal mode oscillation by making the separation of the adjacent longitudinal modes given by Eq. (14.34) wider. Another advantage of this system is that the same pump power can be used for both the laser and the accompanying EDFA. The pumping efficiency is not decreased despite the short laser section because the residually transmitted pump power is used in the EDFA.

The configuration of alternating vertical and horizontal polarizations is not only useful for getting rid of spatial hole burning, but also for preventing pulsewidth spread during pulse mode operation. As indicated in Fig. 14.67e, the pulse is decomposed into horizontal and vertical components. If the two components have different group velocities due to birefringence in the fiber, the pulsewidth spreads as it propagates in the fiber [43]. This can be prevented by alternating the directions of polarization in such a way as mentioned above.

The fiber laser can function in either continuous wave operation or pulse mode operation. Pulse modulation is achieved by either installing an amplitude modulator [42] in the erbium-doped fiber or pumping with pulsed light.

The state of polarization of the output can be made linearly polarized if polarization-maintaining erbium doped fiber [44] is used.

Subpicosecond pulses can be generated if the fiber laser is combined with a nonlinear amplifying loop, which generates an optical soliton wave (see Section 15.6) [45].

14.14 LIGHT-EMITTING DIODE (LED)

Recall that there are two mechanisms of emission of photons from the p-n junction due to the recombination of the electrons with holes. One is stimulated emission, which is the basic mechanism of the already mentioned semiconductor laser diode. The other is spontaneous emission, which is the basic mechanism of the light-emitting diode (LED). The light-emitting diode is the topic of this section.

14.14.1 LED Characteristics

The LED has quite a similar structure to a laser diode except that the LED does not have the laser cavity that is necessary to enhance the stimulated emission. The LED

is structurally much simpler than the laser diode, so that compared to the laser diode, the LED has advantages such as:

- The fabrication cost is lower.
- The LED can operate at higher temperatures, while its operation is less influenced by temperature variations.
- The drive current is moderately low, and the drive circuit is simple.
- The output light power is moderately linear with respect to the drive current.
- Its lifetime is extremely long, of the order of 10^8 hours.

The disadvantages of the LED are:

- The LED has a broader wavelength spectrum (10–50 nm).
- The cutoff frequency for modulation is normally less than 100 MHz.
- The radiation beam is spatially wider ($30^\circ \times 120^\circ$) and harder to couple into an optical fiber.

LEDs with emission in the visible wavelength range are widely used as status indicators in electronic instruments and display panels, mainly because of their long life, low consumption of electric power, and low cost.

14.14.2 LED Structure

There are two LED types: one is the surface-emitting LED and the other is the edge-emitting LED.

The surface-emitting LED is sometimes called the Burrus-type LED. The structure is shown in Fig. 14.68. The structure is quite similar to the surface-emitting laser diode, except for the absence of the cavity reflectors. The advantage of the surface-emitting structure as shown in Fig. 14.68 is that the light generated in the active layer has only to traverse the 1-μm thickness of the active layer in order to exit via the 3-μm-thick cladding layer; the light does not have to travel the entire 20–30 μm length of the active region. Light whose quantum $h\nu$ is larger than the bandgap of the transmission medium is subjected to absorption. The amount of absorption increases with $h\nu$. Since the LED has such a wide spectrum, the longer the light stays in the active region, the more severely absorbed the higher frequency components become.

The refractive index of the GaAs material is large ($n = 3.5$), and the critical angle is small, so that only the light inside a cone with an apex angle of $\pm 16.6^\circ$ can exit the LED from the active layer. The surface-emitting structure also creates a difficulty in coupling the light to an optical fiber. Moreover, the dimension of the emitting area is large, and a short focal length lens is needed to project the image of the active area onto the end of the core of the optical fiber. A good short-focal-length lens is hard to fabricate.

Figure 14.69 shows the structure of the edge-emitting LED. It is quite similar to the electrode stripe laser except that the active layer is not excited along the entire length. For better external efficiency (input electric power to output light power), only the portion of the active region that is near the exit is excited. This structure also reduces the cavity resonance effect and heat sink problems. The cavity resonance effect in the

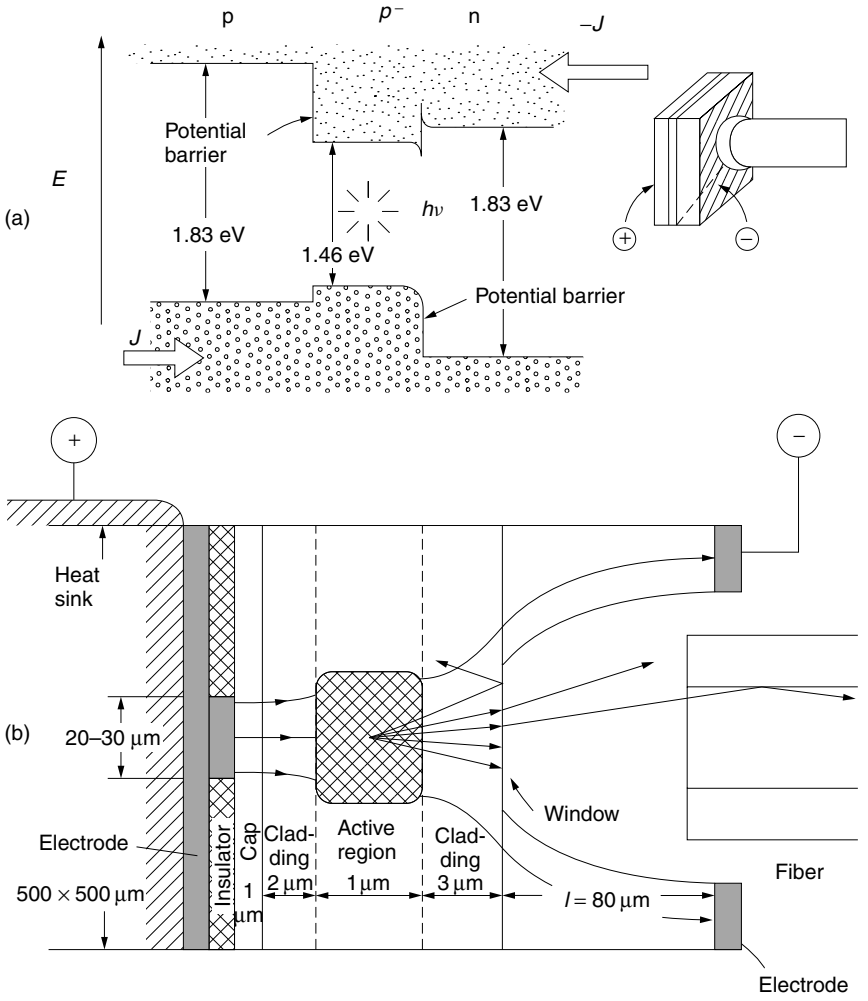


Figure 14.68 Structure of surface-emitting LED. (a) Energy bandgap. (b) Structure: cladding layer, $Al_{0.3}Ga_{0.7}As$; active layer, $Al_{0.03}Ga_{0.97}As$.

surface-emitting LED arises from the cladding interfaces that act as low-reflectivity reflectors. While the "edge" of the edge-emitting LED does act as a reflective boundary, the reflected light is likely to be absorbed before reaching the other end, and a resonant cavity is not established.

14.15 FIBER RAMAN LASERS

A method of amplifying the light by means of the enhancement of the Stokes radiation has already been described in Section 12.8.1. Fiber Raman lasers [46–48] are a natural extension of this technique. When the pump light with frequency ω_L is injected into the Raman fiber, the light acquires an increase in amplitude as well as a shift in frequency to $\omega_L - \omega_R$, where ω_R is the Raman shift frequency.

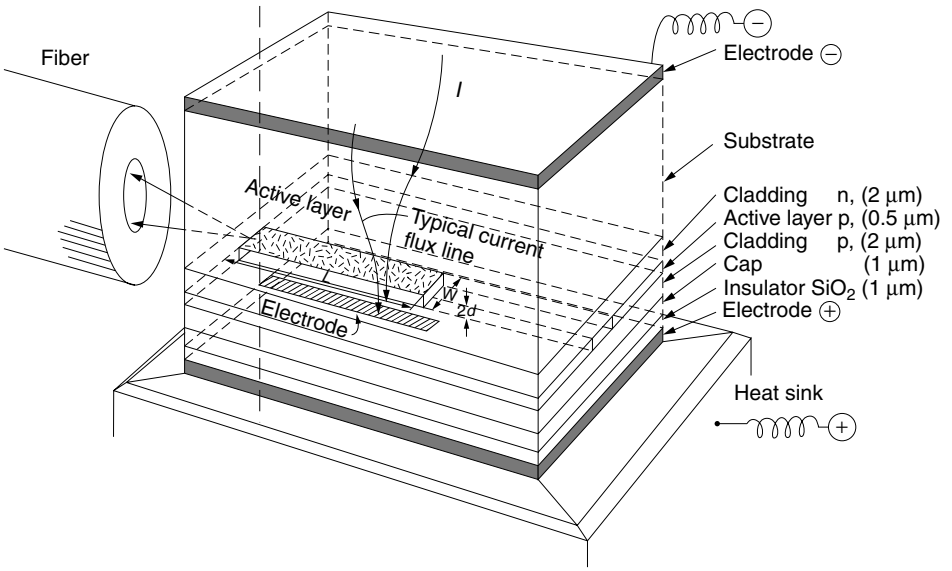


Figure 14.69 Structure of an edge-emitting LED. Cladding layers, $Al_{0.3}Ga_{0.7}As$, $E_g = 1.83$ eV. Active layer, $Al_{0.03}Ga_{0.97}As$, $E_g = 1.46$ eV.

Figure 14.70a shows a diagram of sets of fiber Bragg gratings (fiber Bragg gratings are described in Section 11.5). The center frequencies match the Raman shifted frequencies. Gratings $B_0, B_1, B_2, \ldots, B_n$ are placed on the right side of the GeO_2 fiber and $B_1', B_2', B_3', \ldots, B_n'$ on the left side, in order to form a GeO_2 fiber cascaded Raman laser resonator. All of the gratings are highly reflective at their respective frequencies except the outermost grating B_n which is only partially reflective. For better input coupling, grating B_0' is absent.

Pump light at ω_L is incident to the resonator. Due to the Raman effect, a portion of the incident light shifts its frequency to $\omega_L - \omega_R$. A portion of the $\omega_L - \omega_R$ component shifts further to $\omega_L - 2\omega_R$, which in turn shifts to $\omega_L - 3\omega_R$, and so forth. All of the components except the ω_L component pass through B_0. The $\omega_L - \omega_R$ component is reflected back and forth inside the cavity by B_1 and B_1', which are highly reflective at $\omega_L - \omega_R$ but pass other frequencies. Similarly, the $\omega_L - 2\omega_R$ component is reflected back and forth by B_2 and B_2', which are highly reflective at $\omega_L - 2\omega_R$ but pass other frequencies. Thus, lower order frequency components go back and forth inside the cavity until their frequency is shifted to the next higher order. This process repeats until the light reaches the outermost cavity gratings. At the outermost grating B_n which is partially reflective, a portion of the $\omega_L - n\omega_R$ component escapes. The escaped light is the laser output.

Figure 14.70b shows a schematic of the Raman fiber laser [46]. It is made of germanosilicate fiber (GeO_2-doped fiber) about 1 km long. The fiber Bragg gratings' center frequencies match the Raman-shifted frequencies up to the fifth order. The frequency of the pump light is ω_L and that of the output is $\omega_L - 5\omega_R$. The Raman shift frequency ω_R is 13.2 THz (450 cm^{-1}) in each order. If the pump frequency is set so that the output wavelength is $\lambda = 1.48$ μm, then the laser can be used as a pump light to an erbium-doped fiber amplifier.

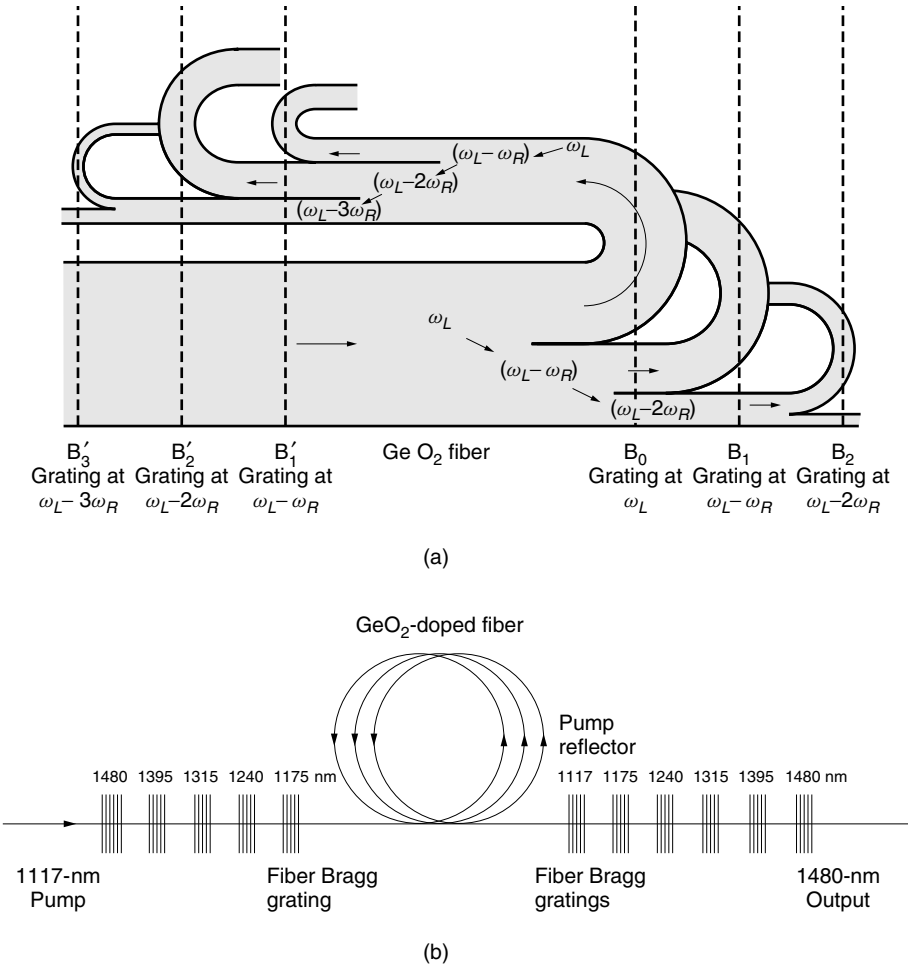


Figure 14.70 Fifth order 1480 nm cascaded Raman fiber laser. (a) Principle of operation. (b) Schematics. (After J. L. Archambault and S. G. Grubb [46].)

The main feature of the fiber Raman laser is simplicity with elegance. The conversion efficiency, which is the ratio of pump power to output power, is as high as 46%. The saturation power is high (0.4 W = 26 dBm). The biggest advantage of all is that even though the amount of frequency shift is set by the material of the fiber, there is no restriction on the pump wavelength. The frequency of the pump can be set according to the desired frequency of the output light.

14.16 SELECTION OF LIGHT SOURCES

The four major types of light sources that are most often used for fiber-optic communication systems have been presented. They are the semiconductor laser diode (LD), the solid-state laser, fiber lasers, and the light-emitting diode (LED). Although all four light sources possess such advantages as being rugged, compact, reliable, long-lived, and low electric consumption, there are some differences in

the characteristics, and proper selection of the source is important. The spectra are compared in Fig. 14.71.

The laser diode is the most widely used because of its high light power output, the reasonable purity of its spectrum, and the ease of direct modulation. The laser diode can be modulated up to tens of gigahertz. The disadvantages are turn-on delay time, temperature-induced instability, and sensitivity to the reflected light from the load. The lifetime of the laser diode, however, is not as long as that of a light-emitting diode.

The Nd:YAG solid-state laser of the MISER configuration (see Section 14.9.4) has the special feature of an extremely narrow spectral width of 10^{-7} nm (or 10–100 kHz) and is suitable for the main and local oscillator lasers of an optical heterodyne system. The output power is more or less the same as the laser diode. The Nd:YAG solid-state laser is optically pumped by an array of the semiconductor laser diodes and is structurally the most complicated. The disadvantages are the difficulty of high-speed direct modulation and the higher cost.

Fiber lasers are relatively simple structurally and compact. They can easily be connected to optical fiber systems. They are useful not only for fiber-optic communication systems but also for other applications as well, such as sources for fiber-optic gyros.

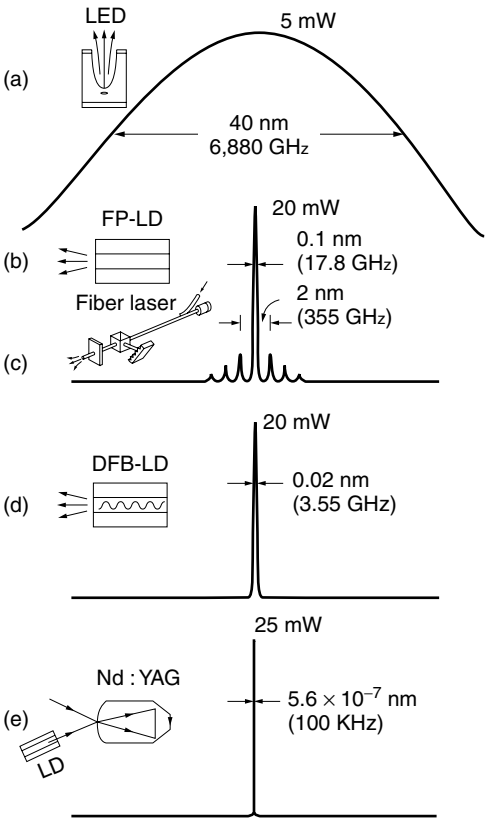


Figure 14.71 Comparison of the spectra of sources. (a) Light-emitting diode (LED). (b) Fabry–perot laser diode (FP-LD). (c) Erbium-doped fiber laser. (d) Distributed feedback laser diode (DFB-LD). (e) Laser diode driven Nd:YAG solid-state laser.

The LED emits radiation in a broad spectrum of light as compared to laser sources. Because of this broadness of the spectrum, the chromatic dispersion in the glass fiber is large and the distance of transmission is much more limited than that of the system with a laser diode. Another shortcoming of the LED is the limit on the modulation frequency, which is limited to less than a few hundred megahertz. The limitation on the modulation frequency is inherent. The cutoff modulation frequency of the LED is $f_c = 1/2\pi\tau_n$, whereas that of the LD is $f_c \propto 1/2\pi\sqrt{\tau_n \tau_s}$, where τ_n is the lifetime of the injected electrons and τ_s is the lifetime of the photons before leaving the laser cavity (see Section 14.3). τ_n is of the order of nanoseconds, while τ_s is of the order of picoseconds. τ_n can be reduced by increasing the doping at the cost of the output light power. Further drawbacks of the LED are the difficulty of coupling the output light into a fiber, because of the spread of the radiation pattern of the LED, and the smaller output light power.

The LED, however, is structurally simple and rugged and has the longest lifetime of 10^8 hours (an estimated one hundred centuries!). It is also less temperature dependent and has simpler drive circuits than a laser diode.

PROBLEMS

14.1 An $Al_xGa_{1-x}As$ laser diode is operated at $J = 1.7J_{th}$. Find the relaxation time $1/\gamma$ of the relaxation oscillation and the upper limit of the modulation frequency f_r. The physical parameters are as follows:

Index of refraction	$n = 3.5$
Length of the cavity	$L = 2 \times 10^{-2}$ cm
Power attenuation constant	$\alpha = 25\ \text{cm}^{-1}$

Cleaved surfaces were used as cavity end mirrors. The lifetime of the spontaneous emission measured by the turn-on delay time was $\tau_n = 2 \times 10^{-9}$ s.

14.2 Find an expression for the decay constant γ of the relaxation oscillation that corresponds to Eq. (14.81) without the assumption that $N_{th} >> N_\alpha$.

14.3 Find the following quantities for a $Ga_{0.47}In_{0.53}As$ semiconductor laser with the dimensions and physical constants given in Fig. P14.3.

(a) The turn-on-delay time at $J/J_{th} = 1.2$.

(b) The length of the laser cavity.

(c) The threshold electron density (use Fig. 14.6).

(d) The threshold current.

(e) The relationship between the light output power and the injection current.

14.4 In connection with Fig. 14.55b, decompose the two complex fields $\tilde{E}_1 = E_1 e^{j\phi_1}$ and $\tilde{E}_3 = E_3 e^{j\phi_3}$, into symmetric $\tilde{E}_s$ and antisymmetric $\tilde{E}_a$ components using a phasor diagram. The tilde symbol represents a phasor vector on the complex plane, and not a spatial vector.

14.5 Determine the values of I_{th0}, η_{th0}, T_0, and T_0' from Fig. 14.26.

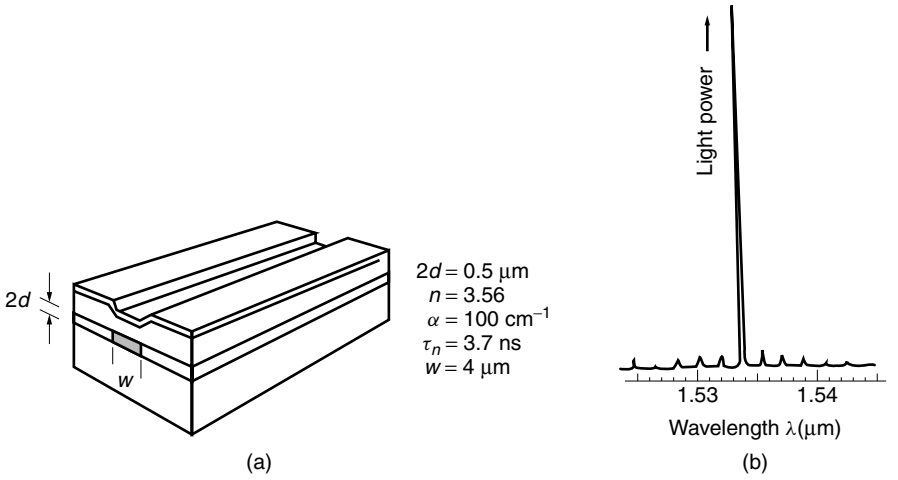


Figure P14.3 Semiconductor laser. (a) Dimensions. (b) Power spectrum.

14.6 The gain of a semiconductor laser diode for the case of $T = 0$ was graphically obtained in Fig. 14.4. The Fermi distribution function used in Fig. 14.4 was a curve with an abrupt ramp. Try the same, but this time, with a Fermi distribution function with a linear ramp representing the case of $T \neq 0$ so as to compare spectral distributions of $g(h\nu)$ for the case of $T \neq 0$ with that for the case of $T = 0$.

REFERENCES

1. A. E. Siegman, *Lasers*, University Science Books, Mill Valley, CA, 1986.
2. T. F. Carruthers and I. N. Duling III, "10 GHz, single-polarization, actively modelocked picosecond erbium-fiber laser," *Optical Fiber Communication Conference Record*, TuB6, San Jose, CA, Feb. 1996.
3. R. A. Chodzko, J. M. Bernard, and H. Mirels, "Coherent combination of multiline lasers," *Proc. SPIE* **1224**, 239–253 (1990).
4. G. P. Agrawal (ed.), *Semiconductor Lasers—Past, Present and Future*, American Institute of Physics, Woodbury, NY, 1995.
5. B. Sapoval and C. Hermann, *Physics of Semiconductors*, Springer-Verlag, Berlin, 1995.
6. C. Kittel, *Introduction to Solid State Physics*, 7th ed., Wiley, New York, 1996.
7. M. Asada and Y. Suematsu, "Density-matrix theory of semiconductor lasers with relaxation broadening model-gain and gain-suppression in semiconductor lasers," *IEEE J. Quantum Electron.* **QE21**(5), 434–442 (1985).
8. H. Yonezu, *Elements of Fiber Optical Communication Systems Emitter and Detectors*, 6th ed., Kogakutosho, Tokyo, 2000.
9. R. Tucker, C. Lin, and C. A. Burrus, "High-frequency small-signal modulation characteristics of short-cavity InGaAsP lasers," *Electron. Lett.* **20**(10), 393–394 (1984).
10. H. C. Casey, Jr. and M. B. Panish, *Heterostructure Lasers*, Vols. I and II, Academic Press, New York, 1978.
11. D. Botez, "Laser diodes are power-packed," *IEEE Spectrum*, 43–53 (June 1985).

12. T. K. Woodward, B. Tell, W. H. Knox, and J. B. Stark, "Suppressed photocurrent multiple-quantum-well optical modulator by proton implantation," *Appl. Phys. Lett.* **60**(6), 742–744 (1992).

13. E. Oomura, H. Higuchi, Y. Sakakibara, R. Hirano, H. Namizaki, W. Susaki, K. Ikeda, and K. Fujikawa, "InGaAsP/InP buried crescent laser diode emitting at 1.3 μm wavelength," *IEEE J. Quantum Electron*. **QE20**(8), 866–874 (1984).

14. H. Kawanishi, T. Morimoto, M. Yamaguchi, S. Kaneiwa, N. Miyauchi, T. Yoshida, H. Hayashi, S. Yano, and T. Hijikata, "High-power CW operation in V-channeled substrate inner-stripe lasers with "Torch"-shaped waveguide," *Jpn. J. Appl. Phys. 2 Lett.* **27**(7), L1310–L1312 (1988).

15. Y. Suematsu, *Semiconductor Laser and Optic ICs* (in Japanese), Ohm-Sha, Tokyo, 1984.

16. N. Chinone, T. Kuroda, T. Ohtoshi, T. Takahashi, and T. Kajimura, "Mode-hopping noise in index-guided semiconductor lasers, and its reduction by saturable absorbers," *IEEE J. Quantum Electron.* **QE21**(8), 1264–1270 (1985).

17. P. W. Smith, Y. Silberberg, and D. A. B. Miller, "Mode locking of semiconductor diode lasers using saturable excitonic nonlinearities," *J. Opt. Soc. Am. B.* **2**(7), 1228–1236 (1985).

18. M. Kitamura, M. Yamaguchi, S. Murata, I. Mito, and K. Kobayashi, "High-performance single-longitudinal-mode operation of InGaAsP/InP DFB-DC-PBH LDs," *J. Lightwave Technol.*, **2**(4), 363–369 (1984).

19. Y. Huang, H. Yamada, T. Okuda, T. Torikai, and T. Uji, "External optical feedback resistant characteristics in partially-corrugated waveguide laser diodes," *Optical Fiber Communication Conference Record*, TuH1, San Jose, CA, Feb. 1996.

20. T. Hamano, K. Iga, and T. Baba, "Spontaneous emission behavior and its injection level dependence in three-dimensional microcavity surface emitting lasers," *Electronics and Communications in Japan* (Translation into English by Scripta Technica), Part 2, Vol. 79, pp. 46–54, Wiley, New York, 1996.

21. R. T. Chen, L. Lin, C. Choi, Y. J. Liu, B. Bihari, L. Wu, S. Tang, R. Wickman, B. Picor, M. K. Hibbs-Brenner, J. Bristow, and Y. S. Liu, "Fully embedded board-level guided-wave optoelectronic interconnects," *Proc. IEEE* **88**(6), 780–793 (2000).

22. T. Kjellberg, S. Nilsson, T. Klinga, B. Broberg, and R. Schatz, "Investigation on the spectral characteristics of DFB lasers with different grating configurations made by electron-beam lithography," *J. Lightwave Technol.* **11**(9), 1405–1415 (1993).

23. H. Kogelnik and C. V. Shank, "Coupled-wave theory of distributed feedback lasers," *J. Appl. Phys.* **43**(5), 2327–2335 (1972).

24. M. Yamaguchi, T. Numai, Y. Koizumi, I. Mito, and K. Kobayashi, "Stable single-longitudinal-mode operation in $\lambda/4$ shifted DFB-DC-PBH laser diodes," *Optical Fiber Communication Conference Record*, TUC4, Reno, NV, Jan. 1987.

25. K. Kobayashi and I. Mito, "Progress in narrow-linewidth tunable laser sources," *Optical Fiber Communication Conference Record*, WC-1, Reno, NV, Jan. 1987.

26. T. J. Kane and R. L. Byer, "Monolithic, unidirectional single-mode Nd:YAG ring laser," *Opt. Lett.* **10**, 65–67 (1985).

27. K. Kobayashi and I. Mito, "Single frequency and tunable laser diodes," *J. Lightwave Technol.* **6**(11), 1623–1633 (1988).

28. K. Kobayashi, "Semiconductor lasers for optical fiber networks," *Conference on Lasers and Electrooptics (CLEO)*, Baltimore, pp. 10–11, May 1995.

29. A. Yariv and P. Yeh, *Optical Waves in Crystals: Propagation and Control of Laser Radiation*, Wiley, New York, 1984.

30. Y. Tohmori, Y. Yoshikuni, H. Ishii, F. Kano, T. Tamamura, Y. Kondo, and M. Yamamoto, "Broad-range wavelength-tunable superstructure grating (SSG) DBR lasers," *IEEE J. Quantum Electron.* **29**(6), 1817–1823 (1993).
31. G. Sarlet, G. Morthier, and R. Baets, "Control of widely tunable SSG-DBR lasers for dense wavelength division multiplexing," *J. Lightwave Tech.* **18**(8), 1128–1138 (2000).
32. L. Fan, M. C. Wu, H. C. Lee, and P. Grodzinski, "4 × 4 reconfigurable optical interconnection network using beam-steering vertical cavity surface emitting lasers," *Optical Fiber Communication Conference Record*, TuC2, San Jose, CA, Feb. 1996.
33. A. Husain and J. N. Lee (Eds.), "Special Issue on Interconnections for information Processing," *J. Lightwave Technol.* **13**(6), 985–1183 (1995).
34. D. Botez and D. R. Scifres (Eds.), *Diode Laser Arrays*, Cambridge University Press, Cambridge, 1995.
35. P. S. Zory Jr. (Ed.), *Quantum Well Lasers*, Academic Press, New York, 1993.
36. M. Asada, A. Kameyama, and Y. Suematsu, "Gain and intervalence band absorption in quantum-well lasers," *IEEE J. Quantum Electron.* **QE20**(7), 745–753 (1984).
37. M. Asada, Y. Miyamoto, and Y. Suematsu, "Gain and the threshold of three dimensional quantum-box lasers," *IEEE J. Quantum Electron.* **QE-22**(9), 1915–1921 (1986).
38. H. Hirayama, J. Yoshida, Y. Miyake, and M. Asada, "Carrier capture time and its effect on the efficiency of quantum-well lasers," *IEEE J. Quantum Electron*. **30**(1), 54–62 (1994).
39. E. Yoshida, Y. Kimura, and M. Nakazawa, "20 GHz, 1.8 ps pulse generation from a regeneratively modelocked erbium-doped fibre laser and its femtosecond pulse compression," *Electron. Lett.* **31**(5), 376–378 (1995).
40. T. F. Carruthers and I. N. Duling III, "10 GHz single-polarization, actively mode-locked picosecond erbium-fiber laser," *Optical Fiber Communication Conference Record*, TuB6, San Jose, CA, Feb. 1996.
41. I. N. Duling III and R. D. Esman, "Single-polarization fiber amplifier," *Electron. Lett.* **28**(12), 1126–1128 (1992).
42. G. A. Ball and W. H. Glenn, "Design of a single-mode linear cavity erbium fiber laser utilizing Bragg reflectors," *J. Lightwave Technol.* **10**(10), 1338–1343 (1992).
43. T. Morioka, H. Takara, K. Mori, and M. Saruwatari, "Ultrafast reflective optical Kerr demultiplexer using polarizatioin rotation mirror," *Electron. Lett.* **28**(6), 521–522 (1992).
44. K. Tajima, "$Er^{3\pm}$ doped single-polarization optical fibers," *Electron. Lett.* **26**(18), 1498–1499 (1990).
45. I. N. Duling III, "Subpicosecond all-fiber erbium laser," *Electron. Lett.* **27**(6), 544–545 (1991).
46. J.-L. Archambault and S. G. Grubb, "Fiber gratings in lasers and amplifiers," *J. Lightwave Technol.* **15**(8), 1378–1390 (1997).
47. E. M. Dianov, M. V. Grekov, I. A. Bufetov, S. A. Vasiliev, O. I. Medvedkov, V. G. Plotnichenko, V. V. Koltashev, A. V. Belov, M. M. Bubnov, S. L. Semjonov, and A. M. Prokhorov, "CW high power 1.24 μm and 1.48 μm Raman lasers based on low loss phosphosilicate fiber," *Electron. Lett.* **33**(18), 1542–1544 (1997).
48. M. Prabhu, N. S. Kim, and K. Ueda, "Simultaneous double-color continuous wave Raman fiber laser at 1239 nm and 1484 nm using phosphosilicate fiber," *Optical Review* **7**(4) 277–280 (2000).

15

STATIONARY AND SOLITARY SOLUTIONS IN A NONLINEAR MEDIUM

This chapter presents the behavior of electric fields in nonlinear media, with an emphasis on Kerr media. Examples of materials that exhibit the Kerr effect are fused silica (SiO_2), benzene (C_6H_6), carbon disulfide (CS_2), water (H_2O), nitrotoluene ($C_6H_4(CH_3)NO_2$), and nitrobenzene ($C_6H_5NO_2$). In Kerr media, the refractive index is nonlinear and depends on the light intensity. The first half of the chapter discusses fields established by a continuous sinusoidal wave after the steady state has been reached. Such solutions are called stationary solutions.

Optical devices based on the Kerr effect change either the amount or the direction of transmission by altering the light intensity. The upper speed limit of these devices is much higher than that of devices based on the electrooptic effect due to the absence of electronic circuits to drive the electrodes. For instance, an optical switch utilizing the Kerr effect can be switched by the flick of a 1-ps light pulse. It is not possible to achieve this level of performance in an electrooptic switch, which would require an electronic circuit that could put out several volts with a bandwidth of 1 THz.

The second half of the chapter is devoted to the description of soliton waves propagating in a dispersive nonlinear medium. When the transmitted pulse has a particular shape, the dispersion effect of an optical fiber is counterbalanced by the nonlinearity of the core glass of the fiber. Such pulses are called solitons and can travel for an extraordinary distance while still preserving the pulse shape. Solitons are particularly attractive for long-haul transmission through optical fibers such as transoceanic fiber-optic submarine communication.

15.1 NONLINEAR (KERR) MEDIUM

The fields have to satisfy Maxwell's equations,

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}$$
$$\nabla \cdot \mathbf{D} = \rho$$
$$\nabla \cdot \mathbf{B} = 0 \tag{15.1}$$

regardless of whether the fields are in a linear or nonlinear medium.

As explained in Section 4.1, **D** in a linear medium is

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \tag{15.2}$$

When the **E** field is not intense, the polarization **P** is linearly proportional to the applied field **E** and

$$\mathbf{P} = \epsilon_0 \chi^{(1)} \mathbf{E} \tag{15.3}$$

where $\chi^{(1)}$ is the linear portion of the susceptibility. Combining the above two equations gives

$$\mathbf{D} = \epsilon_0 \epsilon_r \mathbf{E}, \quad \text{where} \quad \epsilon_r = 1 + \chi^{(1)} \tag{15.4}$$

When the **E** field becomes intense, it is necessary to include the higher order terms of **E** in order to represent **P** adequately.

$$\mathbf{P} = \epsilon_0 [\chi^{(1)} \cdot \mathbf{E} + \chi^{(2)} : \mathbf{EE} + \chi^{(3)} \vdots \mathbf{EEE} \cdots] \tag{15.5}$$

where $\chi^{(1)}$ is the linear, $\chi^{(2)}$ is the second order nonlinear, and $\chi^{(3)}$ is the third order nonlinear electric susceptibility.

Practical applications of nonlinear phenomena to optical devices are categorized by which order of $\chi^{(i)}$ of the substance is predominant [1].

Substances such as KDP (KH_2PO_4), ADP ($NH_4H_2PO_4$), and BNN ($Ba_2NaNb_5O_{15}$), with a large $\chi^{(2)}$ are used to generate the beat of two frequencies. If E_1 and E_2 are at the frequencies ω_1 and ω_2 and are expressed as $E_1 \cos \omega_1 t$ and $E_2 \cos \omega_2 t$, the presence of $\chi^{(2)}$ generates $\chi^{(2)} E_1 E_2 \cos(\omega_1 - \omega_2)t$ and $\chi^{(2)} E_1 E_2 \cos(\omega_1 + \omega_2)t$, which contains the beat frequencies of $\omega_1 - \omega_2$ and $\omega_1 + \omega_2$. In particular, when $\omega_1 = \omega_2$, an output signal with twice the frequency of the input is generated. This is known as *second harmonic generation* (SHG). The SHG devices extend the wavelength range of semiconductor lasers. Wavelengths that the semiconductor laser diode cannot generate itself can be supplemented by the wavelengths generated by SHG devices.

Another major application of $\chi^{(2)}$ is in devices based on the Pockels effect, also known as the electrooptic effect. Many kinds of devices such as phase modulators, amplitude modulators, directional couplers, and polarization controllers are based on the Pockels effect and were dealt with extensively in Chapters 4 and 5. In these Pockels effect devices, of the two fields E_1 and E_2, E_1 is the external electric field ε from dc up to microwave frequencies, and E_2 is the signal light beam. The second order nonliearity $\chi^{(2)}(\varepsilon + E_2 \cos \omega_2 t)^2$ generates a term $\varepsilon \chi^{(2)} E_2 \cos \omega_2 t$ and the emergent light is modulated by ε.

The nonlinearity that this chapter deals with is primarily the third order susceptibility $\chi^{(3)}$. Of the three fields E_1, E_2, and E_3, we are especially interested in the case when $E_1 = E_2^*$ and $|E_1|^2$ is the intensity of the controlling light used to change the index of refraction of the medium, and E_3 is the probe or signal light. When $E_1 = E_2^* = E_3$, the signal beam itself is used to change the index of refraction of the nonlinear medium.

The nonlinear effect based on $\chi^{(3)}$ is known as the *Kerr effect*. Chapter 8 has already described the applications of the Kerr effect to four-wave mixing and optical phase conjugation. This chapter devotes itself to the Kerr nonlinear effect, which is the basis of the generation of soliton waves in the fiber. A special advantage of the Kerr nonlinear effect is also the realization of optical devices controlled by optics. Devices controlled by photons are ultimately expected to be much faster than those controlled by electrons.

Typical devices based on the Kerr nonlinear effect are shown in Fig. 15.1.

Figures 15.1a and 15.1b are two basic configurations for which closed-form solutions are obtainable and are explained in detail in this chapter. The geometry in Fig. 15.1a consists of a nonlinear half-space combined with a linear half-space. A guided nonlinear boundary wave becomes a unique carrier of light energy along the nonlinear boundary and plays an important role in adjusting the amount of coupling of light energy into an adjacent element.

The geometry of Fig. 15.1b is that of an optical guide whose linear core layer is sandwiched by nonlinear cladding layers. It will be shown that there are modes that have a bulge in the nonlinear cladding region. These bulges are responsible for the function of the optically controlled directional couplers such as those illustrated in Figs 15.1c and 15.1d [2].

For the nonlinear directional couplers in Figs. 15.1c and 15.1d, the incident light energy at port 1 is switched between ports 3 and 4, depending on the incident light intensity at port 1. Unlike the electrooptic switch, no external electric field is required.

Figure 15.1e is a diagram of a Mach–Zehnder interferometer. The incident light is divided into linear and nonlinear arms. The relative phase of the two arms is determined by the intensity of the light. When the recombined light from the two arms is out of phase, the interferometer output is null; when the recombined light is in phase, the output reaches a maximum. Such an interferometer can be used as either an optically controlled on–off switch or an amplitude modulator.

Figure 15.1f is an illustration of a nonlinear distributed feedback grating or DFB grating. The grating is formed by a nonlinear medium. The refractive index modulation is a function of the light intensity, and the passband characteristics of the grating are controlled by the light intensity. If the DFB grating is incorporated into a semiconductor laser diode, a chirped frequency light pulse is generated whose carrier frequency is swept with the intensity profile of the light.

Figure 15.1g shows a bistable device made of multi-quantium-well (MQW) layers (see Section 14.12). The MQW layer displays a strong nonlinearity. With an increase in the light intensity, the light absorption of the substance is decreased, but beyond a certain intensity it becomes saturated and the absorption does not decrease any further.

The MQW layer is sandwiched between dielectric coated mirrors forming a Fabry–Pérot resonator (see Section 3.2). The output light intensity from the Fabry–Pérot cavity is plotted as a function of the input light intensity in the P_{out}–P_{in} graph. The cavity length is set so that the wavelength of the incident light is near resonance. When the intensity of the input light is low, as indicated in the figure by a–b, the absorption is too high for the cavity to resonate. When the input power is raised to point b, the absorption becomes small enough for the cavity to resonate. At resonance, the light intensity in the cavity significantly increases and the light intensity goes beyond the saturation of the nonlinearity, which is represented by point d.

What happens if the light intensity is reduced from point d? The output light intensity does not trace back the original route d–c–b, but takes the route d–e–a instead. One

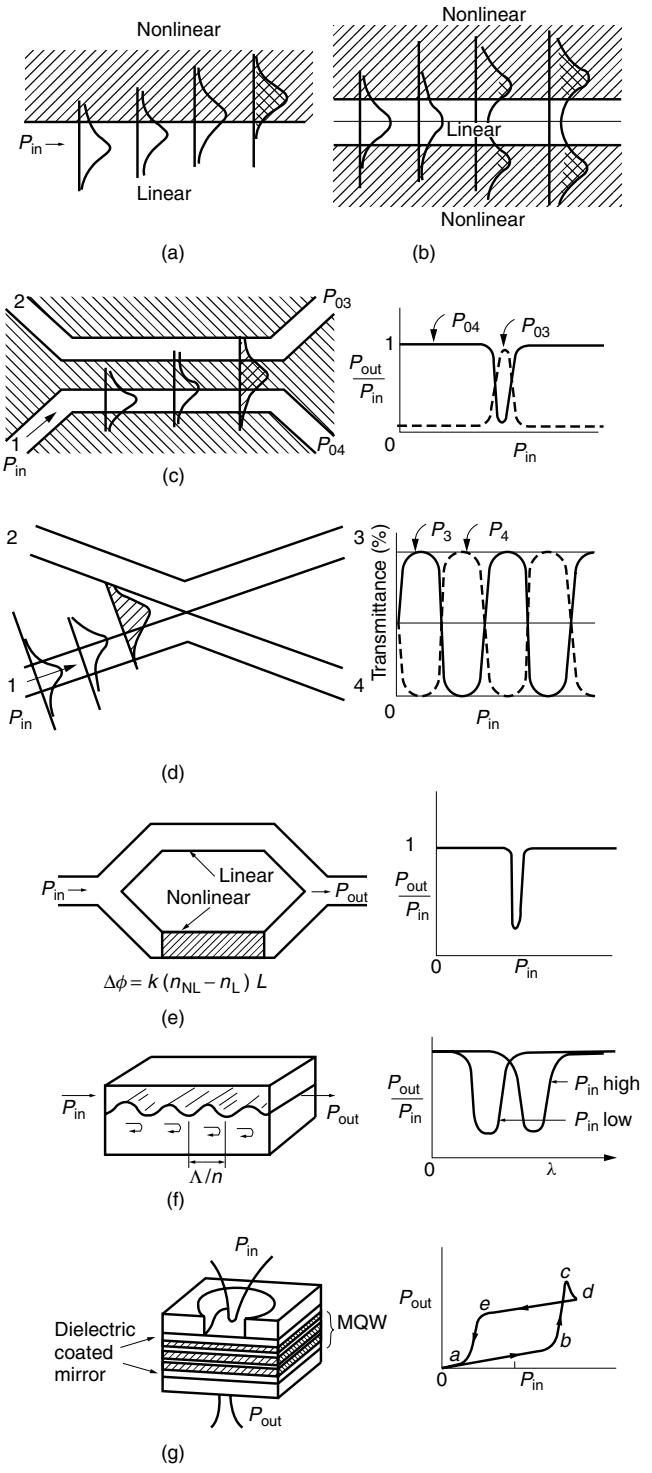


Figure 15.1 Devices based on Kerr nonlinear optics. (a) Guided nonlinear boundary wave. (b) Nonlinear guide. (c) Nonlinear directional coupler. (d) Nonlinear X junction. (e) Mach–Zehnder interferometer. (f) Nonlinear distributed feedback grating. (g) Optical bistability.

reason for the different route down is that the Fabry–Pérot cavity is at resonance, which means the light intensity is high and the absorption is kept low in the cavity. Another reason is that point d is well into the saturation region, and a reduction of the incident light does not significantly increase the absorption in the cavity. Only when the incident light intensity becomes really low does the output reduce to point a. This trace of output versus input is called the hysteresis curve of the nonlinear cavity.

The hysteresis curve shows two possible outputs for a given amount of input light. The amount of output light depends on whether or not the input exceeded the value at d just before this measurement. If the input light has exceeded d before this measurement, the amount of output light is large while if the input light has not exceeded d before this measurement, the amount of output light is small. In other words, by measuring the amount of output light for a fixed input power, the previous value of the input light can be decoded. A device with this ability to memorize the previous value of the input light is potentially useful as a memory element of a "light computer."

15.2 SOLUTIONS IN THE UNBOUNDED KERR NONLINEAR MEDIUM

For simplicity, the medium will be treated as isotropic and instead of a vector, a scalar quantity is used to express the magnitude of the polarization in the Kerr medium:

$$P = \epsilon_0[\chi^{(1)}E + \chi^{(3)}\{E(\omega, t)\}^3] \tag{15.6}$$

In a Kerr nonlinear medium, $\epsilon_0\epsilon_r$ in Eq. (15.4) should be replaced by

$$\epsilon_0\epsilon_{r\mathrm{NL}} = \epsilon_0(\epsilon_r + \alpha|E|^2) \tag{15.7}$$

where $\alpha = \frac{3}{4}\chi^{(3)}$ and α is the *nonlinear Kerr constant*.
The nonlinear wave equation obtainable from the first two Maxwell's equations in Eq. (15.1) becomes (see Section 8.10.1)

$$\nabla^2 E + \mu\epsilon_0\epsilon_{\mathrm{rNL}}\frac{\partial^2 E}{\partial t^2} = 0 \tag{15.8}$$

This differential equation will be solved using the trial solution

$$E(x, z, t) = E_y(x)e^{j\beta z - j\omega t} + \text{c.c.} \tag{15.9}$$

where c.c. stands for complex conjugate, and y is the direction of polarization.

The direction of propagation is in the z direction. There is no variation in the y direction, which is the direction transverse to the direction of propagation. Putting the trial solution Eq. (15.9) into Eq. (15.8) gives

$$\frac{\partial^2 E_y(x)}{\partial x^2} - \beta^2 E_y(x) + k^2[\epsilon_r + \alpha E_y^2(x)]E_y(x) = 0 \tag{15.10}$$

where $k^2 = \omega^2\mu\epsilon_0$. The amplitude function $E_y(x)$ is assumed to be real [3]. The subscripts y and L will be suppressed. Throughout the analysis, a Kerr medium with positive α will be assumed.

Let us see how the absolute value square symbol $|\ |^2$ shows up in Eq. (15.7). Let us focus our attention on the second term of Eq. (15.6).

$$P_{NL} = \epsilon_0 \chi^{(3)} \{E(\omega, t)\}^3 \tag{a}$$

The method of analytic signal cannot be applied to a nonlinear case (see Appendix B in Vol. II).

Expressions such as

$$E(\omega, t) = \mathrm{Re}\{E(\omega) e^{-j\omega t}\} \tag{b}$$

or

$$E(\omega, t) = \tfrac{1}{2}[E(\omega) e^{-j\omega t} + E^*(\omega) e^{j\omega t}] \tag{c}$$

instead of the analytic signal

$$E(\omega, t) = E(\omega) e^{-j\omega t} \tag{d}$$

have to be used to represent the signal.

Putting Eq. (c) into (a) gives

$$P_{NL} = \epsilon_0 \chi^{(3)} \tfrac{1}{8}[E(\omega) e^{-j\omega t} + E^*(\omega) e^{j\omega t}]^3 \tag{e}$$

The cubic multiplication generates many a term.

Only the terms with frequency ω will be kept.

$$P_{NL}(\omega, t) = \epsilon_0 \chi^{(3)} [|E(\omega)|^2 E(\omega) e^{-j\omega t} + |E(\omega)|^2 E^*(\omega) e^{j\omega t}] \tag{f}$$

$$P_{NL}(\omega, t) = \mathrm{Re}\{P_{NL}(\omega) e^{-j\omega t}\} \tag{g}$$

where

$$P_{NL}(\omega) = \epsilon_0 \tfrac{3}{4} \chi^{(3)} |E(\omega)|^2 E(\omega) \tag{h}$$

The $|\ |^2$ term arose from the product operation of Eq. (f). α in Eq. (15.7) is thus,

$$\alpha = \tfrac{3}{4} \chi^{(3)} \tag{i}$$

When, however, the assumption that $E(\omega)$ is a real number is made, the $|\ |^2$ symbol disappears.

Before attempting to solve the nonlinear differential equation (15.10), a short summary of solving the linear differential equation will be presented for the sake of comparison. With $\alpha = 0$, Eq. (15.10) becomes

$$\frac{1}{E_1(x)} \frac{d^2 E_1(x)}{dx^2} + (k^2 \epsilon_r - \beta^2) = 0 \tag{15.11}$$

where the subscript 1 will be used for the solution in the linear medium. The method of separation of variables is used to solve Eq. (15.11). Since the second term of

In order to express the nonlinearity, the majority of soliton literature uses the refractive index expression,

$$n_{\mathrm{NL}} = n_{\mathrm{L}} + n_2 I \tag{a}$$

where n_2 is the *nonlinear-index coefficient* and I is the light power intensity, whereas we use that of the dielectric constant

$$\epsilon_{r\mathrm{NL}} = \epsilon_{r\mathrm{L}} + \alpha |E|^2 \tag{b}$$

where the subscripts L and NL mean linear and nonlinear, respectively.

In order to solve Maxwell's equations, $\epsilon_{r\mathrm{NL}}$ is needed. When using convention (a), $\epsilon_{r\mathrm{NL}}$ is obtained as

$$\begin{aligned} \epsilon_{r\mathrm{NL}} &= (n_{\mathrm{L}} + n_2 I)^2 \\ &= n_{\mathrm{L}}^2 + 2 n_{\mathrm{L}} n_2 I + (n_2 I)^2 \end{aligned} \tag{c}$$

and is approximated as

$$\epsilon_{r\mathrm{NL}} \doteqdot n_{\mathrm{L}}^2 + 2 n_{\mathrm{L}} n_2 I \tag{d}$$

Under normal circumstances, the term $n_2^2 I^2$ is much smaller than the other terms and the approximation is justifiable. This chapter uses convention (b), which does not need this approximation.

In order to convert the expressions in this book to their counterparts in convention (a), simply replace $\alpha |E|^2$ in this chapter by $2 n_{\mathrm{L}} n_2 I$, and replace ϵ_r by n_{L}^2.

Eq. (15.11) is a constant, the first term also has to be a constant. A positive constant should be just as good as a negative constant. The nature of the solutions, however, is significantly influenced by the signs of the constant.

With a positive constant γ^2,

$$\frac{1}{E_1(x)} \frac{d^2 E_1(x)}{dx^2} = \gamma_1^2 \tag{15.12a}$$

with

$$-\gamma_1^2 = k^2 \epsilon_r - \beta^2 \tag{15.12b}$$

and the solution is an exponential function of a real number or hyperbolic function:

$$E_1(x) = C' e^{-\gamma x} + D e^{\gamma x} \tag{15.13}$$

When γ^2 becomes a negative number, the constant is denoted as $-K^2$,

$$\begin{aligned} \frac{1}{E_1(x)} \frac{d^2 E_1(x)}{dx^2} &= -K_1^2 \\ K_1^2 &= k^2 \epsilon_r - \beta^2 \end{aligned} \tag{15.14}$$

and the solution becomes an exponential function of an imaginary number or trigonometric function:

$$E_1(x) = A\cos Kx + B\sin Kx \tag{15.15}$$

The value of $\gamma = 0$ determines the borderline between two significantly different types of solutions, Eqs. (15.13) and (15.15).

Now, we return to solving the nonlinear differential equation (15.10). Equation (15.10) can be rewritten as

$$\frac{1}{E(x)}\left(\frac{d^2E(x)}{dx^2} + k^2\alpha E^3(x)\right) + (k^2\epsilon_{r0} - \beta^2) = 0 \tag{15.16}$$

The second term being constant, the first term has to be a constant. Let's say the constant is γ_0^2, Then Eq. (15.16) becomes

$$\frac{d^2E(x)}{dx^2} - \gamma_0^2 E(x) + k^2\alpha E^3(x) = 0 \tag{15.17a}$$

with

$$-\gamma_0^2 = k^2\epsilon_{r0} - \beta^2 \tag{15.17b}$$

The subscript 0 will be used in the nonlinear medium.

The same nonlinear differential equation, Eq. (15.17a), will be solved by three different methods.

15.2.1 Method by a Trial Solution

The second derivative of the trial function

$$E(x) = \frac{A}{\cosh(\pm\gamma_0 x + \psi_0)} \tag{15.18}$$

is

$$\frac{d^2E(x)}{dx^2} = \gamma_0^2 E(x) - \frac{2\gamma_0^2}{A^2}E^3(x) \tag{15.19}$$

and has all the functional forms that appear in Eq. (15.17a). There is a good chance that Eq. (15.18) satisfies Eq. (15.17a) with some restrictions on the values of the constants.

Insertion of Eq. (15.18) into (15.17a) brings out the necessary condition imposed on the constants as

$$\gamma_0^2 = \tfrac{1}{2}k^2\alpha A^2 \tag{15.20}$$

This means that as long as Eq. (15.20) is accepted as an additional condition, Eq. (15.18) can be treated as the solution of Eq. (15.17a). This additional condition distinguishes the nonlinear case from the linear case. It is worth noting that the solutions in the nonlinear case depend on the intensity of the light by way of γ_0.

A particular feature of this solution given by Eq. (15.18) is that it has only one peak located at $x = \mp\psi_0/\gamma_0$. Later, the location of this sole peak will become the key to classifying modes in the nonlinear guided structure.

It is interesting to note that in the linear case, as soon as $(1/E)d^2E/dx^2$ in Eq. (15.11) changes from a negative value to a positive value, the solution changes from a sinusoidal to a hyperbolic function, and the borderline between the two different types of solution is

$$\frac{1}{E}\frac{d^2E}{dx^2} = 0$$

Let us see if there is a similar borderline in the nonlinear case. The value of γ_0 that makes

$$\frac{1}{E(x)}\frac{d^2E(x)}{dx^2} = 0$$

is, from Eq. (15.17a)

$$\gamma_0^2 = k^2\alpha E^2(x)$$

To be more precise, Eq. (15.20) is the borderline. In short, in the linear case the borderline is $\gamma_1^2 = 0$, but with the nonlinear case the borderline occurs for a particular value of γ_0^2 that is greater than zero.

Regarding sign, the minus sign in Eq. (15.18) will be selected so that the peak of Eq. (15.18) appears in the positive x space for positive values of Ψ_0. γ_0 is a positive number with a positive nonlinear Kerr coefficient. Equation (15.18) can be expressed in terms of the field E_0 at the boundary. For $x = 0$, Eq. (15.18) becomes

$$A = E_0 \cosh \Psi_0 \tag{15.21}$$

and

$$E_0(x) = \frac{E_0 \cosh \Psi_0}{\cosh(-\gamma_0 x + \Psi_0)} \tag{15.22}$$

Before going into detail, we will examine whether or not Eq. (15.22) approaches the expression for the linear case as α is reduced. Inserting Eq. (15.21) into Eq. (15.20) gives

$$\cosh^2 \Psi_0 = \frac{2\gamma_0^2}{k^2\alpha E_0^2} \tag{15.23}$$

As α in Eq. (15.23) approaches zero, Ψ_0 becomes infinite:

$$\Psi_0 = \pm\infty \quad \text{for } \alpha \to 0 \tag{15.24}$$

For a large magnitude of $|\Psi_0|$, Eq. (15.22) can be expanded into exponential functions and

$$E_0(x) \doteqdot \begin{cases} E_0 e^{-\gamma_0 x} & \text{for } \Psi_0 \to -\infty \\ E_0 e^{\gamma_0 x} & \text{for } \Psi_0 \to \infty \end{cases} \tag{15.25}$$

Thus, the nonlinear solution approaches the linear solution given by Eq. (15.13) as $\alpha \to 0$.

Next, the same nonlinear differential equation will be solved using the method by integration.

15.2.2 Method by Integration

Both sides of Eq. (15.17a) are first multiplied by $2\,dE(x)/dx$ and then integrated with respect to x. The result is

$$\left(\frac{dE(x)}{dx}\right)^2 - \gamma_0^2 E^2(x) + \Lambda E^4(x) = C \qquad \Lambda = \tfrac{1}{2}k^2\alpha \tag{15.26}$$

Let us determine the integration constant C. Since Eq. (15.26) has to hold regardless of the value of x, the location of the boundary condition is chosen to be at the location of the peak of $E(x)$. The condition is

$$\frac{dE}{dx} = 0 \tag{15.27}$$

Let the desired type of solution $E(x)$ have a peak value of A at the extremum. At such a location, Eq. (15.26) becomes

$$C = A^2\left(-\gamma_0^2 + \alpha\frac{k^2}{2}A^2\right) \tag{15.28}$$

If the expected solution is a hyperbolic function like Eq. (15.18), the condition of Eq. (15.20) can be used and Eq. (15.28) becomes

$$C = 0 \tag{15.29}$$

The more general case of $C \neq 0$ will be explained in the next method of solution. With $C = 0$, the differential equation (15.26) becomes

$$\int dx = \frac{1}{\sqrt{\Lambda}}\int \frac{dE}{E\sqrt{\gamma_0^2/\Lambda - E^2}} \tag{15.30}$$

Now, using the following integral from a table of integrals,

$$\int \frac{dx}{x\sqrt{a^2 - x^2}} = \frac{1}{a}\operatorname{sech}^{-1}\left|\frac{x}{a}\right|$$

the integral of Eq. (15.30) becomes

$$x - x_0 = \frac{1}{\gamma_0}\operatorname{sech}^{-1}\left(\frac{\sqrt{\Lambda}}{\gamma_0}E\right)$$

The value of $\sqrt{\Lambda}/\gamma_0$ can be rewritten using Eqs. (15.26), (15.28), and (15.29) as

$$\frac{\gamma_0}{\sqrt{\Lambda}} = A$$

Thus, the final result is

$$E(x) = A\ \operatorname{sech}\left[\gamma_0(x - x_0)\right] \tag{15.31}$$

which is identical to the previous solution, Eq. (15.18).

With the linear differential equation, there were two types of solutions: sinusoidal and hyperbolic. The same is true for the present nonlinear differential equation. So far, we have dealt with only the sech u type of solution, which was obtained by imposing $C = 0$. The other solution, which is $\mathrm{cn}(u|m)$, is obtained by removing this condition of $C = 0$, as will be explained in the next section.

15.2.3 Method Using Jacobi's Elliptic Functions

With the removal of the condition $C = 0$, the differential equation (15.26) becomes

$$\int_{x_0}^{x} dx = \frac{1}{\sqrt{\Lambda}} \int_{E}^{b} \frac{dE}{\sqrt{\dfrac{C}{\Lambda} - (E^2(x) - \dfrac{\gamma_0^2}{\Lambda})E^2(x)}} \tag{15.32}$$

The integral of Eq. (15.32) is a form of the Jacobian elliptic function with modulus m, and is listed in the table of integrals as

$$\int_{t}^{b} \frac{dt}{\sqrt{(t^2 + a^2)(b^2 - t^2)}} = \frac{1}{\sqrt{a^2 + b^2}} \mathrm{cn}^{-1}\left(\frac{t}{b}|m\right) \quad \text{with } m = \frac{b^2}{a^2 + b^2} \tag{15.33}$$

The curves of the cn function are plotted in Fig. 15.2. The function looks like a pinched cosine function and it is periodic except for $m = 1$. When the modulus is zero it is indeed a cosine function; but when the modulus is unity, it becomes the sech function [4]:

$$\mathrm{cn}(u|m) = \begin{cases} \cos u & \text{for } m = 0 \\ \operatorname{sech} u & \text{for } m = 1 \end{cases} \tag{15.34}$$

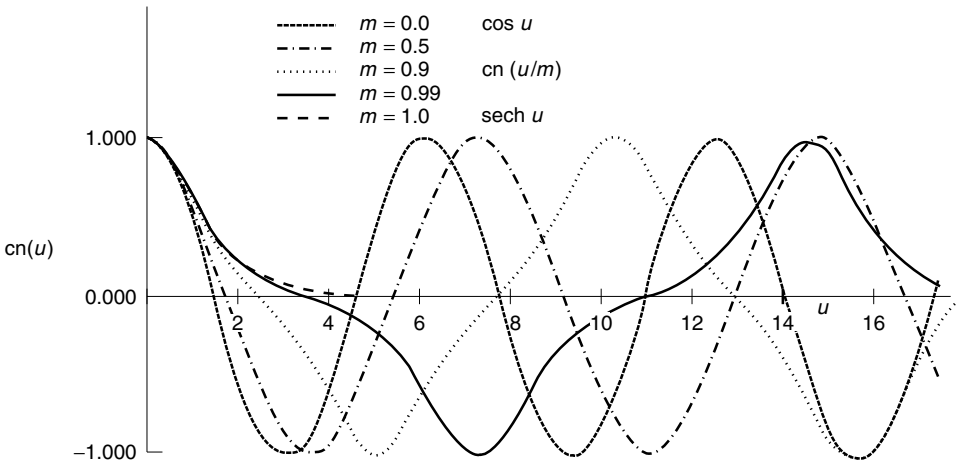


Figure 15.2 Jacobian elliptic function cn as the modulus m is varied from 0 to 1. (Calculations courtesy of H. Anis).

Comparing the coefficients of Eq. (15.32) with those of Eq. (15.33), Eq. (15.32) will be rewritten in terms of the inverse cn function. The coefficients have to satisfy

$$a^2 - b^2 = \frac{-\gamma_0^2}{\Lambda} \tag{15.35a}$$

$$a^2 b^2 = \frac{C}{\Lambda} \tag{15.35b}$$

and the values of a^2 and b^2 will be solved from the simultaneous equations (15.35a) and (15.35b) with the help of the relationship

$$(a^2 + b^2)^2 = (a^2 - b^2)^2 + 4a^2 b^2$$

Using Eqs. (15.35a) and (15.35b) in the right-hand side of the above relationship gives

$$\begin{aligned} a^2 + b^2 &= \sqrt{\left(\frac{-\gamma_0^2}{\Lambda}\right)^2 + \frac{4C}{\Lambda}} \\ &= \frac{q^2}{\Lambda} \end{aligned} \tag{15.36}$$

where

$$q^2 = \sqrt{\gamma_0^4 + 4C\Lambda} \tag{15.37}$$

From Eqs. (15.35a) and (15.36), the coefficients a and b are written in terms of q as

$$a^2 = \frac{-\gamma_0^2 + q^2}{2\Lambda} \tag{15.38}$$

$$b^2 = \frac{\gamma_0^2 + q^2}{2\Lambda} \tag{15.39}$$

and hence from Eqs. (15.38) and (15.39), m and $\sqrt{a^2 + b^2}$ in Eq. (15.33) become

$$m = \frac{\gamma_0^2 + q^2}{2q^2} \tag{15.40}$$

and

$$\sqrt{a^2 + b^2} = \frac{q}{\sqrt{\Lambda}} \tag{15.41}$$

Now, Eq. (15.32) can be rewritten using Eqs. (15.33), (15.40), and (15.41) as

$$x - x_0 = \frac{1}{q} \mathrm{cn}^{-1}\left(\frac{E}{b}\middle| m\right) \tag{15.42}$$

or finally,

$$E = b\ \mathrm{cn}(q(x - x_0)|m) \tag{15.43}$$

Equation (15.43) is the solution in the unbounded Kerr nonlinear medium.

We have treated the positive region of γ_0^2. The same analysis can be extended to the negative region of γ_0^2. Exactly the same formalism can be applied if γ_0^2 is replaced by $-K^2$. Equation (15.40) is changed to

$$m = \frac{-K^2 + q^2}{2q^2} \tag{15.44}$$

where

$$q^2 = \sqrt{K^4 + 4C\Lambda} \tag{15.45}$$

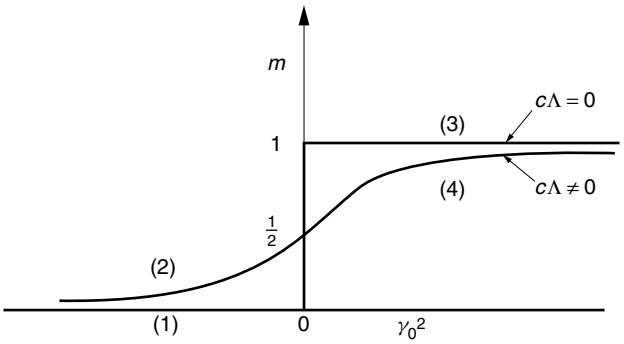
Inserting q^2 in Eq. (15.37) into Eqs. (15.40) and (15.44), the values of m for negative infinity to positive infinity of γ_0^2 are obtained.

The special case where $C\Lambda$ in q is zero is considered. This happens when either the medium is linear from Eq. (15.26), or the medium is nonlinear but the conditions of Eqs. (15.28) and (15.29) are satisfied. In the latter case, from Eqs. (15.37) and (15.40), $m = 0$ for the entire range of negative γ_0^2 (or positive K^2) and the solution is $\cos u$. Likewise, $m = 1$ for the entire range of positive γ_0^2 and the solution is sech u.

The general case of $C\Lambda \neq 0$ occurs when the medium is nonlinear and the conditions of Eqs. (15.28) and (15.29) are removed. In this case, $0 < m < \frac{1}{2}$ for negative γ_0^2 and $\frac{1}{2} < m < 1$ for positive γ_0^2 and the solution is cn $(u|m)$ for the entire region of γ_0^2. The above results are summarized in Table 15.1.

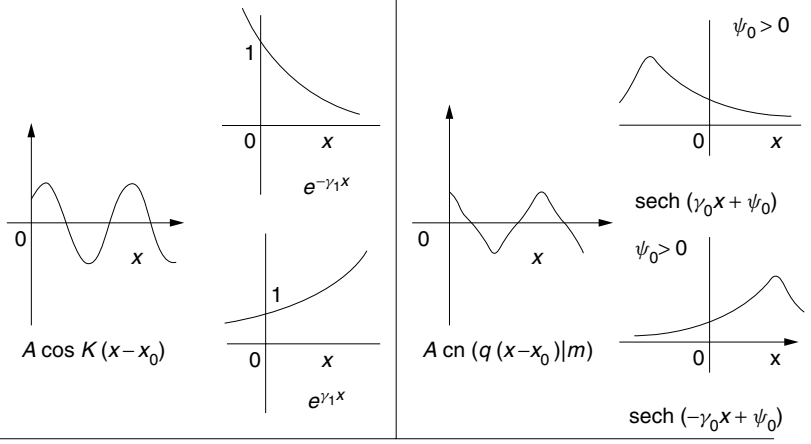
The various types of solutions obtained thus far are summarized in Table 15.2.

Table 15.1 Value of m versus γ_0^2 and solutions for various values of $c\Lambda$



Negative γ_0^2			Positive γ_0^2			
c,Λ	$c\Lambda = 0$	$c\Lambda \neq 0$	$c = 0$ $\Lambda = 0$	$c \neq 0$ $\Lambda = 0$	$c = 0$ $\Lambda \neq 0$	$c\Lambda \neq 0$
m	0	$0 < m < \frac{1}{2}$	1	1	1	$\frac{1}{2} < m < 1$
Solution	$\cos u$	cn $(u\|m)$	$e^{\pm u}$	$\sinh u \quad c > 0$ $\cosh u \quad c < 0$	sech u	cn $(u\|m)$
Region	(1)	(2)	(3)	(3)	(3)	(4)

Table 15.2 Comparison of solutions of linear and nonlinear differential equations

	Linear		Nonlinear	
Differential equation	$\frac{d^2E}{dx^2} + (\epsilon_r k^2 - \beta^2)E = 0$		$\frac{\partial^2 E}{\partial x^2} + [\epsilon_r k^2 - \beta^2]E + 2k^2 \alpha E^3 = 0$	
Type	(I)	(II)	(III)	(IV)
Constant	$K^2 = \epsilon_r k^2 - \beta^2$	$-\gamma_1^2 = \epsilon_r k^2 - \beta^2$	$m \neq 1$	$m = 1$
Shape	Oscillatory	Monotonic	Oscillatory	Monotonic
Graphics				

15.3 GUIDED NONLINEAR BOUNDARY WAVE

When a nonlinear half-space (top of Fig. 15.3) is bordered by a linear half-space (bottom of Fig. 15.3), there exists a so-called guided nonlinear boundary wave [5]. Such a guided wave occurs only on the nonlinear medium boundary. It does not exist if the two bordering media are both linear. The dielectric constants of the two media are

$$\begin{aligned} \epsilon_{\mathrm{NL}} &= \epsilon_{r0} + \alpha E^2(x), \quad x > 0 \\ \epsilon_{\mathrm{L}} &= \epsilon_{r1}, \quad x < 0 \end{aligned} \tag{15.46}$$

and an additional condition is imposed that the linear part of the nonlinear dielectric constant ϵ_{r0} in the nonlinear medium is smaller than the dielectric constant ϵ_{r1} of the linear medium:

$$\epsilon_{r1} > \epsilon_{r0} \tag{15.47}$$

As mentioned earlier, the subscript 0 designates a quantity associated with the top half nonlinear space and the subscript 1 designates a quantity associated with the bottom half linear space. Such a wave as expressed by Eq. (15.9) is used. The wave propagates in the z direction and is polarized in the y direction or in the direction transverse to the direction of propagation. Such a wave is called a *transverse electric wave* or TE wave.

The guided wave that can exist along the boundary between linear and nonlinear media will be investigated. The field in the nonlinear medium has to match with that in the linear medium on the boundary. Special attention, however, should be

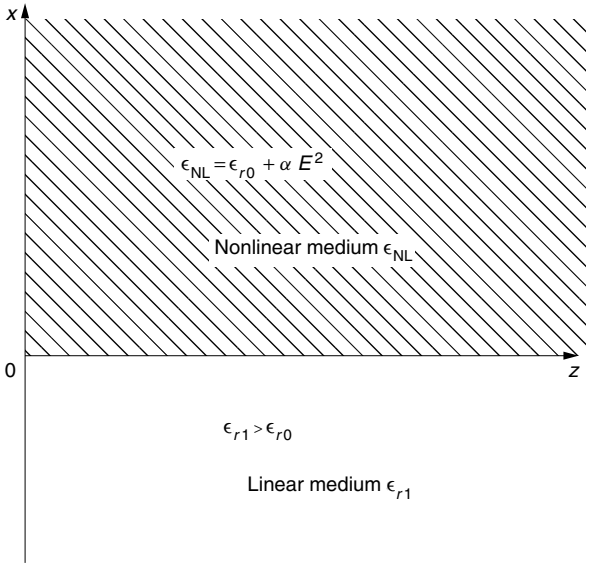


Figure 15.3 Combination of linear and nonlinear half-spaces.

focused on the definition of *guided wave*. The guided wave is defined as a wave whose intensity decays after having at most one bulge over distance and surely approaches zero away from the boundary for *both* the positive x direction and negative x direction. The evanscent wave that appears in the boundary between two linear media does not qualify for our definition of guided boundary wave because the wave decays only in one direction either for $x > 0$ or $x < 0$.

We treat only the type of guided wave that approaches zero for both directions.

First, consider the boundary condition graphically. In choosing the guided wave along the boundary shown in Fig. 15.3, two field distributions have to be selected from Table 15.2. The distributions must match smoothly across the boundary, and, also, they must decay in both large positive and large negative x directions. According to the definition of the guided wave, none of the waves with oscillating functions of x are candidates. It should be noted that $A\text{cn}(q(x - x_0)|m)$ with $m \neq 1$ does not qualify either, because it does not decay with x. Referring to Table 15.2, the candidates are both curves in Column(II) in the linear region, and both curves in Column(IV) in the nonlinear region. These curves, however, have to be smoothly connected at the boundary. Possible ways of connecting the two curves are shown in Fig. 15.4. The combination shown in Fig. 15.4a is smoothly connected, whereas that in Fig. 15.4b is not smoothly connected; hence the connection in Fig. 15.4a is the correct choice.

In other words, the field on the linear side of the boundary has to decay in the negative direction and in order to connect with this field smoothly at the boundary, the field in the nonlinear region has to increase first and then decrease in the positive direction. Only the field that has a bulge in the nonlinear region or in $x > 0$ can meet this boundary condition.

The field in the top half-space is, from Eqs. (15.21) and (15.22),

$$E_0(x) = \frac{A}{\cosh(-\gamma_0 x + \Psi_0)} \tag{15.48}$$

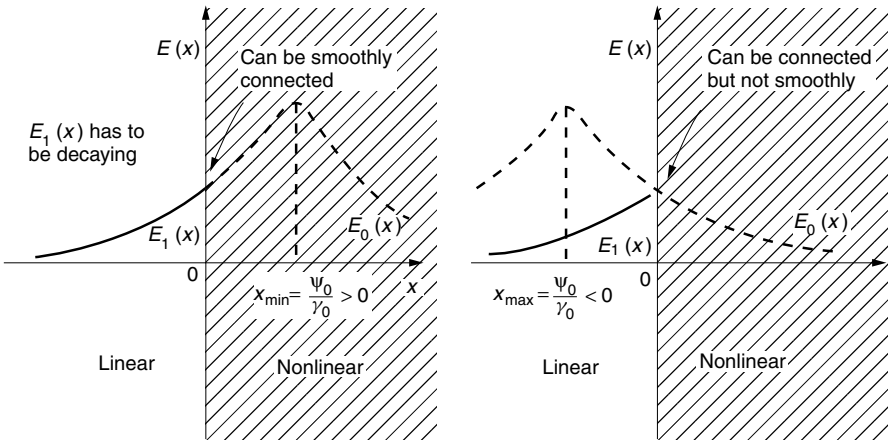


Figure 15.4 Possible combinations of solutions. Case (a) satisfies the boundary condition but case (b) does not.

and that in the bottom half-space is

$$E_1(x) = De^{\gamma_1 x} \tag{15.49}$$

where γ_1 is a positive number.

Next, consider the boundary condition analytically. From the continuity of the field on the boundary at $x = 0$,

$$E_0(0) = E_1(0)$$

which leads to

$$\frac{A}{\cosh \Psi_0} = D \tag{15.50}$$

The second necessary boundary condition is the continuity of the field derivatives at $x = 0$. Applying this condition to Eqs. (15.48) and (15.49) gives

$$A\gamma_0 \tanh \Psi_0 \text{ sech } \Psi_0 = D\gamma_1 \tag{15.51}$$

Dividing Eq. (15.51) by (15.50) gives

$$\gamma_1 = \gamma_0 \tanh \Psi_0 \tag{15.52}$$

Equation (15.52) is the characteristic equation. Since both γ_0 (see Section 15.2.1) and γ_1 are positive, as mentioned earlier,

$$\Psi_0 > 0 \tag{15.53}$$

and from Eqs. (15.48) and (15.53), the bulge is present in the region $x > 0$.

The characteristic equation results from the continuity of the field in the x direction. On the other hand, the continuity of the field in the z direction results in the synchronization of the phase constant β's in the linear and nonlinear media.

From Eqs. (15.12b) and (15.17b),

$$\gamma_0^2 - \gamma_1^2 = \nu^2 \tag{15.54}$$

where

$$\nu^2 = k^2(\epsilon_1 - \epsilon_0) \tag{15.55}$$

and where $\epsilon_1 > \epsilon_0$ as specified at the beginning. The subscript r in the relative dielectric constant will be suppressed. The three equations needed to determine γ_0, γ_1 and Ψ_0 for a given A have been derived and are repeated here for convenience:

$$\gamma_0^2 = \tfrac{1}{2}k^2\alpha A^2 \tag{15.20}$$

$$\gamma_1 = \gamma_0 \tanh \Psi_0 \tag{15.52}$$

$$\gamma_0^2 - \gamma_1^2 = \nu^2 \tag{15.54}$$

Plots of Eqs. (15.20) and (15.54) are shown in Fig. 15.5. The intersections of Eqs. (15.20) and (15.24) first provide γ_0 and γ_1, and then ψ_0 is obtained from Eq. (15.52). Only the intersections in the first quadrant are physically valid because if γ_1 is negative the field becomes infinite at $x = -\infty$.

The procedures for obtaining the analytical solutions are now described. As soon as A is given, γ_0 is found from Eq. (15.20), and then γ_1 is calculated using Eq. (15.54):

$$\gamma_1 = \sqrt{\tfrac{1}{2}k^2\alpha A^2 - \nu^2} \tag{15.56}$$

So that γ_1 will be real, there is a minimum value of A given by

$$A_{\min}^2 = \frac{2(\epsilon_1 - \epsilon_0)}{\alpha} \tag{15.57}$$

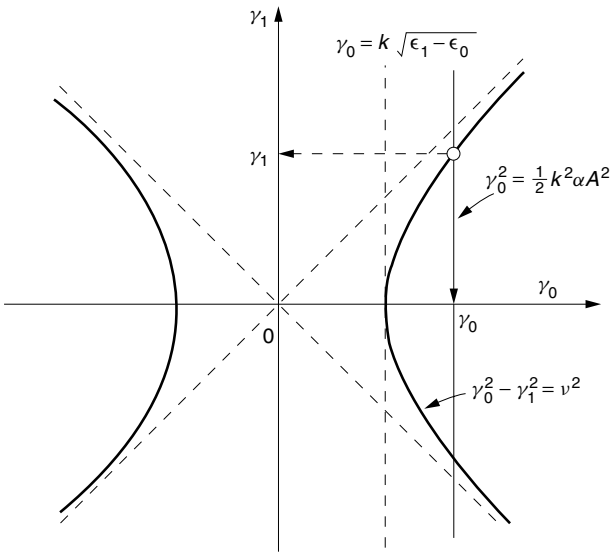


Figure 15.5 Plots of the characteristic equation of the guided nonlinear boundary wave.

where Eq. (15.55) was used. $A_{\min}$ is the minimum amplitude of light required for the guided nonlinear boundary wave to be excited. The existence of the minimum required value of A can be seen graphically in Fig. 15.5. $A_{\min}$ occurs when the γ_0 line becomes the tangent of the characteristic curve.

Next, the values of Ψ_0 and E_0 will be found. Equations (15.52) and (15.54) are combined as

$$\tanh^2 \Psi_0 = 1 - \left(\frac{\nu}{\gamma_0}\right)^2 \tag{15.58}$$

and from the trigonometric relationship,

$$\operatorname{sech}^2 \Psi_0 = 1 - \tanh^2 \Psi_0 \tag{15.59}$$

Equation (15.59) simplifies to

$$\operatorname{sech} \Psi_0 = \frac{\nu}{\gamma_0} = \frac{1}{A}\sqrt{\frac{2(\epsilon_1 - \epsilon_0)}{\alpha}} \tag{15.60}$$

From Fig. 15.5 note that as A approaches $A_{\min}$, γ_1 in Eq. (15.54) approaches zero and ν/γ_0 approaches unity, and Ψ_0 is decreased and finally reaches zero.

The electric field E_0 at the boundary is given by Eqs. (15.21) and (15.60),

$$E_0^2 = \frac{2(\epsilon_1 - \epsilon_0)}{\alpha} \tag{15.61}$$

Finally, $x_{\max}$ is the location of the peak of the field bulge. From Eq. (15.48),

$$x_{\max} = \frac{\Psi_0}{\gamma_0} \tag{15.62}$$

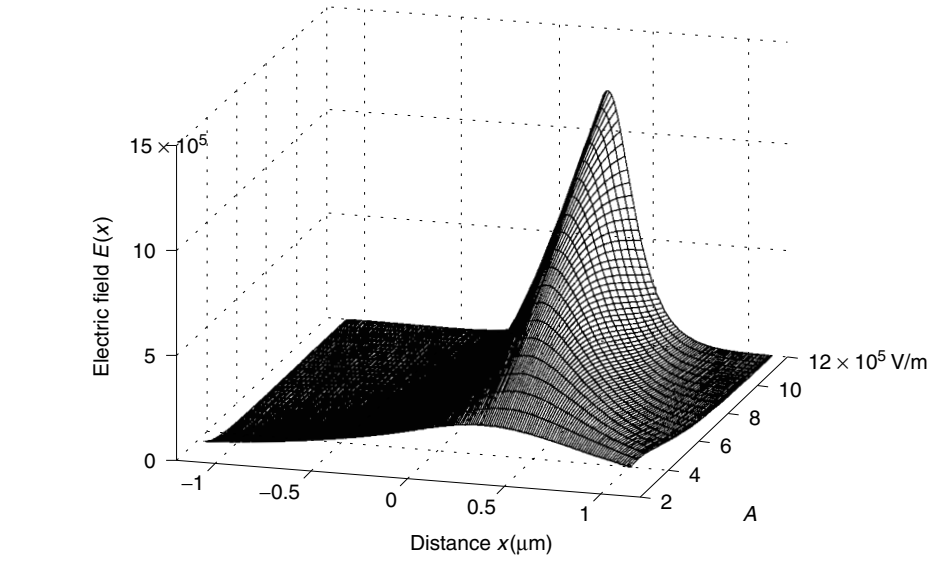
From Eqs. (15.20), (15.21), and (15.62), $x_{\max}$ is

$$x_{\max} = \frac{\sqrt{2}\cosh^{-1}(A/E_0)}{k\sqrt{\alpha}A} \tag{15.63}$$

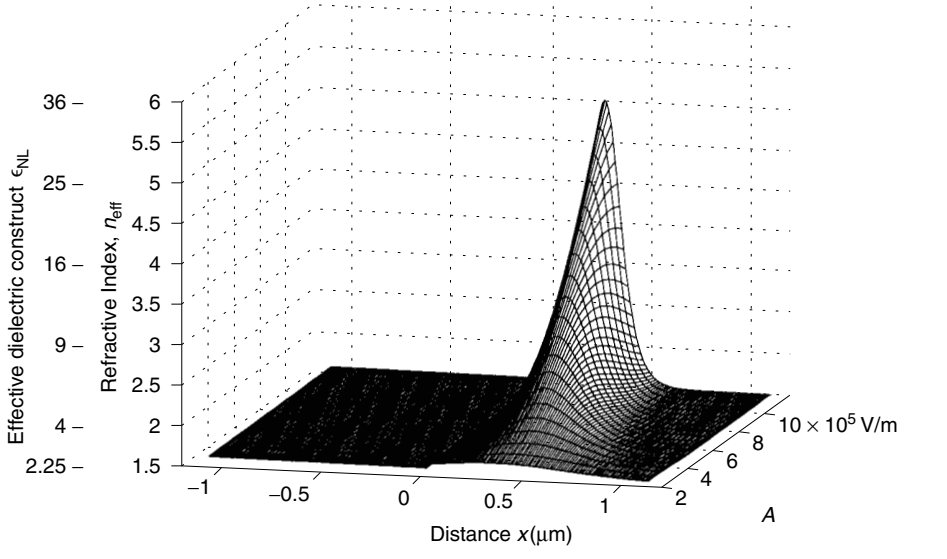
The increase of the numerator with respect to A is slower than that of the denominator. As A increases, the position of $x_{\max}$ slowly approaches the interface. Thus, all the parameters for the field distribution function of the guided nonlinear boundary wave have been calculated. The general shape of the field is shown in Fig. 15.6a with A as a parameter.

The excitation of the guided wave can be interpreted from the shape of the field. The profile of the dielectric constant ϵ_{NL} in the nonlinear region is given by Eq. (15.7) and is shown in Fig. 15.6b. The dielectric constant distribution is similar in shape to the field distribution. It is this bell-shaped distribution of the refractive index created inside the nonlinear medium that acts as a light channel, the same as a graded-index optical guide.

Example 15.1 A guided wave is excited along the boundary between nitrobenzene, which is a nonlinear Kerr medium, and glass, which is considered a linear medium. Calculate the following:



(a)



(b)

Figure 15.6 Field and refractive index distribution of the nonlinear boundary wave. (Calculations courtesy of D. Khomarlou.) (a) Electric field distribution as a function of x for values of A. (b) Refractive index as a function of x for values of A.

(a) The propagation constants γ_0 and γ_1 in the x direction.
(b) The propagation constant β in the z direction.
(c) The electrical field E_0 on the interface.
(d) The position of the maximum field in the nitrobenzene.
(e) The value of ψ_0.

The parameters used are as follows:

$A = 5 \times 10^5$ V/m
$\alpha = 2.44 \times 10^{-12}$ (V/m)$^{-2}$
$\epsilon_0 = 2.253$ (nitrobenzene)
$\epsilon_1 = 2.403$ (glass)
$\lambda = 1.55$ μm

Solution

(a) From Eq. (15.20), γ_0 is

$$\begin{aligned}\gamma_0 &= k\sqrt{0.5\alpha}A \\ &= \frac{2\pi}{1.55 \times 10^{-6}}\sqrt{(0.5)(2.44 \times 10^{-12})}(5 \times 10^5) \\ &= 2.24\ \mu\text{m}^{-1}\end{aligned}$$

From Eq. (15.54), γ_1 is found from

$$\gamma_1 = \sqrt{\gamma_0^2 - \nu^2}$$

where

$$\begin{aligned}\nu^2 &= k^2(\epsilon_1 - \epsilon_0) \\ \nu^2 &= \left(\frac{2\pi}{1.55}\right)^2 (2.403 - 2.253) = 2.46\end{aligned}$$

Thus,

$$\gamma_1 = \sqrt{2.24^2 - 2.46} = 1.60\ \mu\text{m}^{-1}$$

(b) From Eq. (15.17b), β is

$$\begin{aligned}\beta &= \sqrt{\epsilon_0 k^2 + \gamma_0^2} \\ &= \sqrt{2.253\left(\frac{2\pi}{1.55}\right)^2 + 2.24^2} \\ &= 6.48\ \text{rad}/\mu\text{m}\end{aligned}$$

(c) From Eq. (15.61), E_0 is

$$\begin{aligned}E_0 &= \sqrt{\frac{2(\epsilon_1 - \epsilon_0)}{\alpha}} \\ &= \sqrt{\frac{2(2.403 - 2.253)}{2.44 \times 10^{-12}}} \\ &= 3.5 \times 10^5\ \text{V/m}\end{aligned}$$

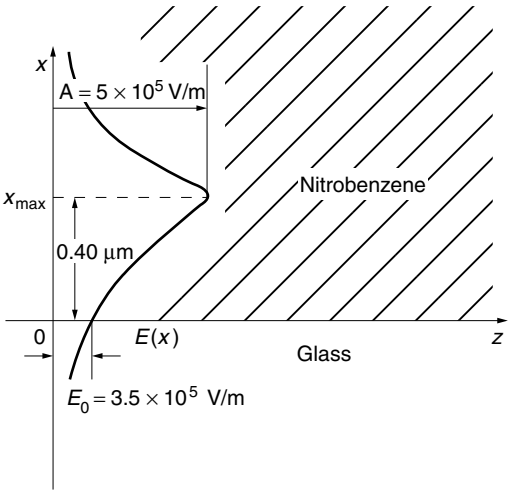


Figure 15.7 Guided wave excitation at the interface between linear and nonlinear media: the results of Example 15.1.

(d) From Eqs. (15.21) and (15.62), $x_{\max}$ is

$$x_{\max} = \frac{\cosh^{-1}\left(\dfrac{A}{E_0}\right)}{\gamma_0} = \frac{\cosh^{-1}\left(\dfrac{5 \times 10^5}{3.5 \times 10^5}\right)}{2.24} = 0.40 \ \mu\text{m}$$

(e) $\psi_0 = \cosh^{-1}\left(\dfrac{A}{E_0}\right) = \cosh^{-1}\left(\dfrac{5 \times 10^5}{3.5 \times 10^5}\right) = 0.90$

The results are summarized in Fig. 15.7. □

15.4 LINEAR CORE LAYER SANDWICHED BY NONLINEAR CLADDING LAYERS

A slab optical guide consists of an optically dense core layer sandwiched by optically less dense cladding layers. This section presents the changes in the behavior of the slab guide caused by the replacement of the linear cladding media with nonlinear media [5–10]. The cross-sectional distribution of the field is altered with the propagating light intensity. As the intensity is increased, three new modes appear, which never existed in the linear case. Some of these modes have their maximum intensity outside the core layer.

15.4.1 General Solutions

The geometry considered is shown in Fig. 15.8. The thickness of the core layer is $2d$, and the wave propagates in the z direction with propagation constant β. The TE wave is excited. It is assumed that there are no field variations in the y direction. The quantities associated with each layer are indicated by subscripts 0, 1, 2, from the top to the bottom layer. The relative dielectric constant ϵ_1 of the core is chosen always higher than the linear part of the nonlinear relative dielectric constant ϵ_0 of the cladding layers.

For TE modes with $E_z = 0$, the differential equations for the y component of the field are given by

$$\frac{d^2E_0(x)}{dx^2} + \left(\epsilon_0 k^2 - \beta^2\right) E_0(x) + k^2 \alpha E_0^3(x) = 0 \tag{15.64}$$

$$\frac{d^2E_1(x)}{dx^2} + \left(\epsilon_1 k^2 - \beta^2\right) E_1(x) = 0 \tag{15.65}$$

$$\frac{d^2E_2(x)}{dx^2} + \left(\epsilon_2 k^2 - \beta^2\right) E_2(x) + k^2 \alpha E_2^3(x) = 0 \tag{15.66}$$

where

$$\epsilon_2 = \epsilon_0 \quad \text{for the symmetrical case} \tag{15.67}$$

and

$$\epsilon_{\mathrm{NL}} = \epsilon_0 + \alpha E^2(x) \tag{15.68}$$

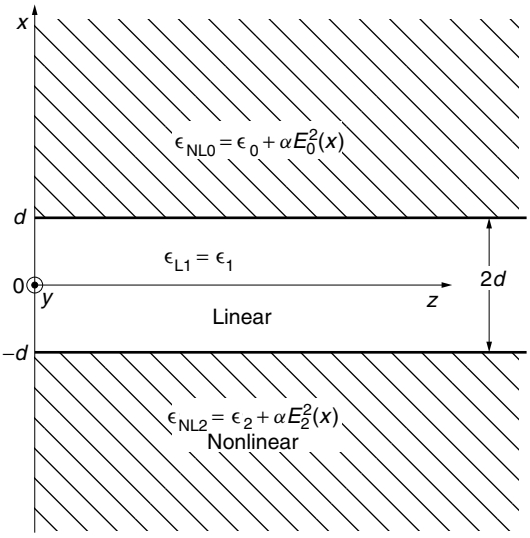


Figure 15.8 Geometry of a three-layer optical guide with a linear core and nonlinear identical claddings.

The general solutions for each layer are

$$E_0(x) = A_0 \text{ sech } [-\gamma_0(x-d) + \Psi_0], \quad x \geq d \tag{15.69}$$

$$E_1(x) = B\cos(K_1 x + \phi), \quad -d \leq x \leq d \tag{15.70}$$

$$E_2(x) = A_2 \text{ sech } [\gamma_2(x+d) + \Psi_2], \quad x \leq -d \tag{15.71}$$

The linear parts of the propagation constant in the claddings are

$$\begin{aligned} -\gamma_0^2 &= \epsilon_0 k^2 - \beta^2 \\ -\gamma_2^2 &= \epsilon_2 k^2 - \beta^2 \end{aligned} \tag{15.72}$$

Note that $\gamma_0^2 = \gamma_2^2$ simply because Eq. (15.67) was assumed. The propagation constant in the core is

$$K_1^2 = \epsilon_1 k^2 - \beta^2 \tag{15.73}$$

From Eqs. (15.72) and (15.73),

$$\begin{aligned} (\gamma_0 d)^2 + (K_1 d)^2 &= V^2 \\ (\gamma_2 d)^2 + (K_1 d)^2 &= V^2 \end{aligned} \tag{15.74}$$

$$V = kd\sqrt{\epsilon_1 - \epsilon_0} \tag{15.75}$$

In the linear medium, instead of using A and B in $A\cos K_1 x + B\sin K_1 x$, B and ϕ are chosen as arbitrary constants. As mentioned in the previous section, these general solutions, Eqs. (15.69) and (15.71), satisfy the differential equations Eqs. (15.64) and (15.66) on the condition that

$$\gamma_0^2 = \tfrac{1}{2}k^2\alpha A_0^2 \tag{15.76}$$

$$\gamma_2^2 = \tfrac{1}{2}k^2\alpha A_2^2 \tag{15.77}$$

are satisfied.

The next section uses the boundary conditions to determine the constants A_0, A_2, B, ϕ, K_1, γ_0, β, Ψ_0, and Ψ_2.

15.4.2 Characteristic Equations from the Boundary Conditions

The continuity of the tangential field across the boundaries at $x = d$ and $x = -d$ gives

$$A_0 \text{ sech } \Psi_0 = B\cos(K_1 d + \phi) \tag{15.78}$$

$$A_2 \text{ sech } \Psi_2 = B\cos(-K_1 d + \phi) \tag{15.79}$$

Similarly, from the continuity of the field derivative at both boundaries, the following equations are obtained:

$$\gamma_0 A_0 \text{ sech } \Psi_0 \tanh \Psi_0 = -K_1 B\sin(K_1 d + \phi) \tag{15.80}$$

$$\gamma_2 A_2 \text{ sech } \Psi_2 \tanh \Psi_2 = K_1 B\sin(-K_1 d + \phi) \tag{15.81}$$

The ratio of Eqs. (15.78) and (15.80) and that of Eqs. (15.79) and (15.81) yields

$$\gamma_0 \tanh \Psi_0 = -K_1 \tan(K_1 d + \phi) \tag{15.82}$$

$$\gamma_2 \tanh \Psi_2 = K_1 \tan(-K_1 d + \phi) \tag{15.83}$$

which can be rewritten as

$$\tan(K_1 d + \phi) = \frac{\tilde{\gamma}_0}{K_1} \tag{15.84}$$

$$\tan(-K_1 d + \phi) = -\frac{\tilde{\gamma}_2}{K_1} \tag{15.85}$$

where $\tilde{\gamma}_0$ and $\tilde{\gamma}_2$ are defined as

$$\tilde{\gamma}_0 = -\gamma_0 \tanh \Psi_0 \tag{15.86}$$

$$\tilde{\gamma}_2 = -\gamma_2 \tanh \Psi_2 \tag{15.87}$$

Using the identity

$$\tan(A - B) = \frac{\tan A - \tan B}{1 + \tan A \tan B}$$

on Eqs. (15.84) to (15.87) leads to

$$\tan 2K_1 d = \frac{K_1(\tilde{\gamma}_0 + \tilde{\gamma}_2)}{K_1^2 - \tilde{\gamma}_0 \tilde{\gamma}_2} \tag{15.88}$$

which is the characteristic equation for the case of the nonlinear cladding layer.

We soon realize that this is exactly the same as the characteristic equation of an asymmetric linear three-layer guide with three different indices of refraction (see Section 9.6 for the TE mode and Section 9.7.2 where the case of the TM mode is treated), if $\tilde{\gamma}_0$ and $\tilde{\gamma}_2$ are replaced by γ_0 and γ_2, respectively.

Let us examine how Eq. (15.88) reduces to the expression of the linear case as the nonlinear constant of the electric field intensity is reduced. From Eq. (15.86), we have

$$\tilde{\gamma}_0^2 = \gamma_0^2(1 - \operatorname{sech}^2 \Psi_0) \tag{15.89}$$

From Eq. (15.69), the field E_0 at the top boundary is

$$E_0 = A_0 \operatorname{sech} \Psi_0 \tag{15.90}$$

Inserting Eq. (15.90) into (15.89) gives

$$\tilde{\gamma}_0^2 = \gamma_0^2 - \gamma_0^2 \left(\frac{E_0}{A_0}\right)^2 \tag{15.91}$$

By using Eq. (15.76), Eq. (15.91) becomes

$$\tilde{\gamma}_0^2 = \gamma_0^2 - \tfrac{1}{2} k^2 \alpha E_0^2 \tag{15.92}$$

and similarly,

$$\tilde{\gamma}_2^2 = \gamma_2^2 - \tfrac{1}{2}k^2\alpha E_2^2 \tag{15.93}$$

Thus, when $\alpha E_0^2 = 0$, then

$$\tilde{\gamma}_0 = \pm\gamma_0$$

and Eq. (15.88) becomes identical to the characteristic equation of the TE modes of the linear case.

At this time Eq. (15.86) becomes

$$\tanh \Psi_0 = \mp 1 \quad \text{and} \quad \Psi_0 = \mp\infty \tag{15.94}$$

Now, the value of Ψ_0 for an arbitrary αE_0^2 will be obtained for later use. From Eqs. (15.76) and (15.90), we have

$$\cosh^2 \Psi_0 = \frac{2\gamma_0^2}{k^2\alpha E_0^2} \tag{15.95}$$

and from Eqs. (15.92) and (15.95), we have

$$\sinh^2 \Psi_0 = \frac{2\tilde{\gamma}_0^2}{k^2\alpha E_0^2} \tag{15.96}$$

Expressions for the bottom cladding are obtained by changing the subscript from 0 to 2.

15.4.3 Normalized Thickness of the Nonlinear Guides

In order to solve Eq. (15.88), two more equations involving $\tilde{\gamma}_0$ and $\tilde{\gamma}_2$ are necessary. A nonlinear normalized thickness of the guide will be defined in a form easily reduced to Eqs. (15.74) and (15.75). Let a new parameter W_0, called the *power level parameter*, be defined from Eqs. (15.92) and (15.93) as

$$W_0^2 = (\gamma_0 d)^2 - (\tilde{\gamma}_0 d)^2 = (kd)^2 \tfrac{1}{2}\alpha E_0^2 \tag{15.97}$$

$$W_2^2 = (\gamma_2 d)^2 - (\tilde{\gamma}_2 d)^2 = (kd)^2 \tfrac{1}{2}\alpha E_2^2 \tag{15.98}$$

Equation (15.74) gives

$$(\tilde{\gamma}_0 d)^2 + (K_1 d)^2 = V^2 - W_0^2 \tag{15.99}$$

and similarly,

$$(\tilde{\gamma}_2 d)^2 + (K_1 d)^2 = V^2 - W_2^2 \tag{15.100}$$

The two quantities $\sqrt{V^2 - W_0^2}$ and $\sqrt{V^2 - W_2^2}$ are the *nonlinear normalized thicknesses*. They represent the radius of a circle in the $\tilde{\gamma}_0 d - K_1 d$ plane or $\tilde{\gamma}_2 d - K_1 d$ plane.

Hence, the nonlinear normalized thickness decreases with an increase in the field intensity. One interpretation is that the increase in the field raises the refractive index of the cladding because of the nonlinearity, and the difference in the refractive indices between the core and cladding decreases, and hence the normalized thickness is reduced.

15.4.4 Fields at the Top and Bottom Boundaries

With the nonlinear case, the solution depends on the field intensity and the fields have to be specified. The field in the top cladding is connected with that in the bottom cladding by the sinusoidal field in the core. The fields E_0 and E_2 at the top and bottom boundaries of the core are deduced from Eq. (15.70) by inserting $x = d$ and $-d$, respectively:

$$E_0 = B\cos(K_1 d + \phi) \tag{15.101}$$

$$E_2 = B\cos(-K_1 d + \phi) \tag{15.102}$$

The ratio of these two equations gives

$$E_0^2 \cos^2(-K_1 d + \phi) = E_2^2 \cos^2(K_1 d + \phi) \tag{15.103}$$

Rewriting the cosine function in terms of the tangent function and using Eqs. (15.84) and (15.85) gives

$$E_0^2(K_1^2 + \tilde{\gamma}_0^2) = E_2^2(K_1^2 + \tilde{\gamma}_2^2) \tag{15.104}$$

From Eqs. (15.92) and (15.93), Eq. (15.104) becomes

$$E_0^2(K_1^2 + \gamma_0^2 - \tfrac{1}{2}k^2\alpha E_0^2) = E_2^2(K_1^2 + \gamma_2^2 - \tfrac{1}{2}k^2\alpha E_2^2) \tag{15.105}$$

New quantities a_0 and a_2 that represent the intensities at the top and bottom interfaces are put as

$$a_0 = \tfrac{1}{2}\alpha E_0^2 \quad \text{and} \quad a_2 = \tfrac{1}{2}\alpha E_2^2 \tag{15.106}$$

Note that a_0 and a_2 are one-half of the nonlinear part of the dielectric constant. Inserting Eqs. (15.74), (15.75), and (15.106) into Eq. (15.105) gives

$$a_0(\epsilon_1 - \epsilon_0 - a_0) = a_2(\epsilon_1 - \epsilon_0 - a_2) \tag{15.107}$$

Equation (15.107) can be rewritten in final form as

$$(a_0 - a_2)[\epsilon_1 - \epsilon_0 - (a_0 + a_2)] = 0 \tag{15.108}$$

Thus,

$$a_0 = a_2 \tag{15.109}$$

or

$$a_0 + a_2 = \epsilon_1 - \epsilon_0 \tag{15.110}$$

These two conditions on the boundary field intensities will play an important role in classifying the different modes in the guide.

15.4.5 Modes Associated with Equal Boundary Field Intensities ($a_0 = a_2$)

From Eq. (15.106), Eq. (15.109) implies that

$$E_0^2 = E_2^2 \tag{15.111}$$

or

$$E_0 = \pm E_2 \tag{15.112}$$

Since the physical parameters of the top and bottom cladding are identical, from Eq. (15.72),

$$\gamma_0^2 = \gamma_2^2 \tag{15.113}$$

Equations (15.92), (15.93), (15.111), and (15.113) lead to

$$\tilde{\gamma}_0^2 = \tilde{\gamma}_2^2 \tag{15.114}$$

or

$$\tilde{\gamma}_0 = +\tilde{\gamma}_2 \tag{15.115}$$

$$\tilde{\gamma}_0 = -\tilde{\gamma}_2 \tag{15.116}$$

The plus and minus signs of the $\tilde{\gamma}$'s influence the final results and the treatment will be separated into two cases.

(a) Case of $\tilde{\gamma}_0 = \tilde{\gamma}_2$ With Eq. (15.115), Eq. (15.88) becomes

$$\tan 2K_1 d = \frac{2K_1\tilde{\gamma}_0}{K_1^2 - \tilde{\gamma}_0^2} \tag{15.117}$$

Equation (15.117) is a quadratic equation in $\tilde{\gamma}_0^2$ and can be solved as

$$\tilde{\gamma}_0 d = K_1 d \tan K_1 d \tag{15.118}$$

and

$$\tilde{\gamma}_0 d = -K_1 d \cot K_1 d \tag{15.119}$$

The characteristic equation has become quite similar to that of the linear case (see Section 9.6). Solutions of Eq. (15.118) are even modes or E modes, and those of Eq. (15.119) are odd modes or O modes of the nonlinear guide. The boundary fields

with the plus sign in Eq. (15.112) belong to the E mode, and those with the minus sign belong to the O modes.

The solutions $\tilde{\gamma}_0 d$ and $K_1 d$ are the intersections of the characteristic equation with the nonlinear normalized thickness of the guide. After $\tilde{\gamma}_0$ is obtained, the value of γ_0 is determined from Eq. (15.92) as

$$\gamma_0 = \pm\sqrt{\tilde{\gamma}_0^2 + \tfrac{1}{2}k^2\alpha E_0^2} \tag{15.120}$$

The sign of γ_0 has to be the same as that of $\tilde{\gamma}_0$ because $\tilde{\gamma}_0$ has to approach γ_0 as αE_0^2 is reduced and

$$\gamma_0 > 0 \tag{15.121}$$

The value of β is, from Eq. (15.72),

$$\beta = \pm\sqrt{\epsilon_0 k^2 + \gamma_0^2} \tag{15.122}$$

The position $x_{\max 0}$ of the peak of the bulge in the top cladding is, from Eq. (15.69),

$$x_{\max 0} = d + \frac{\Psi_0}{\gamma_0} \tag{15.123}$$

An example of obtaining the solutions of the TE_0 and TE_1 modes with $V = \sqrt{5.0}$ for various values of W_0 is shown in Fig. 15.9. With $W_0^2 = 2.5$ the first two modes TE_0 and TE_1 indicated by points E and O are excited with

$$\tilde{\gamma}_0 d > 0 \tag{15.124}$$

From Eqs. (15.86) and (15.121), Eq. (15.124) implies

$$\Psi_0 < 0 \tag{15.125}$$

Thus, the position $x_{\max 0}$ of the peak of the bulge is obtained from Eq. (15.123) and is

$$x_{\max 0} < d \tag{15.126}$$

The peak is not inside the top cladding, as shown in Fig. 15.10a and also by the drawing noted as double decay (in the cladding) in the upper section of Fig. 15.9. It is important to note that Eq. (15.69) is applicable only inside the top cladding. Such a peak outside the top cladding, as indicated by the dashed line in Fig. 15.10a is not legitimate; only its tail section is legitimate. The distribution in the top cladding is of a decaying shape similar to the linear case. It is the sign of $\tilde{\gamma}_0 d$, hence the sign of Ψ_0, that determines whether or not a bulge of the field is present in the cladding.

Next, the case of a larger W_0 is considered. With an increase in E_0, the nonlinear normalized thickness reduces and the TE_1 mode starts having a solution with

$$\tilde{\gamma}_0 d < 0 \tag{15.127}$$

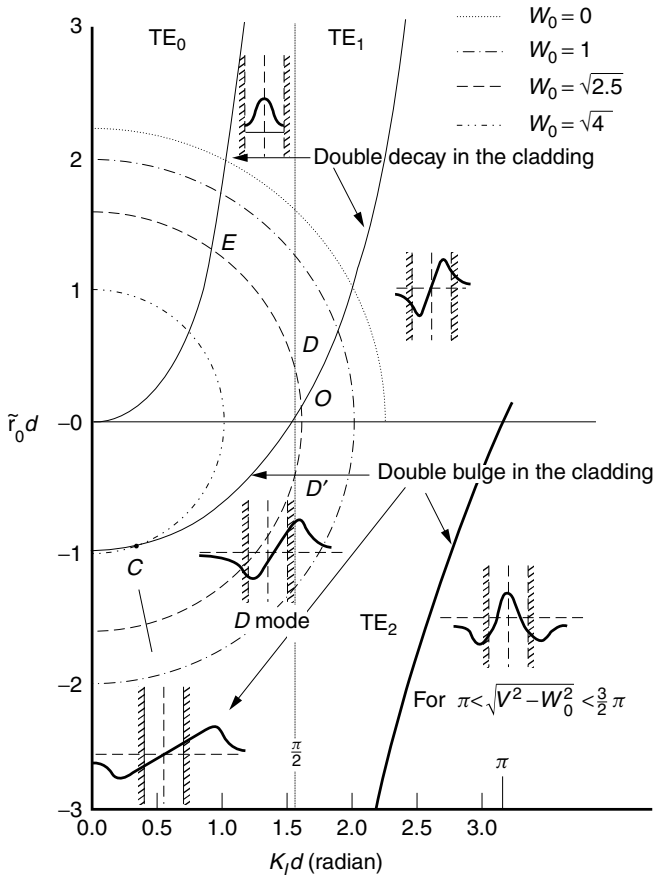


Figure 15.9 Solutions in a waveguide with $V = \sqrt{5.0}$: even and odd modes for both linear and nonlinear cases. (After Y. F. Li and K. Iizuka [11].)

as indicated by point C. From Eqs. (15.86) and (15.121), Eq. (15.127) implies

$$\Psi_0 > 0 \tag{15.128}$$

and the peak of the bulge starts to exist in the top cladding layer, as shown in Fig. 15.10b and also by the drawing noted as the double bulge (in the cladding) in the lower section of Fig. 15.9.

The value of A_0 is found by first projecting the intersecting point on the $\tilde{\gamma}_0 d$ coordinate from which the value of γ_0 is calculated by using Eq. (15.97) for a given value of W_0 and then applying Eq. (15.76) to find A_0.

It is interesting to note that the lowest order mode TE_0 can never have an intersection in the negative $\tilde{\gamma}_0 d$ and cannot have its bulge in the cladding. This is understandable due to the shape of the field of the TE_0 mode. The field inside the core is always decaying with distance from the center and this shape of the field cannot be smoothly connected to the shape of the bulge, which needs a rising shape at the boundary.

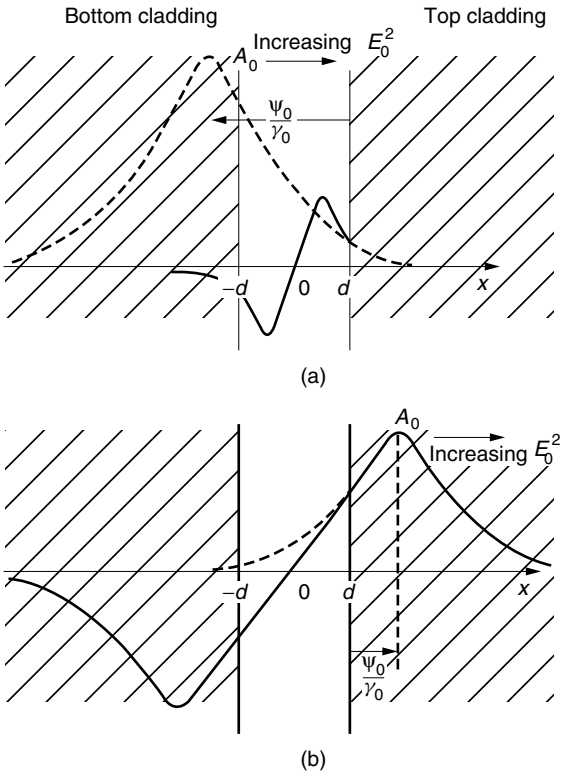


Figure 15.10 Change in the TE_1 mode field pattern with the field intensity E_0^2. (a) TE_1 mode with lower field intensity. (b) TE_1 mode with higher field intensity.

The negative value of γ_0 is not permissible in the linear case because the field becomes infinite as x approaches infinity. In the nonlinear case, however, negative $\tilde{\gamma}_0 d$ is possible and means a bulge is present in the cladding.

Similarly, in the bottom cladding

$$x_{\max 2} = -d - \frac{\Psi_2}{\gamma_2} \tag{15.129}$$

Example 15.2 For the TE_1 mode, at what value of E_0 does a bulge begin to exist in the cladding layer? The normalized thickness of the guide is $V = \sqrt{5.0}$. The other parameters are

$$\alpha = 2.44 \times 10^{-12} \text{ (V/m)}^{-2}$$
$$\epsilon_0 = 2.253$$
$$\lambda = 1.55 \text{ μm}$$
$$2d = 10 \text{ μm}$$

Solution The bulge starts to exist as soon as $\tilde{\gamma}_0 d$ begins to be a negative number. Referring to Fig. 15.9, as soon as the nonlinear normalized thickness reduces to $\pi/2$,

the value of $\tilde{\gamma}_0 d$ starts to become negative. From Eq. (15.99), the nonlinear normalized thickness is expressed as

$$\sqrt{V^2 - W_0^2} = \frac{\pi}{2}$$

E_0 is obtained from Eq. (15.97):

$$E_0 = \frac{\sqrt{V^2 - \left(\frac{\pi}{2}\right)^2}}{kd\sqrt{0.5\alpha}} = \frac{\sqrt{5 - 1.57^2}}{\left(\frac{2\pi}{1.55}\right)(5)\sqrt{(0.5)(2.44 \times 10^{-12})}}$$
$$= 7.1 \times 10^4 \text{ V/m}$$

Without a detailed calculation, we could guess that if we added a second core layer in the top cladding in close vicinity to the original core layer, the amount of coupled energy from the original core to the second core will depend significantly on whether or not this bulge of the field is present. We could imagine that a light intensity activated optical switch may be constructed out of such a structure. □

Example 15.3 From Eq. (15.97),

$$(\gamma_0 d)^2 = (\tilde{\gamma}_0 d)^2 + W_0^2$$

and from Eqs. (15.95) and (15.97),

$$\cosh \Psi_0 = \frac{\gamma_0 d}{W_0}$$

Describe the procedure for graphically finding $\gamma_0 d$ and $\cosh \Psi_0$ using Fig. 15.11.

Solution

1. Draw a circle with radius V, which is the linear normalized thickness.
2. Determine the radius $\sqrt{V^2 - W_0^2}$ of the circle, which is the nonlinear normalized thickness from the triangle OAC, where $AO = V$ and $AC = W_0$.
3. The intersection between the characteristic equation of the TE_0 mode $\tilde{\gamma}_0 d = K_1 d \tan K_1 d$ and the circle OC (or OD) gives the value $\tilde{\gamma}_0 d = OE$.
4. The triangle EFO gives OF, which is $\gamma_0 d = \sqrt{(\tilde{\gamma}_0 d)^2 + W^2}$, and OF/EF, which is $\cosh \Psi = \gamma_0 d / W_0$. □

(b) Case of $\tilde{\gamma}_0 = -\tilde{\gamma}_2$ If Eq. (15.116) is satisfied, Eq. (15.88) becomes

$$\tan 2K_1 d = 0 \tag{15.130}$$

and hence

$$2K_1 d = q\pi \tag{15.131}$$

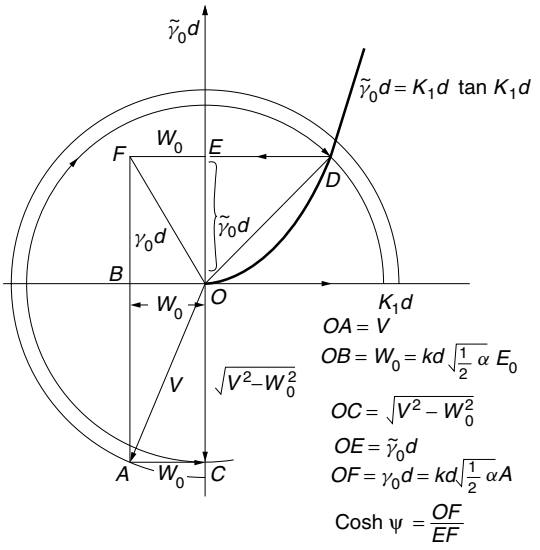


Figure 15.11 Graphically finding $\gamma_0 d$ and cosh ψ.

where q is an integer. The plot of the characteristic equation becomes vertical lines in Fig. 15.9. (The $q = 1$ line is indicated by a fine line.) Such modes are called D modes. The vertical lines intersect the circle given by Eq. (15.99) at two points as indicated by D and D'. The value of $\tilde{\gamma}_0$ at these two points has the same magnitude but with opposite signs. The positive solution has no bulge in the top cladding but has a bulge in the bottom cladding. On the contrary, the negative solution has a bulge in the top cladding but no bulge in the bottom cladding. The distribution of the field of the D modes is asymmetric.

In conclusion, for the case $a_0 = a_2$, the solution with $\tilde{\gamma}_0 = \tilde{\gamma}_2$ generates symmetric modes that are E modes and antisymmetric modes that are O modes, and the solutions with $\tilde{\gamma}_0 = -\tilde{\gamma}_2$ generate asymmetric modes that are D modes. Besides the fact that the field can have a bulge in the cladding when $\tilde{\gamma}_0$ or $\tilde{\gamma}_2$ is negative, the existence of the D mode is unique for the nonlinear optical guide. Classification of the modes is shown in Table 15.3.

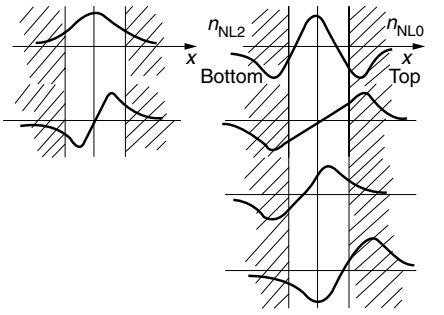
15.4.6 Modes Associated with $a_0 + a_2 = \epsilon_1 - \epsilon_0$

A similar approach as the previous section can be used [7].

First, we express $\tilde{\gamma}_2$ in terms of $\tilde{\gamma}_0$ including condition Eq. (15.110) so that the characteristic equation Eq. (15.88) becomes a function of $\tilde{\gamma}_0 d$ and $K_1 d$. The intersection of this curve plotted in $\gamma_0 d$ versus $K_1 d$ coordinates with the circle of the nonlinear normalized thickness, Eq. (15.99), provides two additional modes called the A and B modes. The field distribution of the A mode is similar to the E mode but with a shift in the x direction. The field distribution of the B mode is somewhat similar to the D mode. In conclusion, the modes that are present only when the cladding is nonlinear are totally the A, B, D, E, and O modes.

For the case when the core is nonlinear rather than the cladding, the problem can be treated in a similar manner but with slightly more complexity [11,12].

Table 15.3 Classification of modes in an optical guide with nonlinear cladding

a'_s	$\tilde{\gamma}'_s$	Mode Name	Fields at Boundaries	Symmetry	Field Distribution: Without Bulge in Cladding / With Buldge in Cladding
$a_0 = a_2$	$\tilde{\gamma}_o = \tilde{\gamma}_2$	E	$E_0 = E_2$	Symmetric	
		O	$E_0 = -E_2$	Antisymmetric	
	$\tilde{\gamma}_o = -\tilde{\gamma}_2$	D	$\gamma_0 > 0$, $\gamma_2 < 0$	Asymmetric	
			$\gamma_0 < 0$, $\gamma_2 > 0$	Asymmetric	

Finally, let us remember that the law of superposition of the solutions does not hold in the case of nonlinear differential equations, except in special cases. As a result, the superposition of the modes has to be treated with special care.

15.5 HOW THE SOLITON CAME ABOUT

One day in 1834, the Scottish scientist John Scott-Russell was sitting on a bench on the bank of a large channel, watching the motion of waves generated by a sudden stop of a barge. Water waves of a certain shape kept on traveling without changing their shape for a distance as far as his eyes could see. He decided to ride his horse to follow the waves. He discovered that the wave of a particular shape travels miles without changing its original shape [13,14].

In 1895 Korteweg and de Vries [15] derived a partial differential equation for such a solitary wave in shallow water as

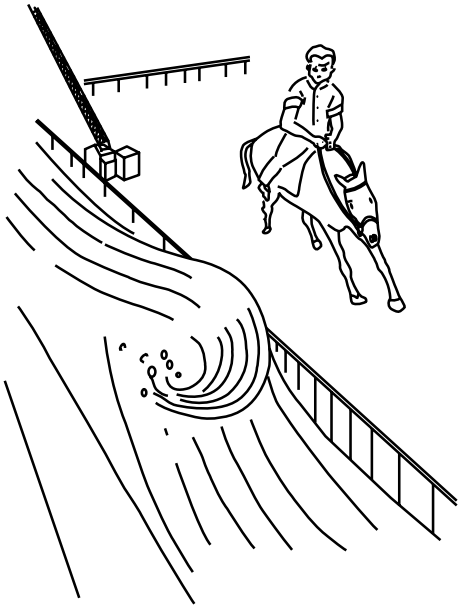
$$U_t - 6UU_x + U_{xxx} = 0 \tag{15.132}$$

where U is the solution of the wave equation and the subscripts t and x denote the derivatives with respect to time and distance. This differential equation is known as the Korteweg–de Vries (KdV) differential equation. The steady-state solution is given by [15]

$$U = -\frac{c}{2}\text{sech}^2\left(\frac{\sqrt{c}}{2}(x - ct)\right) \tag{15.133}$$

where c is the speed of propagation of the water wave. The features of this wave are as follows:

1. The shape of the wave is independent of distance and time, and the wave propagates without changing shape.
2. The height $c/2$ of the wave is proportional to the speed of propagation.



John Scott-Russell galloping after a solitary wave in a Scottish canal in 1838.

3. The width of the wave $2/\sqrt{c}$ is inversely proportional to the square root of the height $c/2$ of the wave.

Such a wave was named a *soliton*.

A lightwave soliton can travel in an optical fiber over a long distance without distortion of the pulse shape. This opens up the possibility of repeaterless transmission of light pulses across a 60,000-km transoceanic fiber-optic submarine cable. In 1972, Zakharov and Shabat [17] derived a differential equation for light propagating in an optical fiber that predicted the existence of solitons. In 1973, Hasegawa demonstrated that the soliton can exist in an optical fiber in the wavelength region longer than 1.3 μm [18]. Solitons in an optical fiber were first used by Mollenauer and Stolen for generating subpicosecond pulses [19].

Additional special features of the soliton are the following:

1. Even after two solitons collide with each other, they recover their individual shapes. This particle-like property has been of interest to many physicists who use the soliton to represent particles.
2. The soliton was named after a wave pulse that does not change its shape after propagation. Some solitons (higher-order solitons) do change their shape and periodically recover their shape as they propagate.

15.6 HOW A SOLITON IS GENERATED

We have learned that the practical distance a light pulse can travel in a single-mode fiber is significantly increased if the fiber's dispersion effect is either suppressed or

compensated. The dispersion countermeasures we have looked at so far are four-wave mixing (Section 8.12), fibers with internal Bragg gratings (Section 11.5.2), dispersion-shifted fibers (Section 11.7), and dual mode fibers (Section 11.8.3).

In this section, the nonlinear Kerr effect of the glass is used to counterbalance the dispersion of the single mode fiber to form a soliton for the long-distance transmission of the light pulse.

Figure 15.12 illustrates how a soliton wave is formed [13,14]. Consider a high-intensity light pulse propagating in the positive z direction inside the core glass of an optical fiber. The pulse intensity distribution with respect to the distance z is shown in Fig. 15.12a. The same pulse intensity distribution is plotted with respect to time t as the pulse crosses a fictitious goal line at $z = L$, as shown in Fig. 15.12b. The region of positive t represents the future, and that of negative t, the past. The leading edge has already reached the goal line and is therefore shown in the negative t region, while the trailing edge has not yet passed the goal line and is shown in the positive t region. Hence, the left and right representations of the leading and trailing edges with respect to time in Fig. 15.12b are reversed from that with respect to distance in Fig. 15.12a.

A phase shift is experienced by an optical field during its propagation in the optical fiber. The phase shift including the Kerr effect is expressed by

$$\phi(L, t) = -2\pi f_c t + k n_{\mathrm{NL}} L \tag{15.134}$$

where f_c is the carrier frequency of the pulse and L is the distance of transmission. From Eq. (a) in the boxed note in Section 15.2,

$$n_{\mathrm{NL}} = n_1 + n_2 I \tag{15.135}$$

Equation (15.134) is represented in Fig. 15.12c. The instantaneous angular frequency $f(L, t)$ of the carrier is

$$f(L, t) = -\frac{1}{2\pi}\frac{d}{dt}\phi(L, t) \tag{15.136}$$

Inserting Eq. (15.134) into (15.136) gives

$$f(L, t) = f_c + \Delta f_p \tag{15.137}$$

where

$$\Delta f_p = -\frac{n_2}{\lambda} L \frac{dI(L, t)}{dt} \tag{15.138}$$

The quantity Δf_p is the frequency shift due to the time dependence of the nonlinear refractive index of the core glass and is shown in Fig. 15.12d.

From Fig. 15.12d, the higher frequency components cluster in the trailing edge of the pulse and the lower frequency components cluster in the leading edge. The above mentioned light intensity dependent phenomenon is called *self-phase modulation* (SPM) and it is caused by the nonlinearity of the fiber.

Next, we should investigate another phenomenon relevant to pulse propagation in the fiber. Consider a self-phase modulated pulse when it propagates through a medium

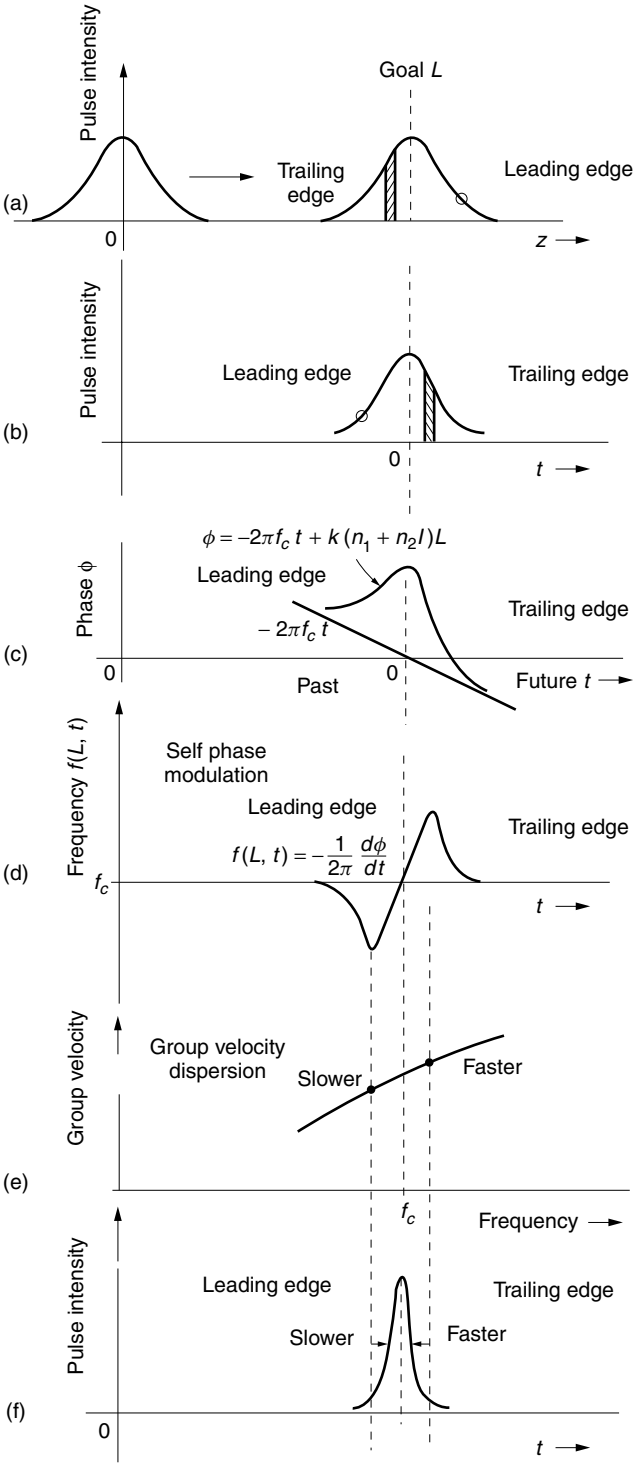


Figure 15.12 Explanation of the formation of a soliton wave.

with group velocity dispersion (GVD), such as in an optical fiber. The interaction of SPM with GVD creates the soliton in the glass fiber.

The dispersion in glass is frequency dependent. In the region of wavelengths shorter than 1.3 μm (normal dispersion region), the higher frequency components travel slower. On the other hand, in the region of wavelengths longer than 1.3 μm (anomalous dispersion region), the higher frequency components travel faster. Now, referring to the frequency distribution inside the pulse shown in Fig. 15.12d, the trailing edge has higher frequency components and if the carrier frequency is in the anomalous dispersion region ($\lambda > 1.3$ μm) as shown in Fig. 15.12e, the trailing edge travels faster than the leading edge and has a chance to catch up with the leading edge. In this manner the pulse can be squeezed and sharpened as it travels, as shown in Fig. 15.12f. This is precisely the soliton wave. The soliton wave is a result of the balancing between SPM and GVD during the propagation of a high-intensity light pulse in the glass.

15.7 SELF-PHASE MODULATION (SPM)

As the pulse is sharpened, the frequency spectrum spreads due to amplitude modulation as well as frequency modulation. The spread of frequency is complicated when these two coexist. An approximate resultant spread is the addition of the two.

The frequency spread of the amplitude modulation alone can be calculated by the Fourier transformation of the pulse shape. To a first order approximation, the spread Δf_a (one side) of a rectangular pulse whose width is $2T$ (see Section 1.4.1) can be approximated as

$$\Delta f_a = \frac{1}{2T} \tag{15.139}$$

The resultant spread Δf_t of frequency due to both the frequency modulation Δf_p and amplitude modulation is

$$\Delta f_t = \Delta f_p + \Delta f_a \tag{15.140}$$

Note that the spectrum of the frequency spread due to the self-phase modulation, which is based on the nonlinear effect, is clustered at the particular location of the pulse as shown in Fig. 15.12d. On the other hand, the frequency spread due to the amplitude or frequency modulation (both of which are linear phenomena) is not clustered.

Example 15.4 Calculate the amount of frequency modulation due to the SPM of the Kerr effect in an optical fiber. The distance of transmission is 200 m. The light is modulated into a trianglar-shaped pulse as shown in Fig. 15.13 with a base width of 200 ps. The peak intensity of the pulse is 160 MW/cm^2. The refractive index of the core glass is 1.55, and the nonlinear-index coefficient is $n_2 = 3.18 \times 10^{-20}$ m^2/W. Assume $\lambda = 1.55$ μm and assume the optical fiber is a dispersion shifted fiber that is free from dispersion at this wavelength.

Solution Let 2T be the base width of the light pulse.

$$\frac{dI}{dt} = \frac{I_0}{T} \tag{15.141}$$

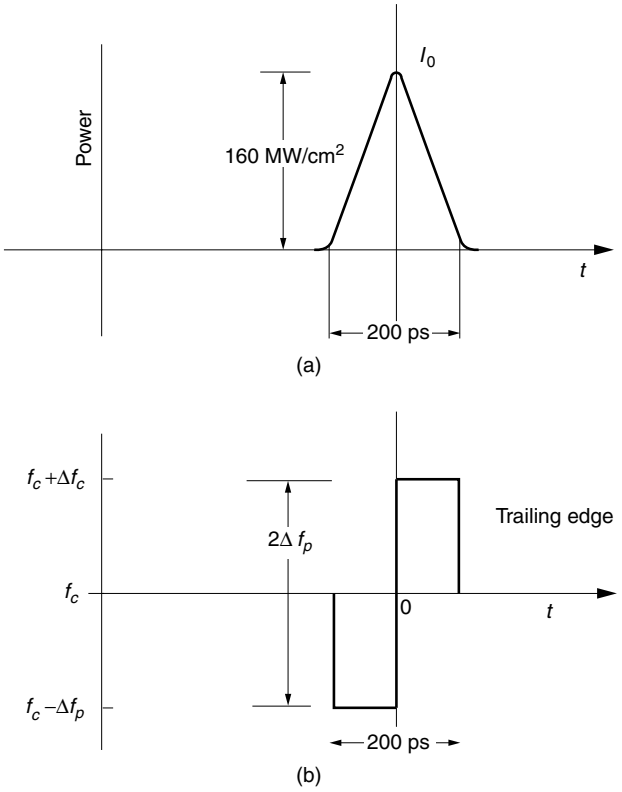


Figure 15.13 Frequency spread due to self-phase modulation (SPM) in a Kerr medium. (a) Triangular pulse with smooth edges. (b) Frequency spread due to SPM effect.

From Eq. (15.138), the frequency shift of one side of the pulse is

$$\Delta f_p = -\frac{n_2}{\lambda} L \frac{I_0}{T} \tag{15.142}$$

$$\Delta f_p = -\frac{3.18 \times 10^{-20}}{1.55 \times 10^{-6}} (200) \frac{1.6 \times 10^{12}}{10^{-10}}$$

$$\Delta f_p = -6.57 \times 10^{10} \text{ Hz} \tag{15.143}$$

Δf_p in terms of $\Delta\lambda$ is calculated approximately using the relationships

$$\lambda = \frac{c}{f} \tag{15.144}$$

$$\Delta\lambda = \frac{d\lambda}{df} \Delta f$$

$$\Delta\lambda = -\lambda^2 \frac{1}{c} \Delta f \tag{15.145}$$

as

$$\Delta\lambda_p = (1.55)^2 \left(\frac{1}{3 \times 10^{14}} \right) (6.57 \times 10^{10}) = 0.53 \text{ nm} \tag{15.146}$$

The spread due to the triangular-shaped amplitude modulation itself is

$$\Delta f_a = \frac{1}{100 \times 10^{-12}} = 1 \times 10^{10} \text{ Hz}$$

The corresponding spread in wavelength is $\Delta\lambda_a = 0.08$ nm. Considering the positive and negative frequency shifts, the total amount of wavelength spread is

$$\Delta\lambda_t = 2(\Delta\lambda_p + \Delta\lambda_a) = 1.22 \text{ nm}$$ □

15.8 GROUP VELOCITY DISPERSION

The velocity of the envelope for an amplitude modulated wave, that is, the group velocity v_g, is given in Section 11.6.1.

$$v_g = \frac{d\omega}{d\beta} \tag{15.147}$$

where β is the propagation constant of the wave in the medium. The inverse of the group velocity is the time taken by the pulse to travel a unit distance and is called the *group delay* τ. The expression for τ is

$$\tau = \frac{1}{v_g} = \frac{d\beta}{d\omega} \tag{15.148}$$

In Section 11.6.2, the dispersion parameter was defined in terms of the spread of the group delay with respect to wavelength and was given by Eq. (11.150), but the dispersion can also be written in terms of frequency as

$$D = \frac{d\tau}{d\lambda} = \frac{d\tau}{d\omega}\frac{d\omega}{d\lambda} \tag{15.149}$$

Denoting the second derivative of the propagation constant as $\beta^{(2)}$, the factor $d\tau/d\omega$ in Eq. (15.149) is expressed with the help of Eq. (15.148) as

$$\frac{d\tau}{d\omega} = \beta^{(2)} \tag{15.150}$$

Using

$$\omega = \frac{2\pi c}{\lambda}$$

The factor $d\omega/d\lambda$ in Eq. (15.149) is expressed as

$$\frac{d\omega}{d\lambda} = -\frac{2\pi c}{\lambda^2} \tag{15.151}$$

Putting Eqs. (15.150) and (15.151) together, D finally becomes

$$D = -\frac{2\pi c}{\lambda^2}\beta^{(2)} \tag{15.152}$$

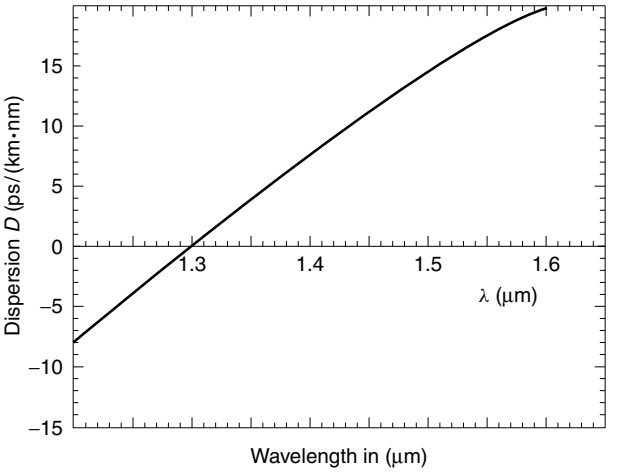


Figure 15.14 Dispersion parameter for calculating group velocity dispersion (GVD) of a single-mode optical fiber.

Note that the sign of D is opposite to that of $\beta^{(2)}$. The dispersion parameter D of the single-mode fiber shown in Fig. 11.9 is replotted on a finer scale in Fig. 15.14.

The dispersion parameter D changes its sign at $\lambda = 1.3$ μm. The region of $D < 0$ or $\beta^{(2)} > 0$ is called the normal dispersion region and that of $D > 0$ or $\beta^{(2)} < 0$ is called the anomalous dispersion region. The significance of the difference between the regions will be investigated by using

$$\beta^{(2)} = \lim_{\Delta\omega \to 0} \frac{\tau(\omega + \Delta\omega) - \tau(\omega)}{\Delta\omega} \tag{15.153}$$

In the normal dispersion region of $\beta^{(2)} > 0$

$$\begin{aligned} \tau(\omega + \Delta\omega) &> \tau(\omega) \\ v_g(\omega + \Delta\omega) &< v_g(\omega) \end{aligned} \tag{15.154}$$

This result means that, in the normal dispersion region, the group velocity of the higher frequency component is slower. The opposite is true in the anomalous dispersion region of $\beta^{(2)} < 0$:

$$\begin{aligned} \tau(\omega + \Delta\omega) &< \tau(\omega) \\ v_g(\omega + \Delta\omega) &> v_g(\omega) \end{aligned} \tag{15.155}$$

The higher frequency component is faster. In the anomalous dispersion region of a single-mode fiber, the soliton wave can exist because the trailing edge of the pulse in Fig. 15.12f, which contains the higher frequency components, has a chance to catch up with the leading edge.

Example 15.5 For the 200 ps base width triangular pulse shown in Example 15.4, find the value of the peak power that can support a soliton in a 6-μm effective diameter

single-mode fiber at $\lambda_0 = 1.55$ μm. The length of the optical fiber is 100 km. Assume $n_1 = 1.55$ and $n_2 = 3.18 \times 10^{-20}$ m^2/W.

Assume that 100 ps is the required differential of arrival times of the trailing and leading edges, and the pulse shape is preserved during transmission.

Solution The condition for the balance between the SPM and GVD at the end of the fiber will be found. The GVD is addressed first. From Fig. 15.14, at $\lambda = 1.55$ μm, the dispersion parameter is

$$D = -17 \text{ ps/(km} \cdot \text{nm)}$$

The spread $\Delta\tau$ of the arrival time is

$$\Delta\tau = |D|\Delta\lambda \cdot L \tag{15.156}$$

The given assumption is that 100 ps is required for the trailing edge to catch up with the leading edge and

$$100 = 17(10^2)\Delta\lambda$$
$$\Delta\lambda = 5.9 \times 10^{-2} \text{ nm}$$

The required wavelength spread on one side of the pulse is $\Delta\lambda/2$, which in terms of frequency is

$$\Delta f_p = -\frac{c}{\lambda^2}\frac{\Delta\lambda}{2}$$
$$= -\frac{3 \times 10^{14}}{1.55^2} 2.95 \times 10^{-5} = -3.7 \times 10^9 \text{ Hz}$$

Next, the required peak intensity I_0 for the generation of Δf_p is obtained.

$$\Delta f_p = -\frac{n_2 L}{\lambda} \cdot \frac{dI}{dt}$$

and

$$I_0 = \frac{dI}{dt} T$$

Combining the above two equations gives

$$I_0 = -\Delta f_p \frac{\lambda T}{n_2 L}$$
$$= 3.7 \times 10^9 \frac{(1.55 \times 10^{-6})(10^{-10})}{(3.18 \times 10^{-20})(10^5)}$$
$$= 180 \text{ MW/m}^2$$

The effective cross sectional area A_{eff} of the core is

$$A_{\text{eff}} = \pi \frac{(6 \times 10^{-6})^2}{4} = 2.8 \times 10^{-11} \text{ m}^2$$

The needed input peak power to the optical fiber is

$$\begin{aligned} P_s &= I_0 \cdot A_{\text{eff}} \\ &= (180 \times 10^6)(2.8 \times 10^{-11}) = 5.0 \text{ mW} \end{aligned}$$

or the energy per pulse is $\frac{1}{2}(200 \times 10^{-12})(5.0 \times 10^{-3}) = 0.5$ pJ. The transmission loss of a 100-km fiber is significant and in order to maintain 5.0-mW average power in this length of fiber, a much larger input power is needed. □

The necessary negative dispersion can be fine tuned externally using a pair of parallel diffraction gratings such as shown in Fig. 15.15. When a chirped pulse is incident onto the bottom diffraction grating at point P_1, the $f - \Delta f$ and $f + \Delta f$ components are diffracted at different angles and they reach points P_2 and P_3 on the upper grating, respectively. If the two gratings are identical, the beams diffracted from the top grating will merge in parallel. There is, however, a path difference between the two component beams. The $f - \Delta f$ component travels the extra distance a–b–c, as shown in the figure, and incurs the phase delay that is necessary to sharpen the chirped pulse. With this arrangement it is even possible to trim the frequency components by placing a spatial filter in between and parallel to the gratings.

It should be noted that in our theoretical approach, the total amount of the spread due to SPM for the transmission length L was used to reshape the pulse by the GVD at the end of the transmission. In reality, however, the SPM and GVD are interacting during the transmission. The spectrum spread causes the reshaping of the pulse by the GVD. The reshaped pulse then has different amounts of SPM and is again corrected by the GVD in an iterative process. The interaction between the SPM and GVD should be considered separately for each small section. It is only after solving the differential equations presented later in Sections 15.9 and 15.10 that the rigorous answer can be obtained.

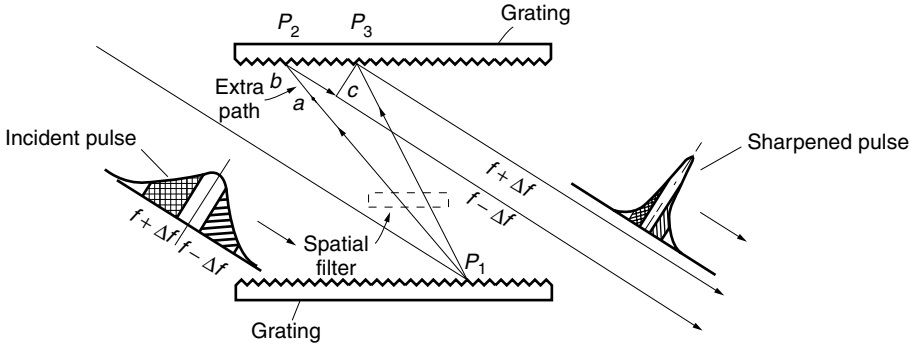


Figure 15.15 A diffraction grating used for sharpening a chirped pulse.

15.9 DIFFERENTIAL EQUATION OF THE ENVELOPE FUNCTION OF THE SOLITONS IN THE OPTICAL FIBER

The prime target of this section is the derivation of a differential equation for the envelope function $u(t, z)$ of the soliton pulse in an optical fiber using the slowly varying envelope approximation [18–24].

A step-by-step approach will be used to derive the results. The following simplifications will be made in order to render the expressions tractable:

1. The envelope is a slowly varying function of t and z compared to the period and the wavelength of the light carrier, respectively.
2. The contribution of the nonlinear portion of the refractive index is small compared to the linear part and $\alpha_1|E|^2$ is treated like a constant during the derivation of the envelope function.
3. The direction of polarization does not change during transmission and scalar solutions that have already been obtained for the linear case in Section 11.2.2 can be employed.
4. The frequency shift caused by the SPM is much smaller than the carrier frequency of the light.
5. The contribution of the higher harmonics of the carrier frequency will not be considered.

The nonlinear wave equation

$$\nabla^2\mathbf{E} + k^2(\epsilon_1 + \alpha|E|^2)\mathbf{E} = 0 \tag{15.157}$$

will be solved. Assume a scalar solution

$$E(x, y, z, t) = F(x, y)u(z, t)e^{-j2\pi f_c t + j\beta_c z} \tag{15.158}$$

E is one of the components of the vector $\mathbf{E}$. Figure 15.16a shows $\mathbf{E}$ with respect to time t for a fixed point. $F(x, y)$ is the transverse field distribution inside the optical fiber. $u(z, t)$ is the variation of the envelope of the light pulse as it propagates along the fiber. The exponential term represents the carrier of the pulse, where f_c is the carrier frequency and β_c is the propagation constant at the carrier frequency in the linear fiber with n_1. The main focus of this section is finding the dynamics of the envelope function $u(z, t)$ that satisfies Eq. (15.157).

The Fourier transform $\tilde{E}(x, y, z, f)$ of $E(x, y, z, t)$ is

$$\tilde{E}(x, y, z, f) = \int_{-\infty}^{\infty} E(x, y, z, t)e^{-j2\pi f t}\, dt \tag{15.159}$$

and the inverse Fourier transform is

$$E(x, y, z, t) = \int_{-\infty}^{\infty} \tilde{E}(x, y, z, f)e^{j2\pi f t}\, df \tag{15.160}$$

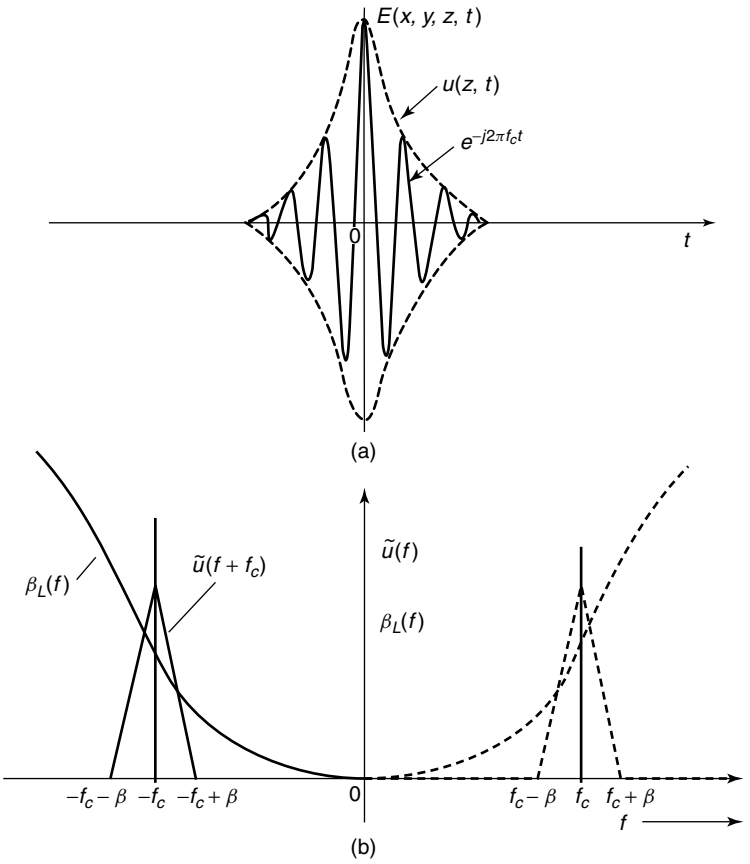


Figure 15.16 Soliton pulse in the time and frequency domains. (a) Soilton pulse $E(x, y, z, t)$. (b) Spectrum of the soliton pulse.

Inserting Eq. (15.158) into (15.159) gives

$$\tilde{E} = F(x, y)\tilde{u}(z, f + f_c)e^{j\beta_c z} \tag{15.161}$$

Inserting Eq. (15.160) into (16.157) proves that $\tilde{E}$ satisfies

$$\nabla^2\tilde{E} + k^2(\epsilon_1 + \alpha|E|^2)\tilde{E} = 0 \tag{15.162}$$

As stated in simplication (2), $\alpha|E|^2$ will be treated as a constant $\delta\epsilon$ even though it is a function of x, y, and z. Equation (15.161) is inserted into Eq. (15.162) to obtain

$$\left(\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2}\right)\tilde{u} + \left(\frac{\partial^2\tilde{u}}{\partial z^2} + j2\beta_c\frac{\partial\tilde{u}}{\partial z} - \beta_c^2\tilde{u}\right)F + k^2(\epsilon_1 + \delta\epsilon)F\tilde{u} = 0 \tag{15.163}$$

When the variation $\tilde{u}$ with distance is much slower than that of the carrier, the approximation

$$\frac{\partial^2 \tilde{u}}{\partial z^2} = 0 \tag{15.164}$$

holds.

Dividing Eq. (15.163) by $\tilde{u}F$ gives

$$\underbrace{\frac{1}{F}\left(\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2}\right) + k^2(\epsilon_1 + \delta\epsilon)}_{\beta_{\mathrm{NL}}^2(f)} + \underbrace{\frac{1}{\tilde{u}}\left(2j\beta_c \frac{\partial u}{\partial z} - \beta_c^2 \tilde{u}\right)}_{-\beta_{\mathrm{NL}}^2(f)} = 0 \tag{15.165}$$

$$= 0 \tag{15.166}$$

The first block is a function of x and y and the second block is a function of z only. For the partial differential equation to be satisfied at any point in space, each block has to be equal to a constant or a function that includes neither x, y, nor z. Using $\beta_{\mathrm{NL}}^2(f)$ and $-\beta_{\mathrm{NL}}^2(f)$ as the function gives

$$\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + [k^2(\epsilon_1 + \delta\epsilon) - \beta_{\mathrm{NL}}^2(f)]F = 0 \tag{15.167}$$

$$2j\beta_c \frac{\partial \tilde{u}}{\partial z} + [\beta_{\mathrm{NL}}^2(f) - \beta_c^2]\tilde{u} = 0 \tag{15.168}$$

The problem reduces to solving the simultaneous equations (15.167) and (15.168). The former equation gives the transverse field distribution in the optical fiber with a nonlinear refractive index when the frequency of the incident light is f. The latter equation is the partial differential equation describing the dynamics of the envelope of the light pulse. The procedure for finding the solution is as follows. We obtain $\beta_{\mathrm{NL}}^2(f)$ from Eq. (15.167) and use it in Eq. (15.168) to find the envelope function.

The solution for an optical fiber with a linear refractive index $\alpha = 0$ has already been obtained by Eq. (11.136) as

$$F(x, y) = J_0(Kr) \tag{15.169}$$

where

$$K^2 = k^2 n_1^2 - \beta_c^2$$
$$r^2 = x^2 + y^2$$

Next, we obtain the propagation constant $\beta_{\mathrm{NL}}(f)$ in an optical fiber with a nonlinear refractive index. If the contribution of the nonlinear part of the refractive index is much smaller than that of the linear part, we may be able to modify the solutions of the linear case to fit the solutions for the nonlinear case. Perturbation theory is used to achieve the fitting between the two situations.

The propagation constant of a guide that is perturbed by a small amount of change $\delta\epsilon$ is obtained by perturbation theory in Appendix C of Volume II as

$$\beta_{\mathrm{NL}}(f) = \beta_{\mathrm{L}}(f) + \Delta\beta \tag{15.170}$$

where

$$\Delta\beta = \frac{k^2 \iint n_1 \delta n |F(x, y)|^2 \, dx \, dy}{\beta_c \iint |F(x, y)|^2 \, dx \, dy} \tag{15.171}$$

and where $F(x, y)$ is given by Eq. (15.169) for the optical fiber and a graph of β_L is shown in Fig. 15.17. Thus, the solution of Eq. (15.167) for a continuous wave at f has in principle been solved.

Before going any further, values associated with β are summarized:

$\beta_{NL}(f)$ is the propagation constant in a nonlinear fiber at frequency f.
$\beta_L(f)$ is the propagation constant in a linear fiber at frequency f.
β_c is the propagation constant in a linear fiber at the carrier frequency f_c.

Next, the solution of Eq. (15.168) will be sought.

If the slowly varying envelope approximation is employed and $\partial\tilde{u}/\partial z$ is small, the second term in Eq. (15.168) is also small, and

$$\beta_{NL}(f) \doteqdot \beta_c$$

Thus, we have

$$\beta_{NL}^2(f) - \beta_c^2 \doteqdot 2\beta_c[\beta_{NL}(f) - \beta_c] \tag{15.172}$$

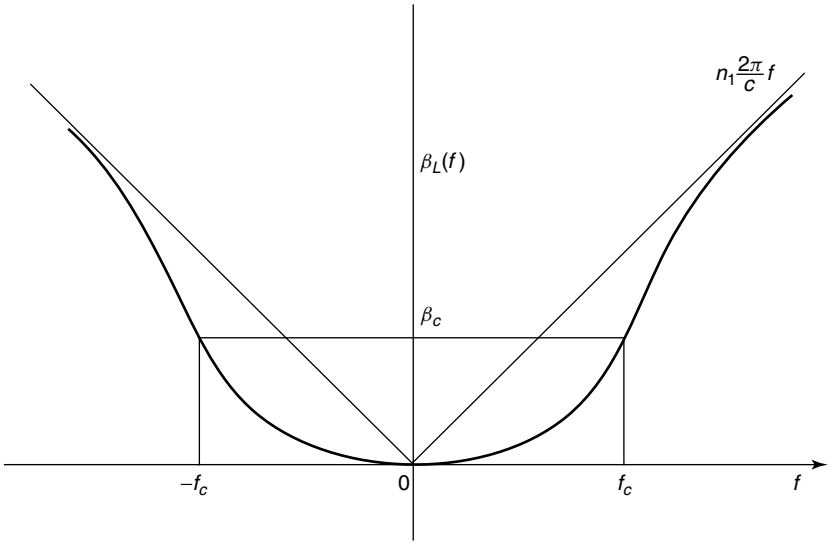


Figure 15.17 Propagation constant $\beta(f)$ of the HE_{11} mode in a step index fiber.

Inserting Eq. (15.172) into (15.168) gives

$$\frac{\partial \tilde{u}}{\partial z} - j[\beta_{\mathrm{NL}}(f) - \beta_c]\tilde{u} = 0 \tag{15.173}$$

Using Eq. (15.170) in (15.173) gives

$$\frac{\partial \tilde{u}}{\partial z} - j[\beta_{\mathrm{L}}(f) - \beta_c + \Delta\beta]\tilde{u} = 0 \tag{15.174}$$

In order to perform the inverse Fourier transform with respect to f, every term in Eq. (15.174) including $[\beta_{\mathrm{L}}(f)\tilde{u}]$ has to be known.

The Taylor series expansion is used to find $\beta_{\mathrm{L}}(f)$. As seen in Fig. 15.16b, $\tilde{u}$ is nonzero only in a small range (<100 GHz) near the carrier frequency (200 THz) and the value of $\beta_{\mathrm{L}}(f)\tilde{u}$ is needed only in this range. Thus, the approximation by the Taylor series expansion is justifiable. The general expression of the Taylor series expansion at $f = f_0$ is

$$\beta_{\mathrm{L}}(f) \doteqdot \beta_{\mathrm{L}}(f_o) + (f - f_o)\beta'(f_o) + \tfrac{1}{2}(f - f_o)^2\beta''(f_o)$$

where primes represent the derivative with respect to f. Our convention has selected the negative side of the spectral region $e^{-j2\pi f_c t}$ rather than $e^{j2\pi f_c t}$ and as shown in Fig. 15.16b, the spectrum is clustered around $-f_c$. Let

$$f_o = -f_c$$

the Taylor series then becomes

$$\beta_{\mathrm{L}}(f) = \beta_{\mathrm{L}}(-f_c) + (f + f_c)\beta'(-f_c) + \tfrac{1}{2}(f + f_c)^2\beta''(-f_c) \tag{15.175}$$

As shown in Fig. 15.17, the curve of $\beta_{\mathrm{L}}(f)$ of the dominant mode in a single mode step index fiber is an even function of f, and

$$\left.\begin{aligned} \beta_{\mathrm{L}}(-f_c) &= \beta_{\mathrm{L}}(f_c) = \beta_c \\ \beta'_{\mathrm{L}}(-f_c) &= -\beta'_{\mathrm{L}}(f_c) \\ \beta''(-f_c) &= \beta''_{\mathrm{L}}(f_c) \end{aligned}\right\} \tag{15.176}$$

Insertion of Eqs. (15.175) and (15.176) into (15.174) gives

$$\frac{\partial \tilde{u}}{\partial z} + j(f + f_c)\tilde{u}\frac{d\beta_{\mathrm{L}}}{df} - j\frac{1}{2}(f + f_c)^2\tilde{u}\frac{d^2\beta_{\mathrm{L}}}{dt^2} - j\Delta\beta\tilde{u} = 0 \tag{15.177}$$

Next, the inversion of Eq. (15.177) into the time domain is sought. Note that

$$u(z, t) = \int_{-\infty}^{\infty} \tilde{u}(z, f')e^{j2\pi f' t}\, df'$$

Putting $f' = f + f_c$,

$$u(z, t) = \int_{-\infty}^{\infty} \tilde{u}(z, f + f_c) e^{j2\pi(f+f_c)t}\, d(f + f_c) \tag{15.178}$$

Taking the derivative of both sides of Eq. (15.178) with respect to t gives

$$\frac{\partial u(z, t)}{\partial t} = \int_{-\infty}^{\infty} j2\pi(f + f_c)\tilde{u}(z, f + f_c) e^{j2\pi(f+f_c)t}\, d(f + f_c) \tag{15.179a}$$

Similarly,

$$\frac{\partial^2 u(z, t)}{\partial t^2} = \int_{-\infty}^{\infty} -(2\pi(f + f_c))^2\tilde{u}(z, f + f_c) e^{j2\pi(f+f_c)t}\, d(f + f_c) \tag{15.179b}$$

Using Eqs. (15.178), (15.179a) and (15.179b), Eq. (15.177) can be transformed back to obtain the expression in the time domain:

$$\frac{\partial u}{\partial z} + \beta^{(1)}\frac{\partial u}{\partial t} + j\frac{1}{2}\beta^{(2)}\frac{\partial^2 u}{\partial t^2} - j\Delta\beta u = 0 \tag{15.180}$$

where

$$\frac{\partial \beta}{\partial f} = 2\pi\beta^{(1)}$$

$$\beta^{(1)} = \frac{d\beta}{d\omega} = v_g^{-1} = \tau$$

$$\beta^{(2)} = \frac{d^2\beta}{d\omega^2}$$

Equation (15.180) is the differential equation that controls the dynamics of the envelope function. The third term is attributed to the dispersion $\beta^{(2)}$ of the fiber and the fourth term to the nonlinearity $\Delta\beta$ of the fiber. In the event that these two terms cancel each other or both of them are absent, Eq. (15.180) becomes

$$\frac{\partial u}{\partial z} + \beta^{(1)}\frac{\partial u}{\partial t} = 0 \tag{15.181}$$

with a general solution of

$$u = g(t - \beta^{(1)}z) \tag{15.182}$$

The solution is a traveling wave with an arbitrary envelope function g at velocity $1/\beta^{(1)}$.

This suggests a possibility of simplifying Eq. (15.180) by using a change of variables, the so-called retarded (time) frame [20]:

$$t' = t - \beta^{(1)}z \tag{15.183}$$

$$z' = z \tag{15.184}$$

The change of variables provides

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial t'} \tag{15.185}$$

$$\frac{\partial u}{\partial z} = -\beta^{(1)} \frac{\partial u}{\partial t'} + \frac{\partial u}{\partial z'} \tag{15.186}$$

The change of variables in the partial derivative is given in the boxed note. The first two terms of Eq. (15.180) are replaced by Eq. (15.185) and (15.186) and, hence, the differential equation Eq. (15.180) finally becomes

$$\frac{\partial u}{\partial z'} + \frac{1}{2} j\beta^{(2)} \frac{\partial^2 u}{\partial t'^2} - j\Delta\beta u = 0 \tag{15.187}$$

Let us see if our change of variables was successful. With $\beta^{(2)} = \Delta\beta = 0$, Eq. (15.187) becomes

$$\frac{\partial u}{\partial z'} = 0$$

The solution of this partial differential equation is $u = g(t')$ which is independent of z' and with Eq. (15.183) it is verified that

$$u = g(t - \beta^{(1)} z).$$

What remains now is the calculation of the perturbation term of Eq. (15.171).

The general formulas for changing the variables in the partial differentiation of composite functions are

$$t' = f(t, z)$$

$$z' = g(t, z)$$

$$\frac{\partial u(t,z)}{\partial t} = \frac{\partial u(t',z')}{\partial t'} \frac{\partial t'}{\partial t} + \frac{\partial u(t',z')}{\partial z'} \frac{\partial z'}{\partial t} \tag{15.188}$$

$$\frac{\partial u(t,z)}{\partial z} = \frac{\partial u(t',z')}{\partial t'} \frac{\partial t'}{\partial z} + \frac{\partial u(t',z')}{\partial z'} \frac{\partial z'}{\partial z} \tag{15.189}$$

with

$$t' = t - \beta^{(1)} z$$

$$z' = z$$

$$\frac{\partial u(t,z)}{\partial t} = \frac{\partial u(t',z')}{\partial t'} + 0 \tag{15.185}$$

$$\frac{\partial u(t,z)}{\partial z} = -\beta^{(1)} \frac{\partial u(t',z')}{\partial t'} + \frac{\partial u(t',z')}{\partial z'} \tag{15.186}$$

$$n = n_1 + n_2 I \tag{15.190}$$

where n_2 is the *nonlinear-index coefficient*.* In terms of the amplitude (see Section 2.4), Eq. (15.190) becomes

$$n_1 + \delta n = n_1 + n_2 \frac{1}{\eta} |uF(x, y)|^2 \tag{15.191}$$

where η is the intrinsic impedance of the fiber core. Inserting Eq. (15.191) into (15.171) gives

$$\Delta\beta = \kappa |u|^2 \tag{15.192}$$

where

$$\kappa = \frac{(n_2/\eta)k \iint |F(x, y)|^4 dx dy}{\iint |F(x, y)|^2 dx dy} \tag{15.193}$$

and where the approximation

$$\beta_c \doteq n_1 k$$

was used.

Inserting Eq. (15.192) into (15.187) gives

$$\frac{\partial u}{\partial z'} + \frac{1}{2} j\beta^{(2)} \frac{\partial^2 u}{\partial t'^2} - j\kappa |u|^2 u = 0 \tag{15.194}$$

The sign of the second term significantly influences the solution of the differential equation. When the dispersion of the fiber is anomalous and consequently $\beta^{(2)}$ is negative,

$$\beta^{(2)} = -|\beta^{(2)}| \tag{15.195}$$

then Eq. (15.194) becomes

$$\frac{\partial u}{\partial z'} - j\frac{1}{2} |\beta^{(2)}| \frac{\partial^2 u}{\partial t'^2} - j\kappa |u|^2 u = 0 \tag{15.196}$$

In order to simplify Eq. (15.196), let

$$\begin{aligned} t' &= \tau_0 T \\ z' &= z_0 Z \\ u &= \mu_0 U \end{aligned} \tag{15.197}$$

where τ_0, z_0, and μ_0 are scaling factors. Inserting Eq. (15.197) into (15.196) results in

$$j\frac{\partial U}{\partial Z} + \frac{|\beta^{(2)}|}{2} \frac{z_0}{\tau_0^2} \frac{\partial^2 U}{\partial T^2} + \kappa \mu_0^2 z_0 |U|^2 U = 0 \tag{15.198}$$

* n_2 rather than α is more widely used in the literature.

If we set conditions:

$$\begin{aligned} |\beta^{(2)}|\frac{z_0}{\tau_0^2} &= 1 \\ \kappa\mu_0^2 z_0 &= N^2 \end{aligned} \tag{15.199}$$

or equivalently,

$$z_0 = \frac{\tau_0^2}{|\beta^{(2)}|} \tag{15.200}$$

$$\mu_0^2 = \frac{1}{\kappa z_0} N^2 \tag{15.201}$$

then Eq. (15.198) finally becomes

$$j\frac{\partial U}{\partial Z} + \frac{1}{2}\frac{\partial^2 U}{\partial T^2} + N^2|U|^2 U = 0 \tag{15.202}$$

The nonlinear differential equation (15.202) with Z and T exchanged is known as the *nonlinear Schrödinger equation*.

15.10 SOLVING THE NONLINEAR SCHRÖDINGER EQUATION

This section will be devoted to solving the nonlinear Schrödinger equation (H. Shimodahira, *private communication*):

$$j\frac{\partial U}{\partial Z} + \frac{1}{2}\frac{\partial^2 U}{\partial T^2} + N^2|U|^2 U = 0 \tag{15.203}$$

Unlike many other nonlinear differential equations, an exact solution is obtainable.

Assume a solution

$$U(Z, T) = f(T)e^{jBZ} \tag{15.204}$$

where $f(T)$ is a real function and B is a positive constant. Inserting Eq. (15.204) into (15.203) gives

$$-Bf + \tfrac{1}{2}f'' + N^2 f^3 = 0 \tag{15.205}$$

Equation (15.205) is multiplied by f' and then integrated with respect to T to obtain

$$-\frac{1}{2}Bf^2 + \frac{1}{4}(f')^2 + \frac{N^2}{4}f^4 = h(z) \tag{15.206}$$

For now, let the initial condition be set as

$$f(T) = f'(T) = 0 \quad \text{as } T \to \pm\infty \tag{15.207}$$

which will be discussed in more detail later. Therefore, Eq. (15.206) becomes

$$-2Bf^2 + (f')^2 + N^2 f^4 = 0 \tag{15.208}$$

Solving for f', we have

$$\frac{df}{dT} = Nf\sqrt{\frac{2}{N^2}B - f^2} \tag{15.209}$$

Substituting

$$f = \frac{\sqrt{2B}}{N}\sin\theta \tag{15.210}$$

into Eq. (15.209) gives

$$\frac{df}{dT} = \frac{2B}{N}\sin\theta\cos\theta \tag{15.211}$$

Next, df/dT will be expressed in terms of $d\theta/dT$. From Eq. (15.210), df/dT is written as

$$\frac{df}{dT} = \frac{\sqrt{2B}}{N}\cos\theta\frac{d\theta}{dT} \tag{15.212}$$

Combining Eqs. (15.211) and (15.212) gives

$$\frac{2B}{N}\sin\theta\cos\theta = \frac{\sqrt{2B}}{N}\cos\theta\frac{d\theta}{dT}$$

$$\sqrt{2B}\sin\theta = \frac{d\theta}{dT} \tag{15.213}$$

$$\int \frac{d\theta}{\sin\theta} = \int \sqrt{2B}dT \tag{15.214}$$

Integrating the left-hand side of Eq. (15.214) gives

$$\int \frac{d\theta}{\sin\theta} = \int \frac{1}{\cos^2(\theta/2)\tan(\theta/2)}d(\theta/2) = \ln|\tan\theta/2|$$

Thus, Eq. (15.214) becomes

$$\ln|\tan\frac{\theta}{2}| = \sqrt{2B}(T - T_0)$$

$$\tan\frac{\theta}{2} = e^{\sqrt{2B}(T-T_0)} \tag{15.215}$$

Eq. (15.210) becomes

$$f = \frac{\sqrt{2B}}{N}\sin\theta = \frac{\sqrt{2B}}{N}\frac{2\tan\theta/2}{1+\tan^2\theta/2} \tag{15.216}$$

Inserting Eq. (15.215) into (15.216), f becomes

$$f = \frac{\sqrt{2B}}{N} \frac{2}{e^{-\sqrt{2B}(T-T_0)} + e^{\sqrt{2B}(T-T_0)}}$$
$$= \frac{\sqrt{2B}}{N} \text{ sech } \sqrt{2B}(T - T_0) \tag{15.217}$$

Inserting Eq. (15.217) into (15.204), $U(Z, T)$ becomes

$$U(Z, T) = \frac{\sqrt{2B}}{N} \text{ sech } [\sqrt{2B}(T - T_0)] \cdot e^{jBZ} \tag{15.218}$$

By letting $\sqrt{2B}/N = \eta$, the final solution is

$$U(Z, T) = \eta \text{ sech } [(N\eta)(T - T_0)] \cdot e^{j\eta^2 N^2 Z/2} \tag{15.219}$$

15.11 FUNDAMENTAL SOLITON

The solution given by Eq. (15.219) with $N^2 = 1$ is called the *fundamental soliton*.

$$U(Z, T) = \text{sech } T \cdot e^{j(Z/2)} \tag{15.220}$$

where $B = \frac{1}{2}$ and $T_0 = 0$ were assumed.

Using Eqs. (15.158), (15.183), (15.184), and (15.197), the full expression of the fundamental soliton $E_1(x, y, z, t)$ becomes

$$E_1(x, y, z, t) = F(x, y)\mu_0 \text{ sech } \left[(t - \beta^{(1)}z)/\tau_0\right] \cdot e^{-j2\pi f_c t + j(\beta_c + 1/2z_0)z} \tag{15.221}$$

It is quite striking to note that the width of the pulse or shape of the pulse is independent of time and distance despite the medium being nonlinear. This feature of the lightwave soliton is highly valued for ultra-long-distance fiber-optic communication. This special feature of invariance of the pulse shape for the fundamental soliton (true only with the *fundamental* soliton) may be explained using Eq. (15.198). The second term, which is the second derivative of the pulse shape, is primarily due to the GVD and the third term is due to the SPM. $N^2 = 1$, which is the condition for the fundamental soliton, means that the effect of the SPM is just enough to cancel that of the GVD. (As mentioned earlier, the envelope function with respect to t' is bell shaped and concave downward; hence $\partial^2 U/\partial T^2$ is negative near the center of the pulse.)

In other words, the proper combination of the pulsewidth (which influences the GVD) and the height of the pulse (which influences the SPM) is necessary to generate the fundamental soliton.

On the other hand, when $N^2 \neq 1$, these two terms do not counterbalance each other and the pulse shape is no longer constant but is distorted with distance. An amazing

fact is that the Nth order (N is an integer) soliton periodically recovers its original pulse shape with a period of $z_c = \frac{\pi}{2} z_0$ [20,21,23]. z_c is called *soliton period* which is immediately obtained from Eq. (15.200) as

$$z_c = \frac{\pi}{2} \frac{\tau_0^2}{\beta^{(2)}} \tag{15.222}$$

Generally speaking, as mentioned earlier, the law of superposition does not hold in nonlinear differential equations. The shape of the overlapped solitons is not simply the sum of the two amplitudes. Unlike this prediction, the shape of the pulse recovers after the collision.

The collision of the solitons displays a peculiar property [25,26]. The amplitude of the resultant wave depends on the direction of the collision. The resultant amplitude when both are propagating in the same direction is different from that when they are

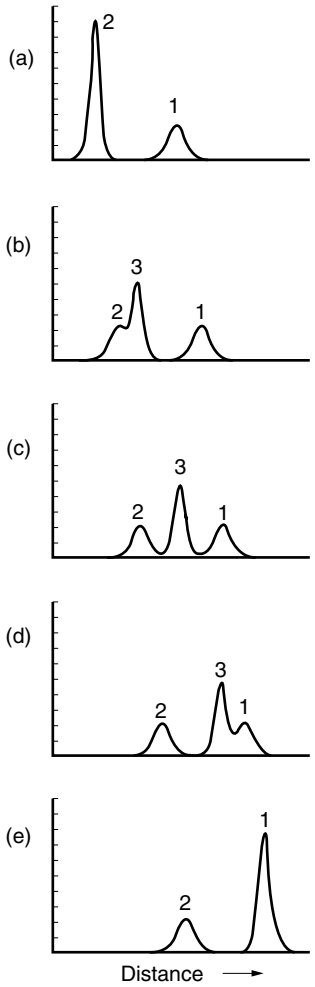


Figure 15.18 Sequence of collision of solitons 1 and 2. (After R. Hirota [26].)

propagating in opposite directions. The distribution of the amplitude in the vicinity of the collision is shown in Fig. 15.18. As the two solitons approach the point of collision, the amplitude is increased but is not the arithmetic sum of the two, and after the collision, they each recover to their former shapes.

Figure 15.18 shows the transition of shapes of two solitons at the time of collision. Soliton 2 with higher intensity and higher speed catches up with soliton 1 of lower intensity in Fig. 15.18a. When soliton 2 arrives within interacting distance, it is split between solitons 2 and 3 in Fig. 15.18b and 15.18c. As soliton 3 approaches soliton 1 in Fig. 15.18d, the energy of soliton 3 is absorbed into soliton 1, and soliton 1 grows in Fig. 15.18e. The exchange of energy of soliton 3 plays an important role at the time of collision. Collision behavior is an important consideration in establishing a soliton-based communication system.

Recall that in the process of solving the nonlinear Schrödinger equation, Eq. (15.203), the rather impractical initial condition that the right-hand side of Eq. (15.206) be zero was used. By using Lax's inverse scattering method [27], the problem of the initial conditions can be resolved.

15.12 PULSEWIDTH AND POWER TO GENERATE A FUNDAMENTAL SOLITON

The pulsewidth and power required to establish the fundamental soliton in an optical fiber will be calculated [20,21,28].

First, the full width at half maximum (FWHM) of such a soliton is calculated. The value of time that gives the half power is from Eq. (15.220)

$$\begin{aligned} \frac{1}{2} &= \operatorname{sech}^2\left(\frac{\tau_s/2}{\tau_0}\right) \\ \left(\frac{\tau_s/2}{\tau_0}\right) &= 0.88 \end{aligned} \tag{15.223}$$

In terms of τ_s of the full width at half maximum (from $-\tau_s/2$ to $\tau_s/2$), τ_0 can be expressed as

$$\frac{\tau_s}{2} = 0.88\tau_0 \tag{15.224}$$

$$\tau_0 = \frac{\tau_s}{1.76} \tag{15.225}$$

Next, the peak power required to establish the fundamental soliton will be derived.

The power transmitted in the fiber is obtained by integrating the light intensity across the cross-section of the fiber as

$$P = \frac{1}{\eta}\iint |u|^2 |F(xy)|^2 dx dy \tag{15.226}$$

The peak power is when the sech function in (15.220) is unity. From Eq. (15.197)

$$P_s = \frac{1}{\eta}|\mu_0|^2 \iint |F(x, y)|^2 dx dy \tag{15.227}$$

Equations (15.200) and (15.201) are the necessary conditions for establishing the soliton.

Inserting Eqs. (15.193) and (15.201) into (15.227) gives

$$P_s = \frac{A_{\text{eff}}}{z_0 n_2 k} \tag{15.228}$$

where

$$A_{\text{eff}} = \frac{\left[\iint |F(x, y)|^2 dx dy\right]^2}{\iint |F(x, y)|^4 dx dy} \tag{15.229}$$

A_{eff} is called *effective core area* and in fact you will see that in the case when $F(x, y)$ is a constant F_0, A_{eff} becomes exactly the geometric core area. Using Eqs. (15.200) and (15.225), Eq. (15.228) is further rewritten as

$$P_s = 3.1\frac{|\beta^{(2)}|}{\Gamma \tau_s^2} \tag{15.230}$$

where a new nonlinearity parameter

$$\Gamma = \frac{n_2 k}{A_{\text{eff}}} \tag{15.231}$$

is introduced.

$\beta^{(2)}$ is rewritten in terms of the dispersion parameter D using Eq. (15.152) to obtain the final result for the peak power P_s needed to generate the fundamental soliton pulse

$$P_s = 0.775\frac{\lambda^3 |D| A_{\text{eff}}}{\pi^2 C n_2 \tau_s^2} \tag{15.232}$$

In order to calculate Eq. (15.232), several units are combined. If the following, practical units are used:

P_s	Watts
D	ps/(km · nm)
A_{eff}	μm^2
τ_s	ps
λ	μm
c	3×10^8 m
$n_2 = 3.18 \times 10^{-20}$	m^2/W (for silica fiber)

Then Eq. (15.232) can be further rewritten

$$\tau_s \sqrt{P_s} = 0.09\lambda^{3/2}\sqrt{DA_{\text{eff}}} \tag{15.233}$$

Thus, a simple expression relating the pulsewidth and the pulse height of the fundamental soliton for given physical parameters is derived. The product of τ_s and $\sqrt{P_s}$ is constant for a given fiber and wavelength of the carrier.

Example 15.6

(a) Find the peak power needed to generate the fundamental soliton in a single-mode fiber with a 60-μm^2 effective cross-sectional core area, a 20-ps FWHM, and a 1.55-μm wavelength.

(b) The above system is applied to a soliton-based optical communication system with a bit rate (number of bits per second) at 10 Gb/s as shown in Fig. 15.19. Calculate the required average light power of the transmitter laser diode. Assume that the "0" bit (absence of the light pulse), and the "1" bit (presence of the light pulse) occur in equal numbers. The optical fiber is assumed lossless.

Solution

(a)

$$\tau_s \sqrt{P_s} = 0.09\lambda^{3/2}\sqrt{DA_{\text{eff}}}$$

$$P_s = \left(0.09(1.55)^{3/2}\sqrt{(17)(60)} \cdot \frac{1}{20}\right)^2$$

$$= 77 \text{ mW} \tag{15.234}$$

(b) The energy of each pulse at $z = 0$ is from Eqs. (15.197) and (15.220)

$$P(t) = P_s \text{ sech}^2\left(\frac{t}{\tau_0}\right) \tag{15.235}$$

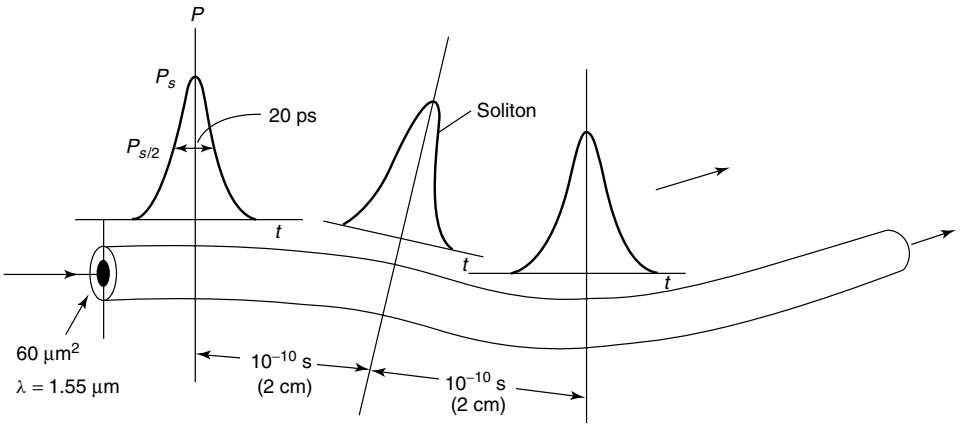


Figure 15.19 Soliton-based optical communication.

The total energy w in each pulse is

$$w = \int_{\infty}^{\infty} P(t)dt = \tau_0 P_s \left[\tanh^{-1}\left(\frac{t}{\tau_0}\right)\right]_{-\infty}^{\infty}$$
$$= 2\tau_o P_s \tag{15.236}$$

Inserting Eqs. (15.225) and (15.234) into (15.236) gives

$$w = 2\left(\frac{20 \times 10^{-12}}{1.76}\right)(0.077) = 1.75 \text{ pJ}$$

The average power with one half of the pulses zero is

$$P_s^a = w\frac{R}{2} = (1.75 \times 10^{-12})\frac{10^{10}}{2}$$
$$= 8.8 \text{ mW} \qquad \square$$

Figure 15.20a shows a block [16,29–31] diagram of an experiment performed on 3000 km of soliton transmission fiber. The fiber consisted of a total of 91 segments connected in series. Each segment was a combination of 33 km of fiber and an erbium-doped fiber amplifier. With a 40-Gb/s pulse, a BER (bit error rate in Section 16.5.9) better than 10^{-10} was obtained.

Figure 15.20b shows an experimental setup that can simulate the long-distance transmission of a soliton. The soliton circles around a three-segment loop. Each segment has 25 km of fiber and an erbium-doped fiber amplifier. Once the pulse is inside the loop, the soliton circles around as many times as desired. For instance, in order to mimic the experiment of the 3000-km fiber in Fig. 15.20a, the soliton in Fig. 15.20b has to circle around the loop 40 times, or approximately 15 ms after the injection of the first soliton.

Figures 15.20c and 15.20d show a pulsetrain at the rate of 10 Gb/s after 50 (250 s after injection) and 180 (15 min. after injection) million kilometers, respectively.

15.13 EVER-EXPANDING SOLITON THEORIES

Applications of the theory of soliton waves are not restricted to the analysis of the lightwave solitons. Applications in connection with solitons have been expanding to fields never before imagined. Some of these applications are introduced here.

The fact that a soliton wave preserves its wave shape even after a collision with another soliton wave intrigued many theoretical physicists. The soliton theories have been used to describe a localized field associated with matter in theoretical physics.

Plasma is highly nonlinear as well as dispersive and fosters an ideal breeding ground for plasma solitons. Plasma solitons can establish a high-temperature plasma, which nuclear physicists are trying to use as a means of triggering nuclear fusion [32].

Solid-state physicists want to use the soliton to analyze the propagation of a dislocation in a crystal. Once a dislocation starts in a crystal lattice, it propagates inside the crystal without stopping for some distance. The propagation of the dislocation [33]

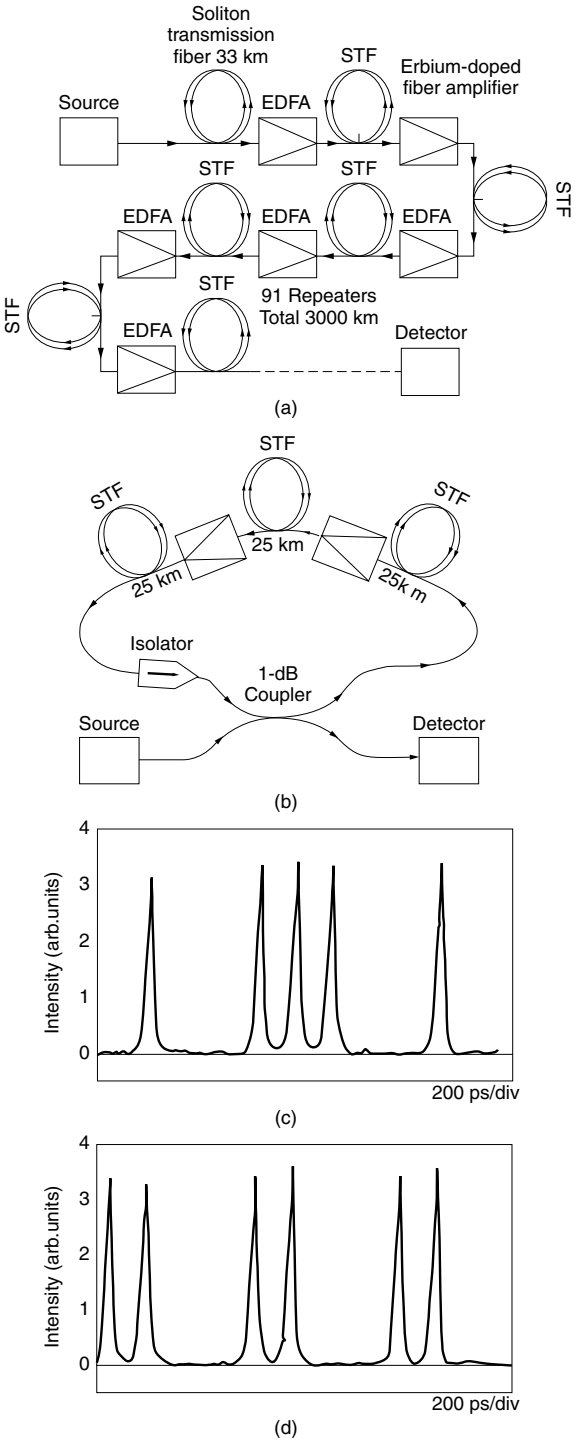


Figure 15.20 Arrangements for soliton experiments. (a) A 3000-km soliton transmission. (b) A loop to simulate long-distance transmission of a soliton. (c) Transmitted data pattern at 10 Gbs over 50 million kilometers of simulated length coded as ⟨0100111001⟩. (d) Transmitted data pattern at 10 Gbs over 180 million kilometers of simulated length coded as ⟨1100110011⟩. (After M. Nakazawa [16].)

can be modeled mathematically by the propagation of an overstressed oscillation along a chain of coil springs connecting small spherical masses. When the magnitude of the oscillation exceeds the Hooke's law limit of the spring, the differential equation to describe the motion becomes nonlinear. The differential equation of the displacement in this case becomes

$$\frac{\partial^2 \phi}{\partial x^2} - \frac{1}{c_s^2}\frac{\partial^2 \phi}{\partial t^2} = \frac{1}{\lambda^2}\sin\phi \tag{15.237}$$

where ϕ is the displacement of a small spherical mass,

$$c_s = \sqrt{\frac{K}{m}}a$$

where a is the lattice constant, m is the mass, and K is the compliance. Also

$$\lambda_0 = \frac{a^2}{2\pi}\sqrt{\frac{K}{A}}$$

where A is the potential energy of the spring. This type of nonlinear differential equation is known as the *sine–Gordon equation*. Its particular solution is

$$\phi = 4\tan^{-1}\left(\exp\left[\frac{1}{\lambda_0}\frac{(x - vt)}{\sqrt{1 - (v/c_s)^2}}\right]\right) \tag{15.238}$$

where v is the propagation speed of the dislocation. The curve of the solution looks like a moving cliff and is called the kink solution. This shape is suitable to represent the progression of the dislocation in the crystal. As a matter of fact, the derivative of this solution becomes identical with that of the soliton wave.

Another example is a phenomenon known as self-induced transparency. *Self-induced transparency* refers to the lossless transmission of an ultra-high-intensity pulse through a crystal. The leading edge of the pulse excites the atoms of the crystal. The trailing edge of the pulse then induces the excited atoms to emit by the process of stimulated emission. The pulse looks as if it is propagating without transmission loss. The multiple sine–Gordon equation

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = \pm\sum_{m=1}^{J}\frac{m}{J}\sin\left(\frac{m}{J}u\right) \tag{15.239}$$

is used as a means of analyzing the self-induced transparency [34] problem.

The usefulness of the soliton wave extends even to the fields of biology and ecology. Similarities between soliton waves propagating through liquid in a rubber tube and the pulsating blood flowing in a human artery have been used to study the dynamics of human blood circulation.

Marine biologists believe that chain reactions of ecological phenomena can be tackled by the same kind of differential equations as used for the light soliton. For example, a sudden decrease in the harvest of seaweed has taken place on the West Coast of the United States [26]. The sea otters have been overhunted for their furs.

A Japanese proverb says:
"A gust of wind brings the pail maker prosperity." Dust picked-up by a gust of wind causes many people to go blind. The blind play the Shamisen (a three-stringed Japanese guitar whose drum is made out of stretched cat skin) for their living. A greater demand for cat skins means that more rats can cheerfully make holes in wood pails, unhindered by the presence of cats. This progression brings unexpected prosperity to the pail makers.

The sea otter's feast is the sea urchin. The sea urchin's feast, in turn, is the seaweed. The marine biologists believe that the propagation of such a chain reaction is of the same nature as that of a soliton wave. The usefulness of soliton theory is limited only by the scientist's imagination.

A special tutorial issue [35] describes future prospects of soliton research including the prediction of a cavity soliton that survives indefinitely: a kind of stopped light that would make it possible to save a soliton in your pocket for a rainy day.

PROBLEMS

15.1 Find the expression for ϕ in the core layer of the three-layer slab optical guide shown in Fig. 15.8. What is the value of ϕ when $\tilde{\gamma}_0 = \tilde{\gamma}_2$?

15.2 Would it be possible for a negative shape pulse ("dark light") such as the one shown in Fig. P15.2 to possess the properties of a soliton?

15.3 A Gaussian pulse with the intensity

$$I(t) = I_0 e^{-(t/\tau)^2}$$

such as shown in Fig. P15.3 propagates inside a glass medium. Calculate the total frequency shift due to the self-phase modulation (SPM) after propagating a distance L in the glass with nonlinear index coefficient n_2 and refractive index n_1. The full width at half-maximum (FWHM) τ_s of the pulse is 200 ps. (Note τ is not the FWHM.)

$I_0 = 160$ MW/cm^2

$n_2 = 3.18 \times 10^{-20}$ m^2/W

$\lambda = 1.55$ μm

$L = 200$ m

$n_1 = 1.55$

15.4 The soliton (pulse width) × (amplitude) product was expressed in practical units by Eq. (15.233). Increase the accuracy of the coefficient by one more digit from 0.09 in the formula. Assume $n_2 = 3.18 \times 10^{-20}$ m^2/W.

15.5 With the soliton based communication presented in Example 15.6, what are the values of

(a) $|\beta^{(2)}|$ in (ps)^2/km.

(b) The nonlinear parameter Γ in W^{-1}/km.

(c) The soliton period z_c in km.

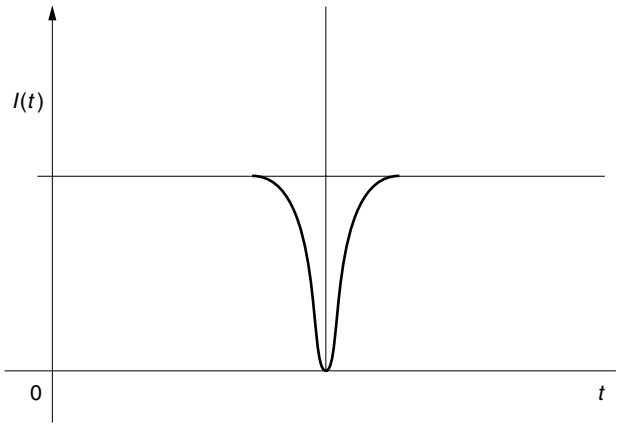


Figure P15.2 Dark light pulse.

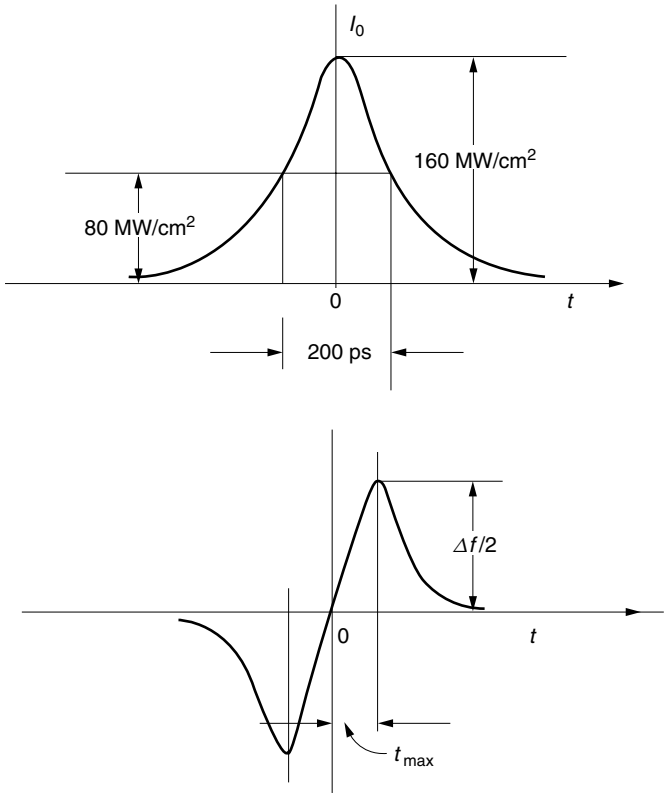


Figure P15.3 Gaussian pulse (top) and self-phase modulation (bottom) due to nonlinearity.

REFERENCES

1. G. I. Stegeman, "Nonlinear optical waveguides," *Tutorial Session, IOOC'89*, Kobe, Japan, July 18–21, 1989.
2. M. Shamonin, M. Lohmeyer, and P. Hertel, "Analysis of power-dependent switching between radiatively coupled planar waveguides," *J. Lightwave Technol.* **15**(6), 983–989, (1997).
3. A. A. Maradudin, "*s*-Polarized nonlinear surface polaritons," *Z. Phys. B. Condensed Matter* **41**, 341–344, (1981).
4. M. Abramowitz and I. A. Stegun (Eds.), *Handbook of Mathematical Functions*, Applied Mathematics Series, NBS, Washington DC, 1972.
5. Y.-F. Li, H. Anis, and K. Iizuka, "General stationary solutions of guided waves in a nonlinear film," *Nonlinear Opt.* **2**, 29–51 (1992).
6. Y.-F. Li, "Nonlinear wave propagation in planar structures," University of Toronto Lecture Notes, 1997.
7. U. Langbein, F. Lederer, H.-E. Ponath, and U. Trutschel, "Analysis of the dispersion relation of nonlinear slab-guided waves, Part II: Symmetrical configuration," *Appl. Phys. B* **38**, 263–268 (1985).
8. A. D. Boardman and P. Egan, "S-polarized waves in a thin dielectric film asymmetrically bounded by optically nonlinear media," *IEEE J. Quantum Electron.* **QE-21**(10), 1701–1713 (1985).
9. V. I. Korneev, N. N. Akhmediev, and Yu. V. Kuz'menko, "Degenerate nonlinear modes of a symmetric layered structure," *Izv. Vyssh. Uchebn. Zaved. Radiofiz.* **29**, 456–461 (1986).
10. L. Leine, C. Waechter, U. Langbein, and F. Lederer, "Evolution of nonlinear guided optical fields down a dielectric film with a nonlinear cladding," *J. Opt. Soc. Am.* **B5**(2), 547–558 (1988).
11. Y.-F. Li and K. Iizuka, "Unified dispersion equation for a general nonlinear waveguide and its applications to symmetrical waveguides," *Nonlinear Opt.* **17**, 99–126 (1997).
12. Y.-F. Li and K. Iizuka, "Dispersion equation for asymmetrical waves in a nonlinear film," *Nonlinear Opt.* **17**, 51–59 (1997).
13. H. A. Haus, "Modeling light into solitons," *IEEE Spectrum*, 48–53 (Mar. 1993).
14. H. A. Haus, "Optical fiber solitons, their properties and uses," *Proc. IEEE* **81**(7), 970–983 (1993).
15. D. J. Korteweg and G. de Vries, "On the change of form of long waves advancing in a rectangular channel and on a new type of long stationary wave," *Philos. Mag.* **39**(5), 422–443 (1895).
16. M. Nakazawa, "Soliton transmission in telecommunication networks," *IEEE Commun. Mag.*, 34–41 (Mar. 1994).
17. V. E. Zakharov and A. B. Shabat, "Exact theory of two dimensional self focussing and one-dimensional modulation of waves in nonlinear media," *Sov. Phys. JETP* **34**(1), 62–69 (1972).
18. A. Hasegawa and Y. Kodama, *Solitons in Optical Communications*, Clarendon Press, Oxford, 1995.
19. L. F. Mollenauer and R. H. Stolen, "The soliton laser," *Opt. Lett.* **9**(1), 13–15 (1984).
20. G. P. Agrawal, *Nonlinear Fiber Optics*, 3rd ed., Academic Press, New York, 2001.
21. G. P. Agrawal, *Fiber-Optic Communication Systems*, 2nd ed., Wiley, New York, 1997.
22. H. A. Haus, *Waves and Fields in Optoelectronics*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

23. V. E. Zakharov and S. Wabnitz (Ed.), *Optical Solitons: Theoretical Challenges and Industrial Perspectives*, Springer Verlag, Berlin and EDP Sciences, Cambridge, Massachusetts, 1999.
24. L. F. Mollenauer, J. P. Gordon, and M. N. Islam, "Soliton propagation in long fibers with periodically compensated loss," *IEEE J. Quantum Electronics* **QE-22**(1), 157–173 (1986).
25. M. Remoissenet, *Waves Called Solitons: Concepts and Experiments*, Springer Verlag, Berlin, 1999.
26. Y. Hirota, "Essay: Solitons from my own experiences," in *Soliton* (in Japanese), edited by Science Publishing Company, Tokyo, 1985.
27. G. L. Lamb, Jr., *Elements of Soliton Theory*, Wiley, New York, 1980.
28. A. Hasegawa, "Soliton-based optical fiber communication," (in Japanese), *Jpn. Appl. Opt.* **62**(3), 227–237 (1993).
29. H. Taga, M. Suzuki, N. Edagawa, Y. Yoshida, S. Yamamoto, S. Akiba, and H. Wakabayashi, "5 Gb/s optical soliton transmission experiment employing 91 cascaded Er-doped fiber amplifier repeaters," *ECOC'92*, Berlin, Germany, PD I.2, pp. 855–858, 1992.
30. M. Nakazawa, Y. Kimura, K. Suzuki, H. Kubota, T. Komukai, E. Yamada, T. Sugawa, E. Yoshida, T. Yamamoto, T. Imai, A. Sahara, H. Nakazawa, O. Yamauchi, and M. Umezawa, "Field demonstration of soliton transmission at 10 Gb/s over 2,500 km and 20 Gb/s over 1,000 km in the Tokyo metropolitan optical network," *IOOC'95*, PD.2–1, Hong Kong, 1995.
31. M. Matsumoto, Y. Akagi, and A. Hasegawa, "Propagation of solitons in fibers with randomly varying birefringence: effects of soliton transmission control," *J. Lightwave Technol.* **15**(4), 584–589 (1997).
32. Y. Ichikawa, "Plasma solitons and chaos," (in Japanese), *J. Phys. Soc. Jpn.* **47**(7), 531–538 (1992).
33. K. Kobayashi, "Sine–Gordon equation and its applications to solid state physics," (in Japanese), *Solid State Phy.* **11**, 477–491 (1976).
34. A. O. Barut (Ed.), *Nonlinear Equations in Physics and Mathematics*, Reidel Publishing, Boston 1978.
35. Special Issue on "Solitons," *Opt. and Photonics News*, **13**(2), 27–76 (2002).

16

COMMUNICATING BY FIBER OPTICS

The prime purpose of communication systems is to transfer data from point A to point B at a remote site. In the past this was accomplished primarily by using copper wires as a conduit for communication. In the present day, however, fibers are now much more often installed than copper wire. This is a result of many significant advantages of fiber-optic communications over its copper counterpart. Some main advantages include lower transmission loss and greater information capacity. These concepts are briefly summarized as follows:

1. Lower Transmission Loss. Fiber and copper wires are used to transmit telephone, television, and computer signals. For both fiber and copper wire communications, this signal level reduces with distance. Before the signal level is attenuated to a value that is lower than the noise level of the detector, the data line is terminated. The signal is then detected and amplified or just amplified in order to be retransmitted to the next repeater station.

The spacing between the repeater stations is one of the parameters used to characterize transmission systems. Long repeater spacings are desirable. While a typical repeater distance for copper wire transmission is every 10–50 kilometers, the optical fiber requires repeater stations at a distance of greater than a few hundred kilometers.

2. Greater Information Capacity. The speed at which data can be transferred (the bit rate) for copper coaxial cables, over a distance of 1 km, ranges only up to a few hundred megabits/second (Mb/s). In comparison, light pulses sent through optical fibers have bit rates of up to tens of gigabits/second (Gb/s). This is an improvement of about one hundredfold over copper transmission.

Furthermore, by making use of a special method of lightwave propagation called soliton waves in the optical fiber (see Section 15.5) combined with optical amplifiers, the light pulses can be sent over a distance as far as a few thousand kilometers with minimal distortion of the transmitted light pulses.

16.1 OVERVIEW OF FIBER-OPTIC COMMUNICATION SYSTEMS

Before going into detail, a brief overview of the entire fiber-optic communication system is presented here [1–11]. It is hoped that this section will be useful in orienting the reader and providing a simple perspective on the subject.

A pictorial block diagram of a fiber-optic communication system is shown in Fig. 16.1. The operation of the system is based simply on three major functions:

1. Conversion of electrical signals into modulated light by the transmitter.
2. Transmission of the modulated light signal through optical fiber networks.
3. Detection of the transmitted light by the receiver in order to recover the original electrical signal.

16.1.1 Transmitters

Semiconductor light sources such as the laser diode (LD) and light-emitting diode (LED) are most frequently used as light transmitters. Their popularity is largely due to such advantages as longevity, reliability, high efficiency, and compactness. The wavelengths of light used are predominantly 1.3 and 1.55 μm. These wavelengths were chosen from the attractive properties of optical fibers at these values. For example, light transmission is distorted least at a wavelength of 1.3 μm. On the other hand, light transmission loss is minimum at 1.55 μm.

16.1.2 Modulation of Light

In order to encode the signal information, the transmitted light must be altered or modulated using modulating circuitry. There are two major methods of

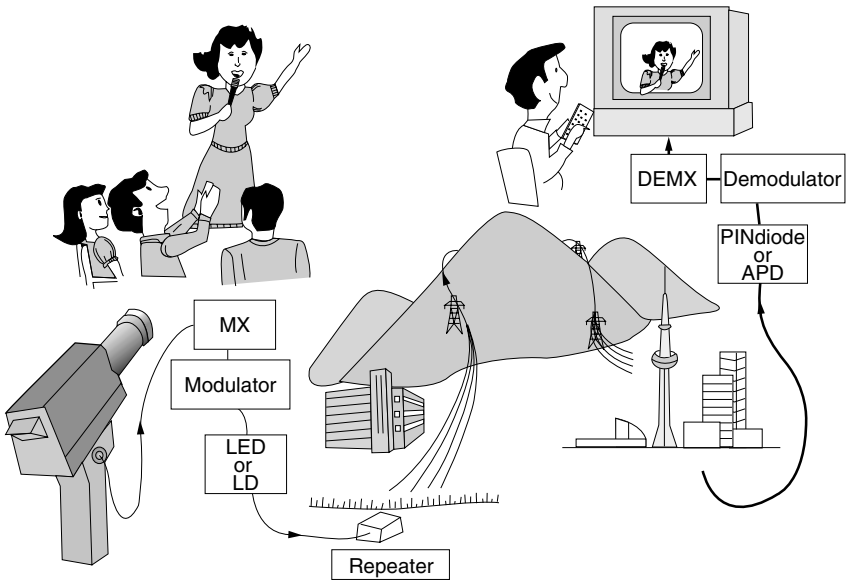


Figure 16.1 Pictorial block diagram of fiber-optic communication system. MX: Multiplexer. DEMX: Demultiplexer. LED: Light emitting diode. LD: Laser diode. APD: Avalanche photodiode.

modulation — analog and digital. Analog modulation is, in general, simpler than digital modulation, but the quality of the transmitted signal has some limitations. Digital modulation provides better quality of the transmitted signal but requires more complicated circuitry and a wider frequency bandwidth. The choice of modulation depends on the type of information being sent, the desired quality of signal, and how signals are combined during transmission.

For instance, transmitting speech through a voice channel does not require a high quality of transmission. Hence, analog modulation is usually used. However, if it is combined with other digitized signals, digital modulation is used to send the voice channel as well.

The intensity of the light from a laser diode can be modulated by varying its bias current, as shown in Fig. 16.2a. The operating characteristics of the laser diode in Fig. 16.2b demonstrate that the power of the light output follows the current of the electrical input.

In contrast, one may want to convert the amplitude of an electrical signal into its binary equivalent. For instance, a TV signal, which is an amplitude modulated signal,

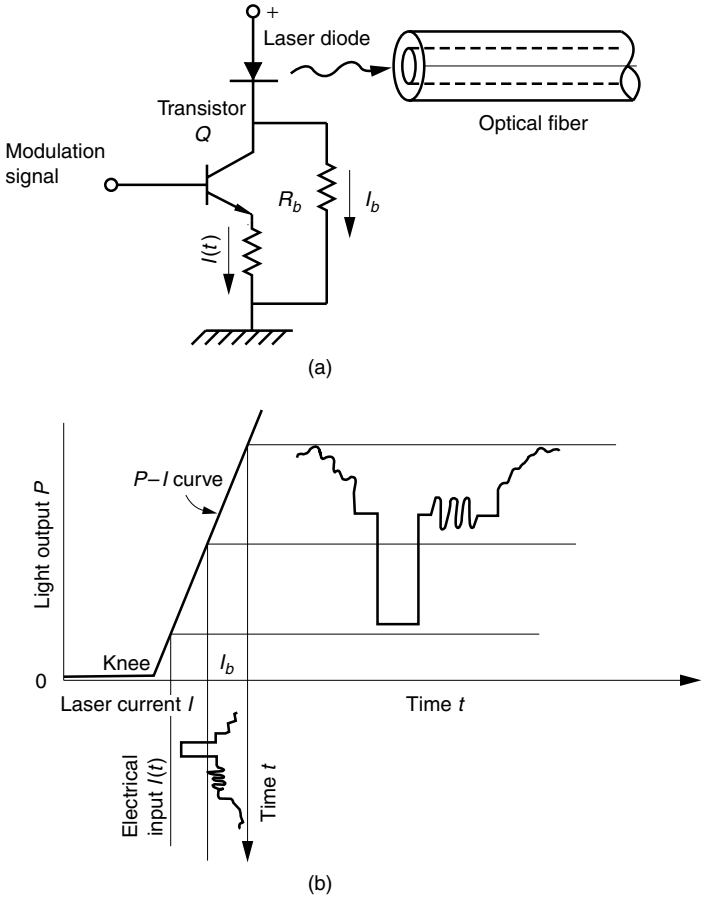


Figure 16.2 Modulation of laser light by an electrical signal. (a) Circuit for modulating a laser diode. (b) Light output versus injection current (*P–I* curve).

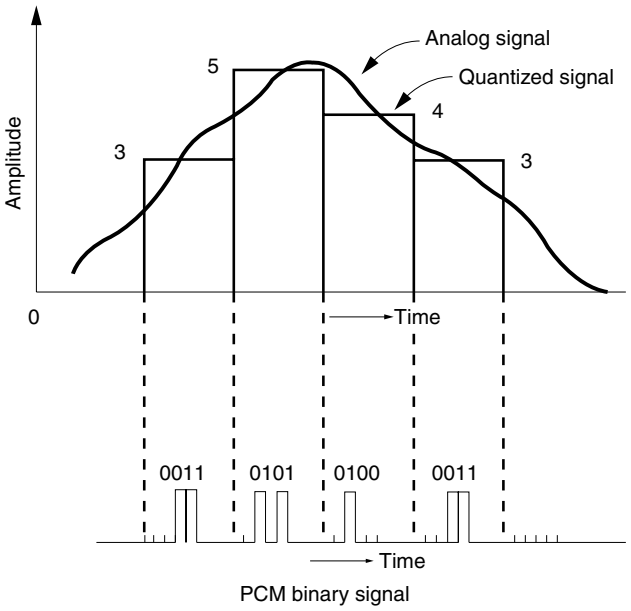


Figure 16.3 Principle of pulse code modulation (PCM).

must be coded in digital form. The most common method of digital modulation is the pulse code modulation (PCM) method, as illustrated in Fig. 16.3. In this figure, the value of the amplitude of the signal is sampled periodically. The sampled values are then converted into a binary form. For instance, if the first four sampled values were 3, 5, 4, and 3, they would be converted into an array of the following binary 0 and 1 combinations of 0011, 0101, 0100, 0011. Such an array of zeros and ones are then used to bias the laser diode to convert into light pulses. The light pulses propagate through the optical fiber network and reach the receiver, where the light pulses are converted back into an electrical signal of zeros and ones. These signals are then converted into analog signals by means of a digital-to-analog converter. Transmission with digital modulation requires much greater bandwidth than analog modulation. However, since the fiber-optic communication system inherently has an abundance of bandwidth, it can cope nicely with this situation.

16.1.3 Transmission Through the Optical Fiber

Next, the transmission of the light signal through the fiber is introduced. There are two types of fibers. One is the multimode fiber and the other is the single-mode fiber. The standard core diameter of the multimode fiber is 50 μm, while that of the single-mode fiber is 9.5 μm. For the multimode fiber, coupling light from the transmitter into the fiber's core is easier than coupling light into a smaller diameter single-mode fiber. However, the amount of the spread or distortion of a light pulse is greater in the multimode fiber than in the single-mode fiber.

The spread of the pulse shape is dealt with by increasing the spacing between the adjacent pulses so as to avoid overlapping and hence corrupting the individual signals. Since the spread is greater for multimode fibers, the spacing in time between adjacent pulses is greater than that for single-mode fibers. Thus, the maximum bit rate that can

be sent through multimode fibers is less than through single-mode fibers. Due to this difference, multimode fibers are used for low-bit-rate communication or in applications that require only short-distance communication. In contrast, single-mode fibers are used for high-bit-rate, long-haul communication systems.

An extremely short distance line-of-sight communication can use space-optic communication. Such systems are convenient for high-rise business buildings within a one kilometer radius. This scheme saves both cost and time of laying fiber optic cables underground of already congested business districts. The transmission condition of the laser beam, however, is subjected to atmospheric conditions, such as, fog, heavy rain, or clear air turbulence. Reliability of the communication system has to be sacrificed unless redundancy or some other countermeasure is incorporated [12].

16.1.4 Received Signal

The light reaching the receiver is first converted into an electrical signal by means of either a PIN photodiode or an avalanche photodiode (APD).

The detector is then followed by an electronic amplifier. Since the internal impedance of the PIN photodiode or APD is very high (of the order of megohms), special impedance transducers are necessary for optimizing the signal transfer from the detector to the preamplifier without sacrificing the frequency bandwidth.

In the event that the transmission fiber is long, optical amplifiers are installed. Erbium-doped fiber amplifiers are used to amplify light signals for 1.55-μm wavelengths. For 1.3-μm systems, semiconductor laser amplifiers (SLAs) are used.

16.1.5 Multiplexing Hierarchies

In reality, optical communication systems never send just one voice channel at a time; instead, they combine as many as over several thousand voice channels together. This combination is called *multiplexing* (MX). Currently, there are three main methods of multiplexing the electrical signal. They are frequency division multiplexing (FDM), time division multiplexing (TDM), and code division multiplexing (CDM). Separate from electrical signal multiplexing, there is optical signal multiplexing of different wavelengths of light, which is called wavelength division multiplexing (WDM). Most frequently, FDM, TDM, or CDM are combined with WDM to make the most of the wideband capability of fiber-optic communication systems.

The following sections will provide more detailed knowledge of the theory and practical implementations of the basic components of an optical communication system.

16.2 MODULATION

Paul Revere, who successfully defended the New England coast of America from a massive British invasion, used a flare signal from an observation post on Beacon Hill, signaling “one if by land, two if by sea.” One flare meant invasion through land; two flares meant invasion through Boston harbor. Modern communication theory would describe Revere’s warning light as being amplitude modulated by the baseband signal of one pulse meaning by land, and two pulses meaning by sea. Information can only be sent after being altered or modulated by a baseband signal to which a meaning is assigned. For instance, in communications, our voice is considered the baseband signal that modulates light around a center frequency called a carrier frequency. In

this section, methods of modulation [13–16] by the baseband signal and methods of demodulation to recover the baseband signal will be summarized.

Methods of modulation are broadly categorized into analog and digital modulations. Digital modulation is further subdivided into binary pulse modulation and multiple state pulse modulation. Each has advantages and disadvantages. Selection is made based on criteria such as frequency bandwidth, robustness against distortion and noise, and economy of light power. A summary of the various types of modulation and their features is presented in Table 16.1. Let's start with analog modulation that includes amplitude, frequency, and phase modulations.

16.2.1 Amplitude Modulation

Amplitude, frequency, and phase are the three parameters that can be used to modulate a lightwave. Amplitude modulation (AM) is the simplest of all. From the top row of Table 16.1, amplitude modulation with a baseband signal $s(t)$ is expressed as

$$E(t) = E_0[1 + ms(t)] \cos 2\pi f_c t \tag{16.1}$$

where f_c is the carrier frequency of the light and m is the modulation index. The frequency spectrum obtained by Fourier transforming this equation ranges from $f_c - B$ to $f_c + B$ and is centered at f_c, where B is the frequency bandwidth of the baseband signal. Thus, the frequency spectrum of the AM stretches over a bandwidth of $2B$. There are two ways of implementing AM in connection with a semiconductor laser diode source. One is internal and the other is external modulation.

With internal modulation, the baseband signal current combined with the proper bias current is directly injected into a laser diode to modulate its output light. In this approach, it is not the amplitude but the intensity of the light that is linearly modulated by the baseband signal current. Luckily, it is not the amplitude but the intensity of the received light that is proportional to the output electrical current from the photodiode detection. Such a method of detection is called *direct detection*.

The problem with this method of internal modulation is that the frequency of the output light is also modulated according to the injection current, resulting in unwanted frequency modulation (FM). This inherent FM of the light signal introduces the problem of additional spreading of the signal pulse within the fiber due to material and waveguide dispersion (see Section 11.1.4).

With external modulation, the laser diode is driven by a constant injection current, and its output light amplitude is externally modulated by devices such as the electrooptic amplitude modulator described in Section 5.2.1. This system is more involved but is free from the unwanted FM. Amplitude as well as intensity modulations can be achieved by the external modulation method.

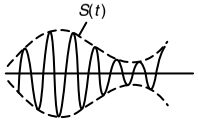
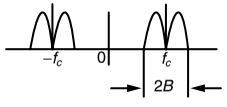
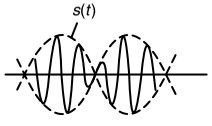
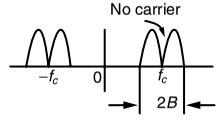
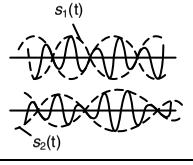
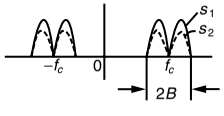
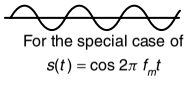
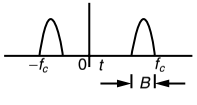
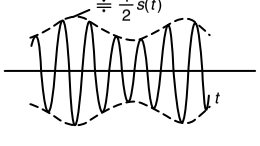
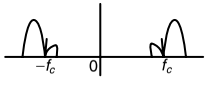
16.2.2 Variations of Amplitude Modulation

There are a few variations in the AM as also shown in Table 16.1. The first is double sideband (DSB) modulation in which the carrier is suppressed and is expressed as

$$E(t) = s(t) \cos \omega_c t \tag{16.2}$$

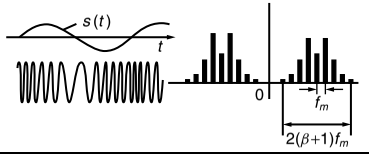
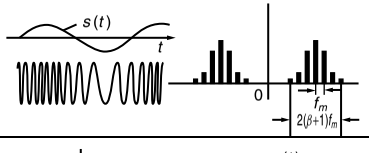
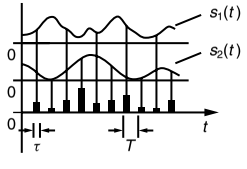
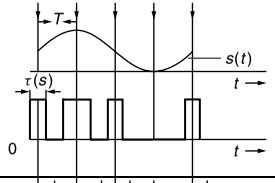
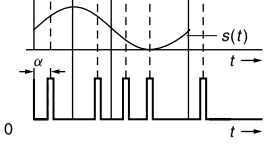
The carrier does not carry information, and its removal saves light energy. However, it requires a more elaborate demodulation scheme.

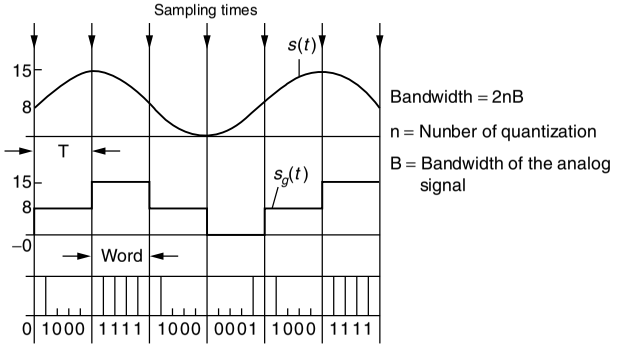
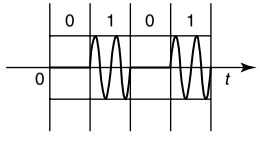
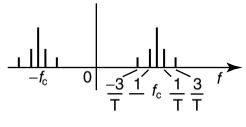
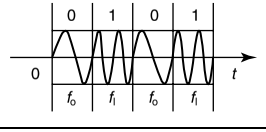
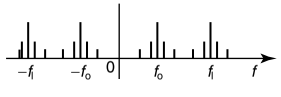
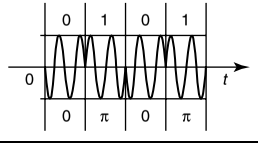
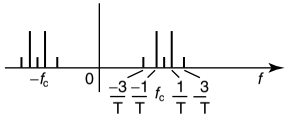
Table 16.1 Table of various types of modulation

Category	Acronym	Full Name	Expression	Spectrum: Time Domain	Spectrum: Frequency Domain	Features
Carrier Amplitude Modulation	AM	Amplitude modulation	$A[1 + ms(t)]\cos\omega_c t$ Spectrum of $s(t)$ is B Hertz			Simplest to modulate and demodulate. Bandwidth is not economized.
	IM	Intensity modulation	$I[1 + ms(t)]\cos\omega_c t$			
	DSB	Double sideband modulation	$s(t)\cos\omega_c t$			Easy to modulate but difficult to demodulate. No energy for carrier frequency. (Required power is 50% less than AM).
	DSBC	Double sideband suppressed carrier modulation				
	QAM	Quadrature amplitude modulation	$s_I(t)\cos\omega_c t + s_2(t)\sin\omega_c t$			Bandwidth is conserved.
	SSB	Single sideband modulation	$\frac{1}{2}[s(t)\cos\omega_c t + \hat{s}(t)\sin\omega_c t]$ $\hat{s}(t)$: Hilbert transform of $s(t)$			Bandwidth is conserved. Difficult to demodulate.
	VSB	Vestigial sideband modulation	$\frac{1}{2}[s(t)\cos\omega_c t - js(t)\sin\omega_c t]$			Bandwidth is conserved. Difficult to demodulate. One and a fraction of sideband is used. No such abrupt filter as used for SSB is needed.

(*continued next page*)

Table 16.1 (*Continued*)

Category	Acronym	Full Name	Expression	Spectrum: Time Domain	Spectrum: Frequency Domain	Features
Angle Modulation	PM	Phase modulation	$A\cos[\omega_c t + s(t)]$			Implementation is complicated. Stringent requirement of LO laser. Robust against distortion and noise.
	FM	Frequency modulation	$A\cos[\omega_c t + \int_{t_o}^{t} s(\lambda)d\lambda]$			Implementation is complicated. Robust against distortion and noise.
Pulse Modulation	PAM	Pulse amplitude modulation	$\sum_{n=-\infty}^{\infty}[as(nT)+K]\Pi\left(\frac{t-nT}{\tau}\right)$			Both modulation and demodulation are simple. The narrow pulse width is susceptible to noise. The example shown is used for the time division multiplexing.
	PDM	Pulse duration modulation	$\sum_{n=-\infty}^{\infty}\Pi\left(\frac{t-nT}{\tau(s)}\right)$ $\tau(s)=as(nT)+K$			The amplitude is not susceptible to noise. The positions of the zero crossing point of both sides of the pulse are susceptible to noise.
	PPM	Pulse position modulation	$\sum_{n=-\infty}^{\infty}\Pi\left(\frac{t-nT-\alpha}{\tau(s)}\right)$ $\alpha=as(nT)$			The pulses are all the same height and width. Only one side of the zero crossing point is susceptible to noise.

Analog Signal in Digital Form	PCM	Pulse code modulation				Commonly used. Robust against distortion and noise. Moderate cost for high-quality transmission.
Binary Modulation (Two Level Modulation)	ASK	Amplitude shift keying	$h(t)\cos\omega_c t$ $h(t) = \begin{cases} 0 \\ 1 \end{cases}$			Simple to modulate and demodulate. Robust against distortion and noise
	FSK	Frequency shift keying	$\cos[\omega_h(t)t]$ $\omega_h(t) = \begin{cases} 2\pi f_0 \\ 2\pi f_1 \end{cases}$			Three-decibel higher energy than ASK. Complicated in implementation. Robust against distortion and noise.
	PSK	Phase shift keying	$\cos[\omega t + \phi(t)]$ $\phi(t) = \begin{cases} 0 \\ \pi \end{cases}$			Three-decibel higher energy than ASK. Stringent requirement on the reference wave to detect.

Next, there is quadratic amplitude modulation (QAM), which can transmit two signals without increasing the frequency bandwidth. The light is split into two parts. The first part (in-phase component) is multiplied by the baseband signal $s_1(t)$ to give

$$E_I(t) = s_1(t)\cos\omega_c t \tag{16.3}$$

The phase of the other part (quadrature component) is shifted from that of the first part by $\pi/2$ radians. The quadrature component is multiplied by the other baseband signal $s_2(t)$ to give

$$E_Q(t) = s_2(t)\sin\omega_c t \tag{16.4}$$

E_I and E_Q are transmitted simultaneously. Without increasing the frequency bandwidth, both $s_1(t)$ and $s_2(t)$ can be transmitted. In the receiver, the signals are selectively demodulated by multiplying $E_I(t)$ by $\cos\omega_c t$, and by multiplying $E_Q(t)$ by $\sin\omega_c t$, and then filtering with a low-pass filter.

In DSB modulation, the two sideband signals contain essentially the same information. Hence, this redundant signal wastes power and bandwidth during its transmission. To economize in both of these areas, single sideband (SSB) modulation was introduced, where only a single sideband is transmitted. A single-sideband signal is mathematically expressed as [14]

$$E(t) = \tfrac{1}{2}[s(t)\cos\omega_c t + \hat{s}(t)\sin\omega_c t] \tag{16.5}$$

where $\hat{s}(t)$ is called the Hilbert transform of $s(t)$, which is realized by passing the baseband signal through a 90° phase shifter. The above expression can be verified by using the example of a sinusoidal signal, namely,

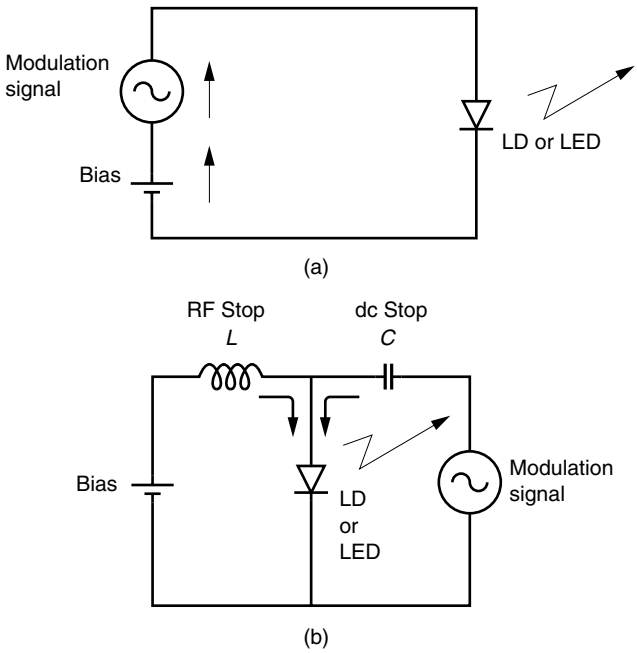


Figure 16.4 Amplitude modulation of a radiant diode. (a) Series circuit modulation. (b) Shunt circuit modulation.

$$s(t) = a\cos\omega_m t$$
$$\hat{s}(t) = a\sin\omega_m t \tag{16.6}$$

This results in a modulated signal of

$$E(t) = \tfrac{1}{2}a\cos(\omega_c - \omega_m)t \tag{16.7}$$

which contains only the lower sideband $(\omega_c - \omega_m)$ but not the higher sideband $(\omega_c + \omega_m)$.

The drawback of SSB modulation is that it requires additional circuitry to remove one of the sidebands. It also requires a more elaborate demodulation.

Finally, in vestigial sideband (VSB) modulation, only a portion of the amplitude modulated signal is removed. No abrupt cutoff filters are needed, and the remaining vestigial fraction of the carrier frequency is used for demodulating the signal.

In the design of the internal AM circuit, one should consider (1) the ease of modulation, (2) the linearization of the characteristic curve, and (3) matching to a low input impedance of typically 1.5–2.0 Ω. The methods of internal AM are classified into series and shunt connections, as shown in Fig. 16.4. Samples of both types of connections for modulating an LED or LD are shown in Fig. 16.5.

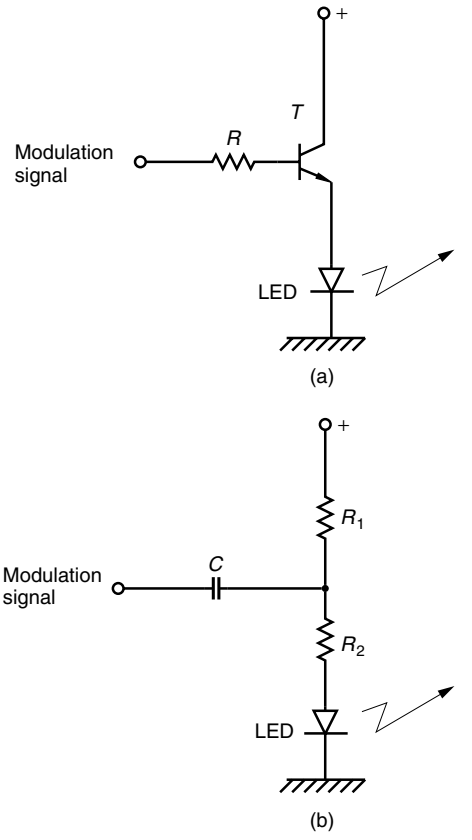


Figure 16.5 Simple circuits to modulate an LED. (a) Emitter follower (series connection). (b) *RC* coupling (shunt connection).

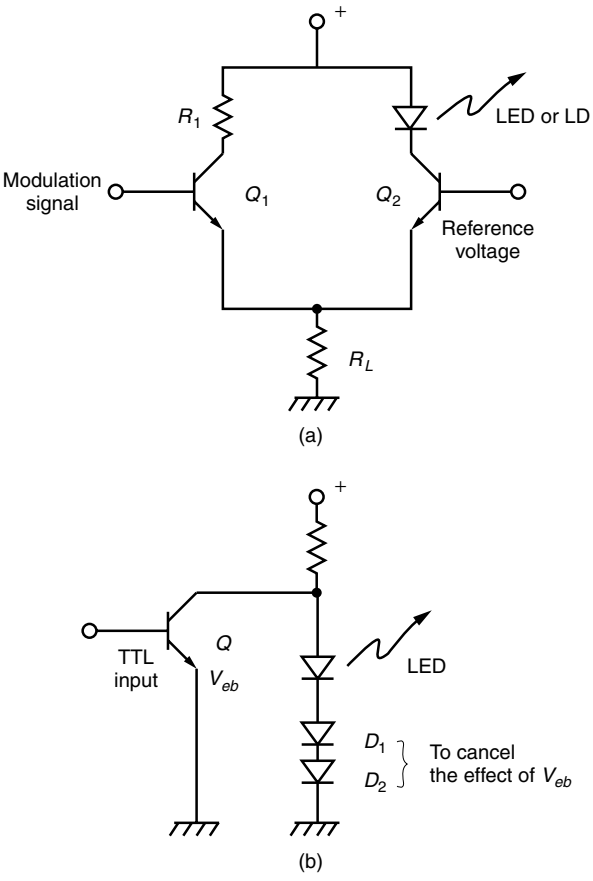


Figure 16.6 Circuits for modulating an LED. (a) Emitter coupled circuit. (b) Parallel switching.

Figure 16.6 shows some additional circuits used in modulating an LD or LED. The circuit shown in Fig. 16.6a is an emitter coupled circuit. By pairing the same type of transistors, the drift of the operating point due to the ambient temperature is suppressed. Another modulation method is the parallel switching shown in Fig. 16.6b. As the transistor Q is turned on and off, the shunt circuit to the LED is also switched on and off. The diodes D_1 and D_2 are to assure the off state of the LED.

Figure 16.7 shows driver circuits with pre-enhancement. In order to operate in the linear region of the characteristic curve of the light power versus bias current of the LD, the driver shown in Fig. 16.7a is prebiased by shunting the transistor Q with R_b. The value of R_b controls the bias point, which is necessary to operate in the linear region. Figure 16.7b shows a driver that compensates for the drop of output light power from an LED at higher modulating frequencies. It uses a capacitor across the resistor R. With this driver, the frequency band of modulation can be stretched.

16.2.3 Angle Modulation

Another means of modulation besides amplitude modulation is angle modulation. Phase modulation (PM) and frequency modulation (FM) are two types of angle modulation.

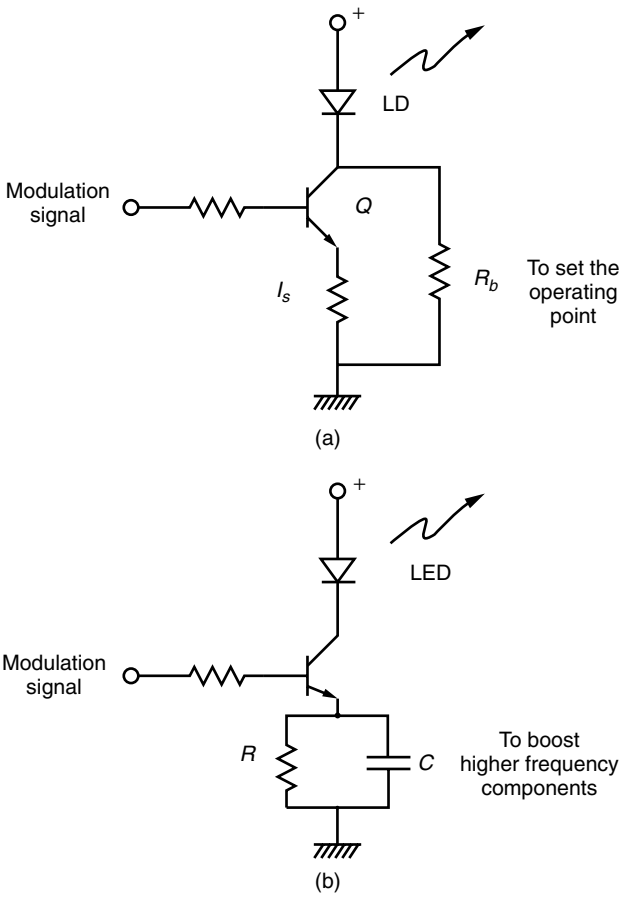


Figure 16.7 Drivers with pre-enhancement. (a) Driver with pre-biasing. (b) Driver with pre-emphasis.

The baseband signal modulates the phase of the signal to give a phase modulated expression of

$$E(t) = A\cos[\omega_c t + s(t)] \tag{16.8}$$

whereas frequency modulation can be expressed as

$$E(t) = A\cos\left(\omega_c t + \int_{t_0}^{t} s(\lambda)d\lambda\right) \tag{16.9}$$

Note that the required frequency bandwidth of the AM signal is $2B$, but in the case of FM, the required bandwidth is much wider, as shown in Table 16.1.

The benefit of PM and FM over AM is significantly improved discrimination against noise and interference. For both PM and FM, the method of demodulation utilizes phase and frequency as a means of retrieving the baseband signal. Thus, they do not depend on the amplitude of the light signal where interference and noise primarily show up.

The phase or frequency modulated signal, however, cannot be demodulated by the direct detection method, which can only detect the light power and cannot discriminate either the frequency or phase of the light carrier. A different detection system called a *coherent detection system* has to be used to demodulate phase or frequency modulated signals. With coherent detection, the received signal is first fed to a mixer diode where the received light is superimposed onto a local oscillator light to generate an intermediate frequency (IF) signal,

$$f_{\mathrm{IF}} = |f_c - f_{\mathrm{LO}}|$$

where f_{LO} is the carrier frequency of the local oscillator light. The FM signal is demodulated from the IF signal using a frequency discriminator, and the PM signal is demodulated from the IF signal by a phase discriminator.

16.2.4 Pulse Modulation

Pulse modulation conveys information in pulse form rather than in a continuous wave form. Various kinds of pulse modulation are outlined in this section, the first of which is called pulse amplitude modulation (PAM). In PAM, the height of the pulse is modulated according to the amplitude of the baseband signal.

The most popular use of PAM is in time division multiplexing (TDM). With TDM, several different baseband signals can be transmitted in a single channel. The graph in Table 16.1 explains how TDM is used for multiplexing two signals $s_1(t)$ and $s_2(t)$ in one channel. Signal 1 is sampled for a short time of τ seconds, and then signal 2 is sampled for another τ seconds. Samples are alternated between signals 1 and 2. The composite signal is a pulse amplitude modulated signal that is sent through a single channel. At the receiver, the two signals are separated by a properly timed switch.

At this point, it is worth briefly mentioning three other multiplexing techniques besides TDM. Frequency division multiplexing (FDM) uses different carrier frequencies for each signal and transmits them simultaneously into a single channel. Code division multiplexing (CDM) uses different codes for each signal, and at the receiver, the signals are separated using autocorrelation according to the difference in codes.

Returning now to the subject of pulse modulation, pulse duration modulation (PDM) and pulse position modulation (PPM) are two other types of pulse modulation and are shown below PAM in Table 16.1.

PDM modulates the pulse duration τ of each pulse in the pulse period T according to the amplitude of the baseband signal. The advantage of this modulation is the constant pulse amplitude. Hence, as with FM, amplitude variation due to either noise or interference can easily be removed.

PPM modulates the time of occurrence of the pulse α after the beginning of the pulse period T. PPM possesses similar advantages to those of PDM.

16.2.5 Pulse Code Modulation

In pulse code modulation (PCM), as previously introduced in Fig. 16.3, the baseband signal is first quantized into discrete amplitude levels that can be represented by binary coded pulses. The chain of binary coded pulses thus generated is used to modulate the transmitter light. The simplest binary code light modulation is to represent "1" by

the presence of light and "0" by the absence. At the receiver, the light is detected to recover the binary coded pulses. The binary signal is then converted into an analog signal by a digital-to-analog converter.

The binary code modulation is special in the sense that it requires two states representing "0" and "1." Since this type of modulation is commonly used, the next section expands on this type of modulation.

16.2.6 Binary Modulation (Two-State Modulation)

Binary modulation is often referred to as "shift keying." In earlier days, communication officers were trained to send messages using Morse code. In Morse code, each letter of the alphabet is represented by a unique set of dots and dashes. The dots and dashes are generated by turning an electrical signal on and off. The switching device for sending Morse code was called a keying device, and the terminology has been carried over to modern communication. The keying device can generate only two states. The electrical signal is either "on" or "off" and there is no in between state.

In communication theory, binary modulation is called shift keying. If the two states are generated by the difference of two levels of amplitude, the modulation is called amplitude shift keying (ASK). If the two states are obtained using two different frequencies of the light carrier, the modulation is called frequency shift keying (FSK), and if the two states are created using two different phases of the carrier, the modulation is called phase shift keying (PSK).

16.2.7 Amplitude Shift Keying (ASK)

The two levels of ASK modulation in optical communication are provided by turning the injection current of the semiconductor laser diode on and off. Since the level of injection current of the "on" state is the same at every pulse, there is no concern about unwanted FM as in the case of direct AM of the semiconductor laser diode.

At the receiver, the ASK modulated light is detected and reduced to the original chain of electrical pulses. The "0" or "1" state is determined by comparing the level of the received signal to that of a comparator. If the signal level is higher than the comparator level, then the state is a "1" state; and if lower, the state is a "0" state.

Thus, with ASK modulation, even when distortion due to noise is almost half as large as the regular pulse height, the pulse can still be correctly decoded as long as the extent of the distortion is less than half the height of the undistorted pulses. Not only is ASK modulation robust against signal deterioration due to distortion and noise, but it is also simple to implement and analyze theoretically.

16.2.8 Frequency Shift Keying (FSK) and Phase Shift Keying (PSK)

FSK shifts the carrier frequency of the light between two different frequencies corresponding to the "1" and "0" states of the binary code. The FSK modulated light can be generated either by using a frequency tunable semiconductor laser diode or by switching between two semiconductor laser diodes of two different wavelengths.

The implementation of FSK modulation is more complicated than that of ASK, and the frequency spectrum bandwidth is $2(\Delta f + B)$, where $2\Delta f$ is the spacing between the two frequencies, and $2B$ is due to the bandwidth of the baseband signal. Compared

to ASK, the required bandwidth of FSK is $2\Delta f$ wider. PSK shifts the phase of the light carrier between two different phases using an electrooptical phase modulator.

One of the great advantages of FSK and PSK, however, is greater robustness against distortion and noise than ASK. The reason for this is that ASK uses amplitude information, which is most easily influenced by noise; whereas FSK uses information about the frequency to determine the state.

The ASK modulated light is in the off state about 50% of the time. The FSK modulated light is always on; thus, the amount of light energy that reaches the receiver is 3 dB higher than that of ASK, which leads to a higher signal to noise ratio for an FSK system.

Compared to FSK, PSK provides an improved signal to noise ratio, but its method of detection is more complicated. The detection method is homodyne detection. It involves creating a local light with a frequency and phase identical to or in the vicinity of those of the received light.

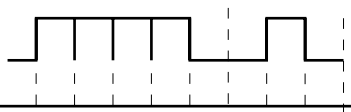
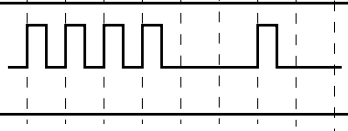
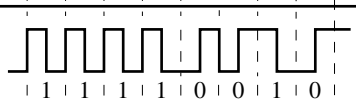
16.2.9 Representation of Bits

The manner of representing the logic states "1" and "0" by the states of an electrical signal will next be illustrated. The state of the electrical signal is represented by either the amplitude of the pulse in ASK, the frequency of the pulse in FSK, or the phase of the pulse carrier in PSK. The coding characteristics are listed in Table 16.2.

NRZ (Nonreturn to Zero) Code The logic "1" is represented by maintaining the electrical signal in the "1" state throughout the bit period. Similarly, the logic "0" is represented by the state "0" of the electrical signal maintained throughout the bit period. Even though coding is simple, there are some drawbacks. When a capacitor coupled circuit is used in the input of a signal amplifier, an uninterrupted chain of continuous "1" bits is seen as a dc signal and the signal is blocked by the capacitor and the dc level drifts. This phenomenon is called *baseline wander*. In an effort to remove this drawback, the next code was invented.

RZ (Return to Zero) Code The logic "1" state is represented by the state "1" of the electrical signal maintained only during the first half of the bit period, and during

Table 16.2 Representation of bits

Acronyn	Full Name	Description	Example
NRZ	Nonreturn to zero code	Logic "1": Electrical state 1 during entire period Logic "0": Electrical state 0 during entire period	
RZ	Return to zero code	Logic "1": Electrical state 1 during half period Logic "0": Electrical state 0 during entire period	
Manchester	Manchester code	Logic "1": Transition from electrical state 1 to 0 Logic "0": Transition from electrical state 0 to 1	

the second half of the period, the signal is turned to the "0" state. Thus, the pulse of the "1" state lasts only one-half of the bit period. The logic "0", on the other hand, is represented by the state "0" of the electrical signal during the entire bit period.

With this coding, even when "1" states appear in a continuous sequence, the signal level changes every midbit and can be amplified by a capacitor coupled amplifier. The required frequency bandwidth of the amplifier, however, doubles that of NRZ.

In a long-distance transmission, however, the noise of the optical amplifiers and the dispersion of the optical fiber affect the performance. A better overall quality is reported with NRZ code than with RZ code [16].

Manchester Code With Manchester code, logic is represented by the transitions of the states of the electrical signal rather than by the state itself.

The logic "1" is represented by the transition of the state of the electric signal from "1" to "0", while the logic "0" is represented by the transition in the opposite direction which is from "0" to "1." The simple rule is that logic "1" starts from the state of "1" and the logic "0" starts from the state of "0." Since Manchester code always compares the levels of two adjacent pulses, it is not affected by the long-term variation of the signal level. Moreover, transitions exist at every midbit, and a clock pulse can easily be generated inside the receiver.

There is a price to be paid for the robustness of binary coding for all three methods and that price is increased frequency bandwidth. Let us calculate quickly the frequency bandwidth needed to send one 4-kHz voice channel using PCM coding. In order to send a 4-kHz voice channel, the amplitude of the voice signal must be sampled at least twice each period of the highest frequency component of the voice channel (Nyquist criterion). This means that the instantaneous value of the voice signal has to be sampled at 8×10^3 times a second. Each sampled value must immediately be converted into 8-bit words or 256 gray scales to represent the sampled value. Thus, the bit rate, equal to the number of bits generated per second to send one voice channel, is $8 \times 8 \times 10^3 =$ 64 kb/s.

Each pulse is a rectangular shape. A rule of thumb for estimating the required frequency bandwidth for an almost perfect recovery of the pulse shape is 10 times the bit rate; but in digital communication, perfect shape recovery is not essential since it is only necessary to distinguish between the presence and absence of a pulse. In order to discern between the presence and absence of a pulse, a frequency bandwidth B equal to the bit rate B_t for RZ code and one-half of the bit rate B_t for NRZ code of the pulses will suffice; namely,

$$B = B_t \quad \text{for RZ}$$

$$B = \tfrac{1}{2}B_t \quad \text{for NRZ}$$

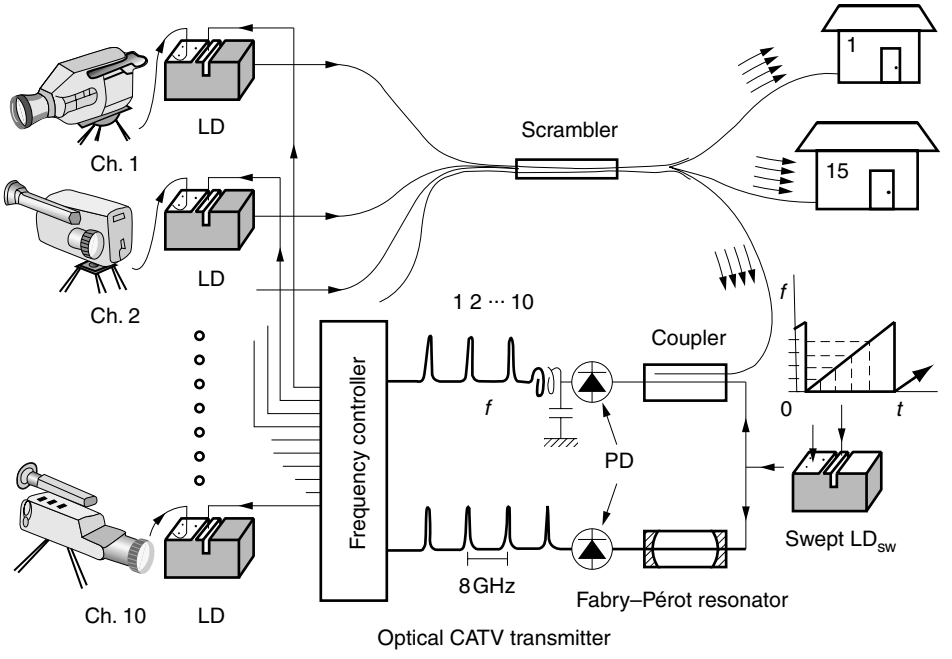
This is explained further in Section 16.6.5. Thus, the PCM of one voice channel of 4 kHz needs a frequency bandwidth of at least 32 kHz.

16.3 MULTIPLEXING

A fiber-optic communication system has such an abundant frequency bandwidth that multiplexing schemes for sending all different channels in the same optical fiber have been implemented.

16.3.1 Wavelength Division Multiplexing (WDM)

The first layer of the hierarchy of multiplexing methods is wavelength division multiplexing (WDM). Figure 16.8a shows an example of a small scale WDM used for sending several TV channels in one fiber [17]. The electrical signal of each TV camera intensity modulates a laser diode emitting at its own specific wavelength. The spacing between adjacent carrier frequencies is maintained at 8 GHz (see Section 3.2.4). The light outputs from the array of laser diodes are then fed into a scrambler, which



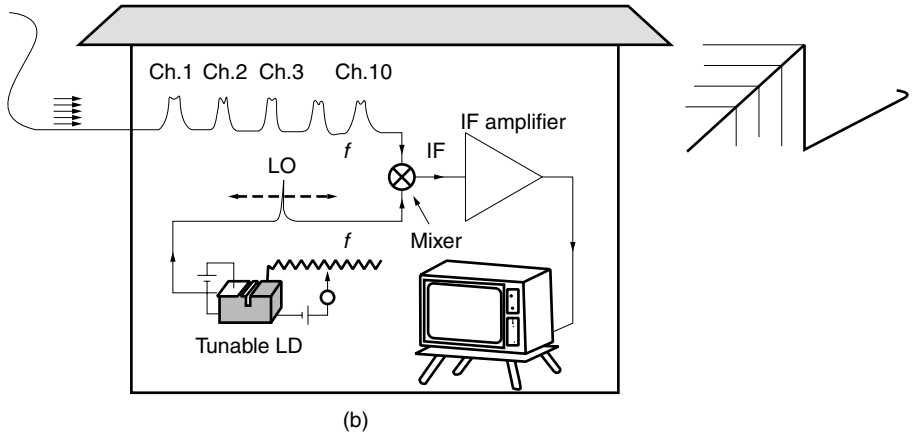


Figure 16.8 An example of wavelength division multiplexing (WDM) used for sending TV channels. (a) Channel space locking by the reference pulse method. (After S. Yamazaki et al. [17].) (b) Optical heterodyne receiver at home.

combines all the wavelengths of the light, producing a WDM system. The combined signal is then shared and sent to its subscribers via fiber.

From there, the combined signal is received by the subscriber's TV. The TV shown in Fig. 16.8b is equipped with a coherent detector. As you may recall (in Section 12.6), a coherent detector uses a local oscillator that can be tuned in order to select a particular frequency of the incoming scrambled light, thereby allowing the subscriber to view a desired channel.

16.3.2 Frequency Division Multiplexing (FDM)

Figure 16.9a illustrates how FDM is combined with the above-mentioned WDM and applied to telecommunications. Each voice channel (baseband) is used to amplitude modulate the output of a fixed frequency oscillator. The oscillator frequency is called the subcarrier frequency. The ith voice channel modulates the ith subcarrier frequency f_i. The outputs of the subcarrier frequencies

$$f_1, f_2, f_3, \ldots, f_i$$

are combined by a combiner T_1 that superimposes the various signals. Typically, a wide frequency band transformer is used to combine the signal for a lower frequency range of carrier frequencies, as shown in Fig. 16.9a. Alternatively, an RC circuit combiner can be used for a higher subcarrier frequency range. This is the first stage of frequency division multiplexing.

The frequency spacing between the subcarriers f_i is 8 kHz for the FDM of a voice channel because the frequency spectrum of $\cos 2\pi f_c t$, where f_c is 4 kHz, has two spectra at ± 4 kHz, occupying a frequency spectrum of 8 kHz centered at the subcarrier frequency. In general, the subcarrier frequency f_i has a spacing of $2B$ for multiplexing B frequency band signals.

Similar arrangements of FDM are made with transformers: $T_2, T_3, T_4, \ldots, T_k$. Let us take T_k as an example. The output from transformer T_k modulates the laser diode LD_k whose carrier wavelength is λ_k. The light outputs of k different wavelengths are focused into a single optical fiber to perform WDM. Thus, only one optical fiber transports $i \times k$ voice channels all at once.

If the multiplexed signals are again FD multiplexed by n times before modulating the laser diode, $i \times n \times k$ voice channels can be multiplexed.

Next, a method of demultiplexing the signal after reaching the receiver will be explained by referring to Fig. 16.9b. The received light first illuminates an optical demultiplexer such as a dispersive prism or a grating. The received light components are diffracted toward their respective locations according to wavelength into a waiting photodiode detector.

The output of each detector contains the frequency division multiplexed electrical signals that are further demultiplexed according to the subcarrier frequencies f_i by a bank of band-pass filters with center frequencies $f_1, f_2, f_3, \ldots, f_i$ with bandwidth $2B$. The demultiplexed signals are detected by each of their own detectors and an array of the original baseband signals is recovered. Figure 16.10 shows a similar but bidirectional case. An arrayed-waveguide grating (AWG) is used as a demultiplexer (DEMX). The AWG is a grating of arrayed waveguides deposited on a slab guide (see Section 10.6.1). It disperses the angle of diffraction in accordance with of the wavelengths [18].

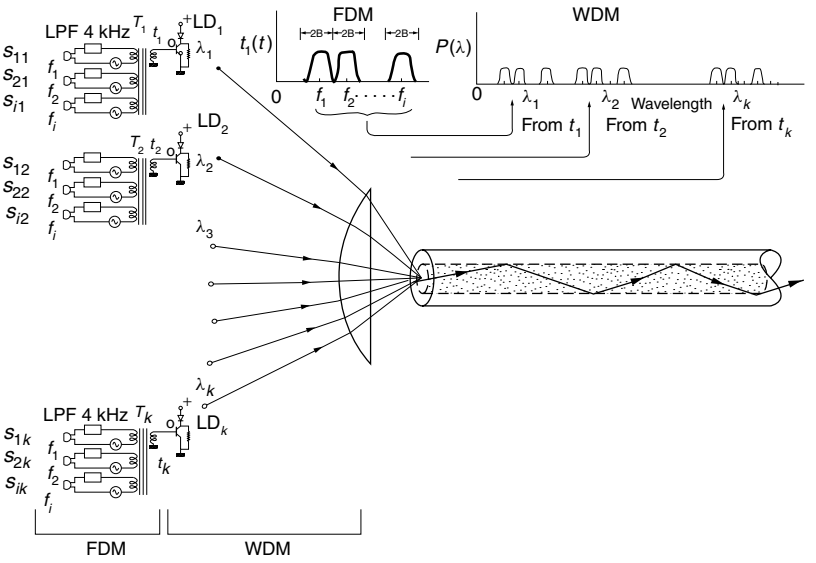


Figure 16.9 (a) WDM combined with FDM. (b) Demultiplexing of WDM combined with FDM.

16.3.3 Time Division Multiplexing (TDM)

Next, time division multiplexing (TDM), which is used for digital modulation, is explained using an example of multiplexing two digital voice channels.

In Fig. 16.11, output pulses from two channels are switched back and forth and two chains of light pulses are interleaved into one. Such a multiplexed signal is sent through

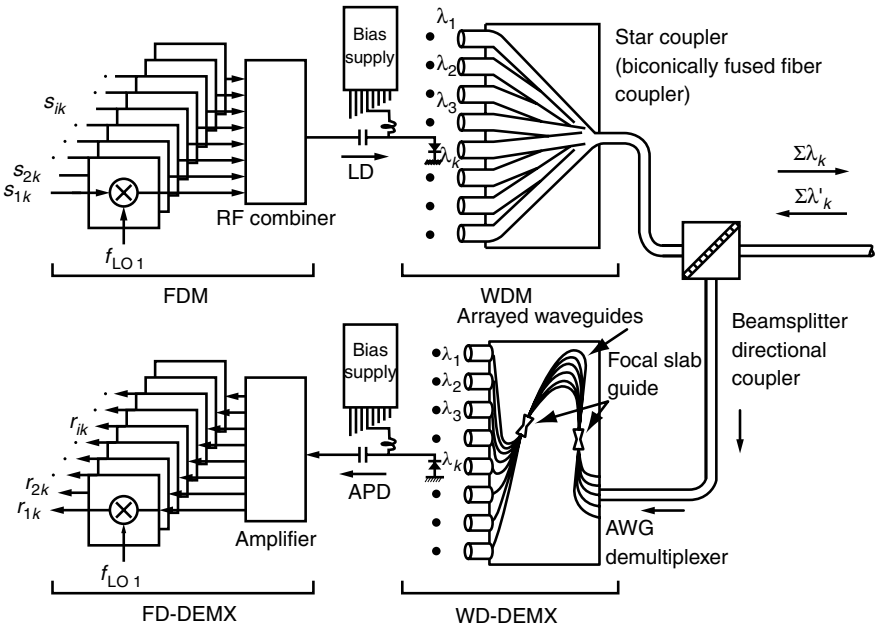


Figure 16.10 Bidirectional WDM combined with FDM.

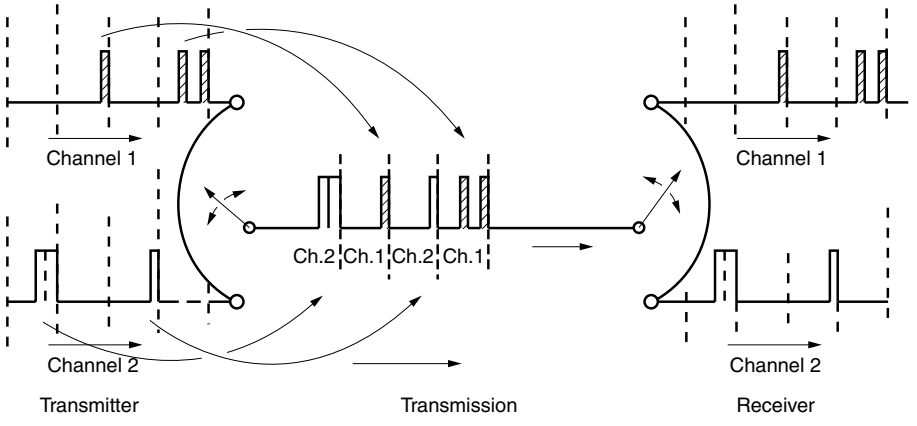


Figure 16.11 Time division multiplexing (TDM) and demultiplexing.

the optical communication system. At the receiver, the signal is demultiplexed into the two original separate channels by a similar switching system. Provisions for proper synchronization of the switches at the transmitter and the receiver are necessary and are derived from the periods of the received signal averaged over time. The illustrated example demonstrates the case of multiplexing two channels but it is not unusual to multiplex over a thousand channels per fiber.

A system that can send 24 voice channels is considered the most primitive (basic) communication system and is called digital system 1 (DS-1) in the American telephone

Table 16.3 Designation of PCM telephone bit rate

System Name	Number of Voice Channels	Data Rate (Mb/s)
	In the United States	
DS-1	24	1.544
DS-1C	48	3.152
DS-2	96	6.312
DS-3	672	44.736
DS-3C	1344	91.053
DS-4	4032	274.175
	In Japan	
F-6M	96	6.312
F-32M	480	32.064
F-100M	1440	97.728
F-200M	2880	198.584
F-400M	5760	397.200

system. Such a system allocates 1.544 Mb/s for the pulse rate as shown in Table 16.3. Let us confirm this allocated bit rate. The required pulse rate is 4 kHz (voice frequency bandwidth) $\times$ 2 (number of samples in one cycle) $\times$ 8 (each sampled value is gray scaled by 8 bits) $\times$ 24 (number of TDM circuits) = 1.536 Mb/s, which is (1.544 − 1.536) = 8 kHz less than the allocated bit rate. This gap is used for synchronizing the switches and signaling the telephone calls. Table 16.3 summarizes telephone bit rates for the standardized U.S. system, as well as the standardized Japanese system.

In the next section, detection of the modulated light in the receiver will be discussed.

16.4 LIGHT DETECTION SYSTEMS

The principles involved in light detection have already been mentioned in Chapter 12. In this chapter, these principles are explored further to describe how the performance of the detector influences the overall performance of the fiber-optic communication system. Such factors as the system's transmission distance, frequency bandwidth, signal to noise ratio for the analog signal, and bit error rate in the digital system all critically depend on the quality and capabilities of the receiver.

16.4.1 Equivalent Circuit of the PIN Photodiode

The PIN photodiode can be considered as a combination of a p-n junction diode and a constant current generator driven by the incident light. The current i_d from the p-n junction diode for the case without incident light (see Appendix A of Vol. II) is

$$i_d = I_{so}\left(e^{eV/kT} - 1\right) \tag{16.10}$$

As shown in Fig. 16.12, the current i_d increases exponentially for positive V. For large negative V, the current approaches the value I_{so}. As seen from Eq. (16.10), I_{so} is the current when the p-n junction is deeply back-biased. It is called the *saturated back-biased current*.

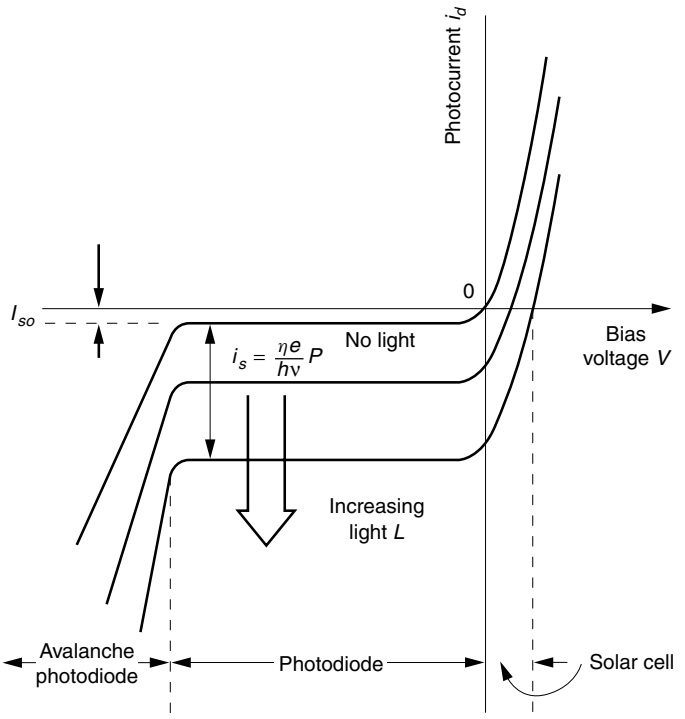


Figure 16.12 Current–voltage characteristic of a p-n junction.

When light whose quantaum $h\nu$ is larger than the bandgap is incident on the junction, there is a chance that an electron in the valence band will absorb the photon energy and move into the conduction band, leaving a hole behind. This process is called *photon-induced pair production*. The current due to the incident photons or photocurrent has to be added to Eq. (16.10). The characteristic curve including the photocurrent is shown in Fig. 16.12. Increasing the number of incident photons causes a downward translation of the I–V curve. In the negative bias region, also referred to as the back-biased region, the n layer is positive and the p layer is negative. The negative bias region is used as a photodetector because the dark current (current in absence of incident light) is at its minimum, having reached the saturated back-biased current. The output current is then primarily due to the photocurrent. The shot noise, which is proportional to the current through the junction, can be minimized. Moreover, in this region, the incremental impedance of the p-n junction is practically independent of the back-bias voltage.

The region of extreme negative bias is that of the avalanche effect. With a large bias voltage, the accelerated electrons start creating additional electron–hole pairs by colliding with the orbital electrons. The photocurrent can be multiplied up to a thousand times depending on the bias current. This is the avalanche photodiode (APD) (see Section 12.4).

Figure 16.13a shows a direct detection circuit using a PIN photodiode such as the one shown in Fig. 12.4. Figure 16.13b shows the simplified equivalent circuit including noise sources and the current generator driven by the incident light. The source and the current therefore consist of the diode current expressed in Eq. (16.10) and the photocurrent expressed in Eq. (12.8). Quite often, the dark current, which is the current

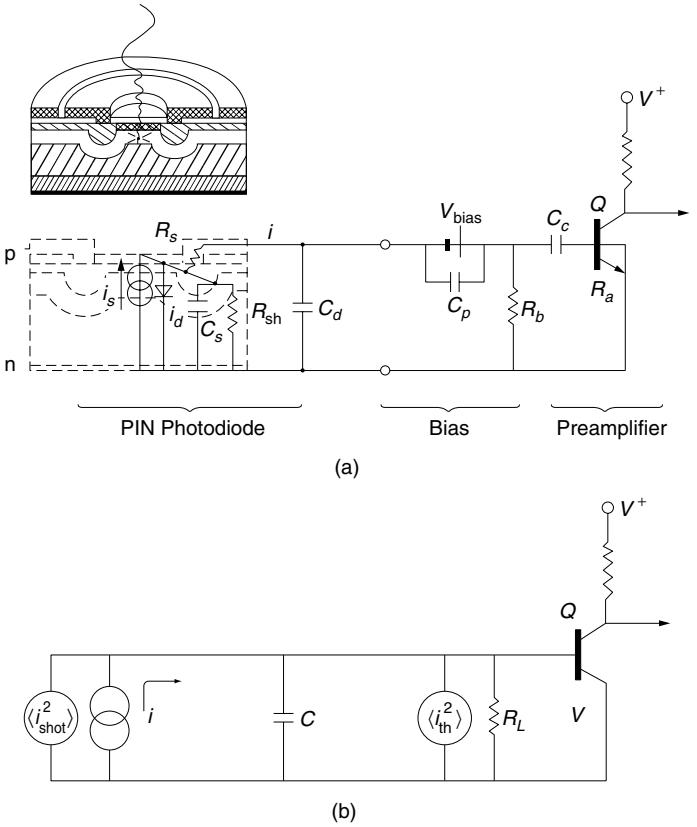


Figure 16.13 Equivalent circuits of a PIN diode. (a) Connection diagram of PIN photodiode. (b) Simplified equivalent circuit including noise sources.

without incident light, is more than the saturated back-biased current. Such excessive dark current is caused by impurities and lattice defects inside, as well as on the surface, of the crystal. This extra current is accounted for by the shunt resistance R_{sh}. The capacitive effect of the immobile charge in the depletion region is represented by C_s; the distributed capacitance associated with the connecting wire, by C_d; the contact resistance of the electrode, by R_s; the input resistance of the amplifier, by R_a; the bias resistor, by R_b; the RF bypass capacitor, by C_p; and the coupling capacitor, by C_c.

16.4.2 Frequency Response of the PIN Diode

The upper limit on the amplitude modulation frequency of the signal light that can be detected by a PIN diode is determined by the following factors:

1. The transit time for an electron generated in the intrinsic-type layer to reach the electrode.
2. The time for the carriers generated in either p-type or n-type layers to reach the electrodes.
3. The RC time constant of the signal circuit.

Ways to raise the upper limit on the frequency response are outlined as follows:

1. The transit time in the intrinsic layer can be shortened by raising the back-bias voltage. With a minimum doping level in the intrinsic layer, the drift velocity is accelerated and the resistivity that is needed to maintain the high electric field can be achieved. The higher electric field means a shorter transit time. Another way of shortening the transit time is to shorten the length of the intrinsic region, but the penalty for this is a reduction in the efficiency of converting the incident photons into electrons.

2. Referring to Fig. 12.5, there is a chance that an incident photon will create an electron–hole pair inside the p^+-type layer before reaching the i-type layer. For such carriers inside the p^+-type layer, the electric field is weak due to the higher conductivity. The carriers must then depend on diffusion to reach the electrodes, which is a much slower process than electron drift in the presence of an external field. These carriers not only slow down the response of the photodiode but also become a source of detector noise. Such problems are minimized by designing the p^+-type layer as thin as possible.

3. The time constant RC associated with the external circuit of the photodiode also limits the frequency response of the detector. Referring to the equivalent circuit in Fig. 16.13a, this time constant will be calculated. The potential drop across R_s, and the current through R_{sh}, and the impedance of the coupling capacitor C_c are assumed negligibly small.

The photocurrent i from Fig. 16.13a is

$$i = i_s - i_d \tag{16.11}$$

where i_s is the signal current, and i_d is the diode current as given by Eq. (16.10).

With the simplified equivalent circuit in Fig. 16.13b, the photocurrent is divided between the capacitor C and resistor R_L, where

$$\begin{aligned} C &= C_s + C_d \\ \frac{1}{R_L} &= \frac{1}{R_a} + \frac{1}{R_b} \end{aligned} \tag{16.12}$$

The output voltage is calculated as

$$\begin{aligned} V &= \frac{R_L / j\omega C}{R_L + 1/j\omega C} i \\ &= \frac{R_L}{\sqrt{1 + (\omega R_L C)^2}} e^{j\phi} \cdot i \end{aligned} \tag{16.13}$$

where

$$\tan\phi = -\omega R_L C \tag{16.14}$$

The cutoff frequency is

$$f_c = \frac{1}{2\pi R_L C} \tag{16.15}$$

The cutoff frequency determines the upper limit of the modulation frequency. One can increase this upper limit by reducing the value of $R_L C$. This can be accomplished by selecting a detector diode with a small value of C, or reducing the load R_L to the diode. However, in the latter case, one should note that reduction of R_L results also in a reduction of the output voltage V as shown by Eq. (16.13).

The above description of the PIN diode characteristics also applies to the APD except for the higher gain.

Next, the coupling of current converted from the incident light to the preamplifier will be explained.

16.4.3 Coupling Circuits to a Preamplifier

Special attention is needed in the design of the circuit that couples the detector with the preamplifier. Like any other circuit design, the following three considerations are important:

1. Minimization of mismatch.
2. Wide frequency band operation.
3. Optimized signal to noise ratio.

Depending on the region of the operating frequency, the circuit design is broadly divided into two approaches. One is the lumped element approach for frequencies below a gigahertz, and the other is the distributed element approach for frequencies above this level. For midrange frequencies, hybrid circuits of the two are used.

16.4.3.1 Coupling Circuits to a Preamplifier at Subgigahertz

Examples of the lumped element approach are described next.

A 50-Ω Circuit with Mismatch Figure 16.14 shows a PIN detector circuit [19]. The photodiode is loaded with a 50-Ω resistor, and its output is fed directly to a 50-Ω input impedance preamplifier. R_b prevents the photocurrent going through the battery as well as prevents damage to the preamplifier in the event that the photodiode becomes short-circuited. An advantage of such a circuit is that any length of 50-Ω coaxial cable can be used to connect the photodiode to a standard 50-Ω amplifier, making it convenient for putting instrumentation together quickly. A disadvantage is the low sensitivity of the circuit to the input voltage due to the low value of the 50-Ω load in comparison to the high internal impedance of the photodiode. Furthermore, at higher frequencies, the reflection at the junction of the diode to the 50-Ω lead wire severely degrades the performance. Another disadvantage of this circuit is the high average square thermal noise current. As will be shown in Section 16.5.2, the average square thermal noise current is inversely proportional to the load resistance (refer to Eq. (16.27)).

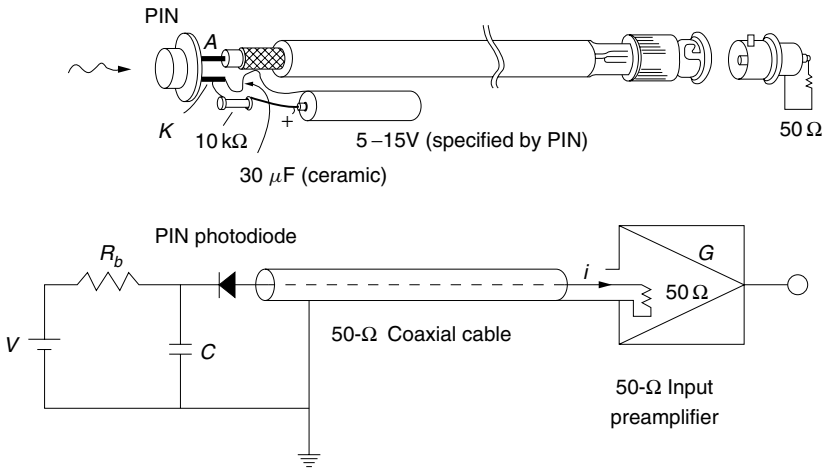


Figure 16.14 PIN photodiode connected to a 50-Ω input preamplifier.

High Impedance with a Compensating Circuit (HZ Circuit) Another circuit design involves using a high load impedance. An advantage of a high load impedance is a low average square thermal noise current, as detailed later. The problem with this design, however, is a low cutoff of the frequency response given by Eq. (16.15). One means of dealing with this problem is to combine the high impedance load with a compensator circuit that raises the cutoff frequency.

Figure 16.15 shows an example of such a circuit. The input to the operational amplifier is essentially a high input impedance low-pass filter. The output circuit of the operational amplifier is made a high-pass filter to compensate the input low-pass filter. The output of the operational amplifier is connected to the R_1C_1 parallel circuit, which is then connected in series with R_L. The output voltage from the operational amplifier is divided between the impedance of the R_1C_1 parallel circuit and R_L. As the frequency is raised beyond a critical frequency f_{c_1} the portion of the potential drop across R_L becomes large compared to that of the R_1C_1 parallel circuit. The critical frequency is the frequency above which the reactance $1/\omega_1 C_1$ is smaller than R_1, and

$$f_{c_1} = \frac{1}{2\pi R_1 C_1}$$

Next, let us design an output circuit that neither overcompensates nor undercompensates the input low-pass filter. If the input impedance of the operational amplifier is assumed infinity, the input voltage V to the operational amplifier becomes the same as the value already given by Eq. (16.13).

The output voltage V_O is

$$\begin{aligned} V_O &= VG \frac{R_L}{R_L + \dfrac{R_1}{1 + j\omega C_1 R_1}} \\ &= \frac{iGRR_L}{(1 + j\omega CR)} \cdot \frac{(1 + j\omega C_1 R_1)}{R_1 + R_L + j\omega C_1 R_1 R_L} \end{aligned} \tag{16.16}$$

where G is the gain of the amplifier and i is the input current.

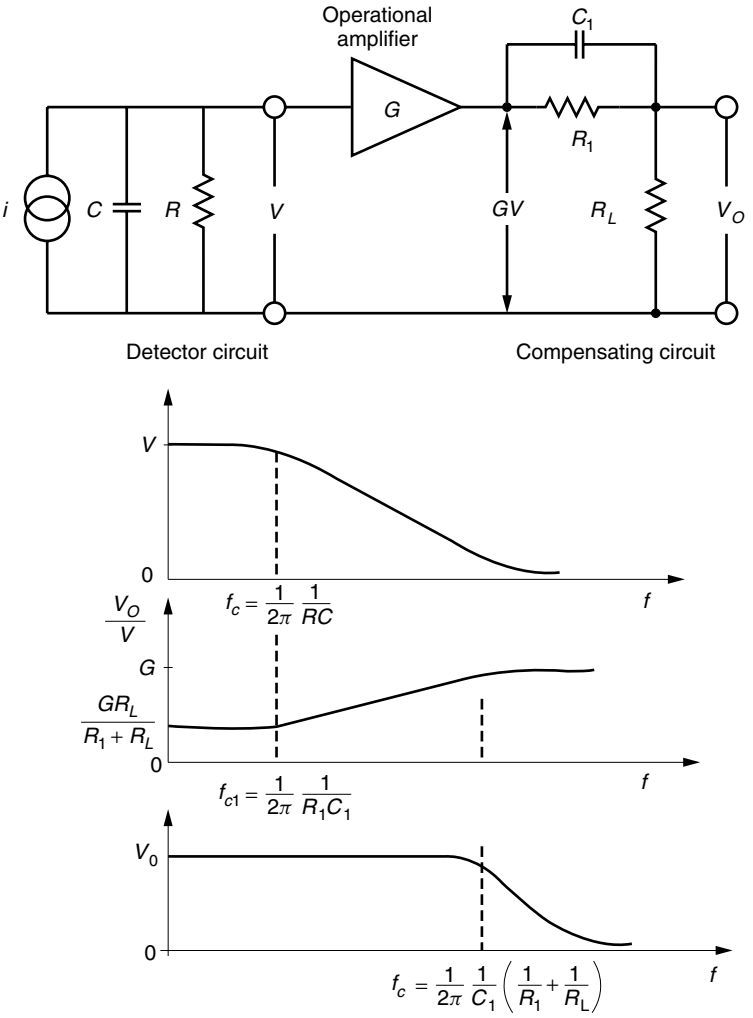


Figure 16.15 High impedance with a compensation circuit.

Proper compensation is achieved if the denominator of the first factor cancels the numerator of the second factor, namely,

$$R_1 C_1 = RC \tag{16.17}$$

The condition, therefore, is that the cutoff frequency of the high-pass filter matches that of the low-pass filter. It follows from Eq. (16.17) that the cutoff frequency of the overall circuit is

$$f_c = \frac{1}{2\pi} \cdot \frac{1}{C_1}\left(\frac{1}{R_1} + \frac{1}{R_L}\right) \tag{16.18}$$

The advantages of such a preamplifier circuit are high sensitivity in a wide frequency band and low average square thermal current due to the high input impedance of the

operational amplifier. The disadvantage, however, is that if f_c is stretched too far, the signal V fed into the operational amplifier becomes too small and soon reaches the noise level of the operational amplifier, causing the minimum detectable power to suffer.

Transimpedance Circuit (TZ) Another well-established method for widening the frequency band of a high input impedance amplifier is the use of a negative feedback circuit, such as the one shown in Fig. 16.16. The current i_f and the voltages V_0 and V are given by

$$i_f = \frac{V_O - \mathrm{V}}{R_f} \tag{16.19}$$

$$V = \frac{i + i_f}{1/R + j\omega C} \tag{16.20}$$

$$V_O = -GV \tag{16.21}$$

Observe that the input current to the operational amplifier is essentially zero.

Briefly explaining the operation, an increase in V_O increases i_f as shown by Eq. (16.19). An increase in i_f increases V across R and C. The increase in V then

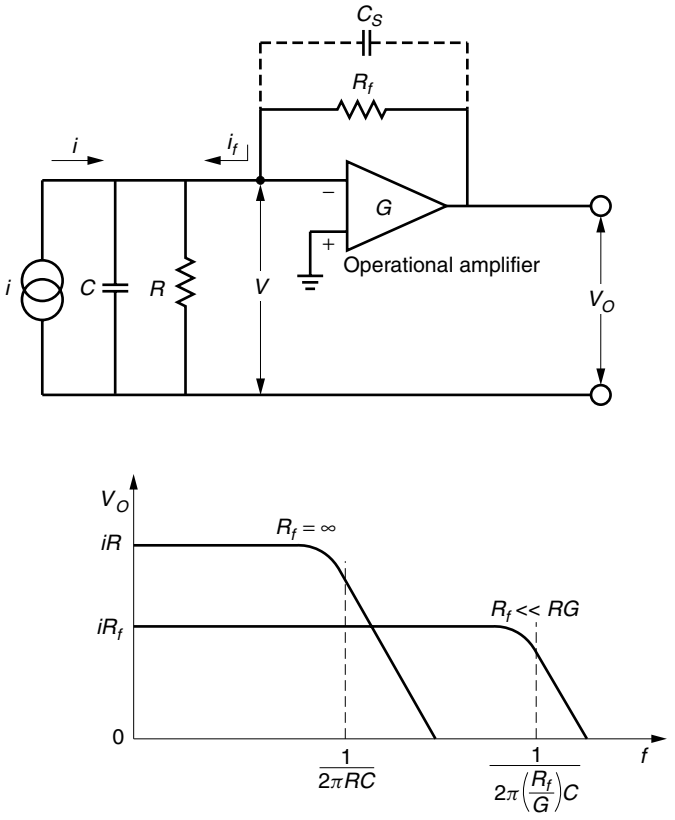


Figure 16.16 Transimpedance coupling.

decreases V_O. First, eliminating i_f from Eqs. (16.19) and (16.20) and then using Eq. (16.21) gives

$$V_O = -\frac{R_f i}{1 + \dfrac{1}{G} + \dfrac{R_f}{RG} + j\omega C\dfrac{R_f}{G}}$$

If one chooses

$$G \gg 1 \quad \text{and} \quad RG \gg R_f \tag{16.22}$$

then V_O becomes

$$V_O = -\frac{R_f i}{1 + j\omega C(R_f/G)} \tag{16.23}$$

Thus, the cutoff frequency is

$$f_c = \frac{1}{2\pi(R_f/G)C} \tag{16.24}$$

Inserting $R_f \ll RG$ into Eq. (16.24) gives

$$f_c \gg \frac{1}{2\pi RC}$$

This cutoff frequency is much higher than the cutoff frequency of Eq. (16.15) without compensation. One can raise f_c further by lowering R_f. The lowering of R_f, however, means lowering the output voltage, which is essentially $R_f i$ from Eq. (16.23).

As far as the output voltage is concerned, compared to the previously mentioned 50-Ω circuit with mismatch in Fig. 16.14, the transimpedance circuit gives a much larger output. If both circuits use an operational amplifier of gain G, the output from the 50-Ω circuit with mismatch is $50iG$. The output from the transimpedance circuit in Fig. 16.16 gives approximately iR_f and if $R_f > 50G$, then the transimpedance circuit gives a higher output.

Any stray capacitance C_s around the feedback resistance R_f severely affects the frequency bandwidth (Problem 16.7).

16.4.3.2 Coupling Circuits to a Preamplifier Above a Gigahertz

The high impedance (HZ) [20,21] and transimpedance (TZ) [22,23] approaches are still used up to few gigahertz, provided that ultrawide frequency band elements are used, and provided that sensor capacitance and stray capacitance are minimized. Above several gigahertz the distributed amplifier approach is used.

HZ and TZ with Special Considerations The stray capacitance or series resistance of the bonding wire or inductor wire should be minimized.

In order to increase the cutoff frequency as given by either Eq. (16.18) or (16.24), it is important to reduce C. For instance, by reducing the diameter of the sensor area of the photodiode to 30 μm, C can be reduced to 0.1 pF.

Operational amplifiers whose cutoff frequencies are normally subgigahertz can no longer be used in the gigahertz frequency range. These have to be replaced by ultrawide frequency band elements such as the GaAs metal–semiconductor field effect transistor

(GaAs MESFET), the high electron mobility transistor (HEMT), or the heterojunction bipolar transistor (HBT). Their cutoff frequencies are several tens of gigahertz. When an ultrawide frequency band element is used, it is important to install an adequately designed low-pass filter in order to prevent the noise spectra outside the required bandwidth from deteriorating the signal.

Distributed Amplifier Approach When the required frequency bandwidth exceeds several gigahertz, another approach such as the distributed preamplifier [24–26] has to be employed. Figure 16.17 shows a schematic diagram of distributed preamplifiers. The distributed FET amplifiers form two artificial transmission lines. The inductors L_g, which are connected to the gate terminals of the FETs and the gate–source capacitance C_{gs} of the FETs, form one transmission line with characteristic impedance $Z_g = \sqrt{L_g/C_{gs}}$, while the inductors L_d, which are connected to the drain terminals and the drain–source capacitors C_{ds}, form another transmission line with characteristic impedance $Z_d = \sqrt{L_d/C_{ds}}$.

A photodiode with the proper value of the load inductor L_p can smoothly be amalgamated into the artificial transmission line as an additional line element.

Each transmission line is loaded with impedances that match the characteristic impedance of the line in order to eliminate reflection for all signal frequencies. Thus, a wide frequency band operation is accomplished.

Referring to Fig. 16.17, there are two possible directions for the output of each FET to propagate along the drain transmission line; one is forward, and the other backward. First, consider only the backward wave in the vicinity of Q_1. There are two ways of reaching point d_1 from point g_1. The signal at g_1 is split into two components. One takes the direct route through Q_1 and the other takes an indirect route through Q_2 via g_1–g_2–Q_2–d_2–d_1. Both components go through the FET amplifier once, but their path lengths are different. By setting this path difference to be one-half wavelength, these two components cancel each other, and the backward wave is suppressed. As for the forward wave from g_1 to d_2, there is no difference between the path lengths

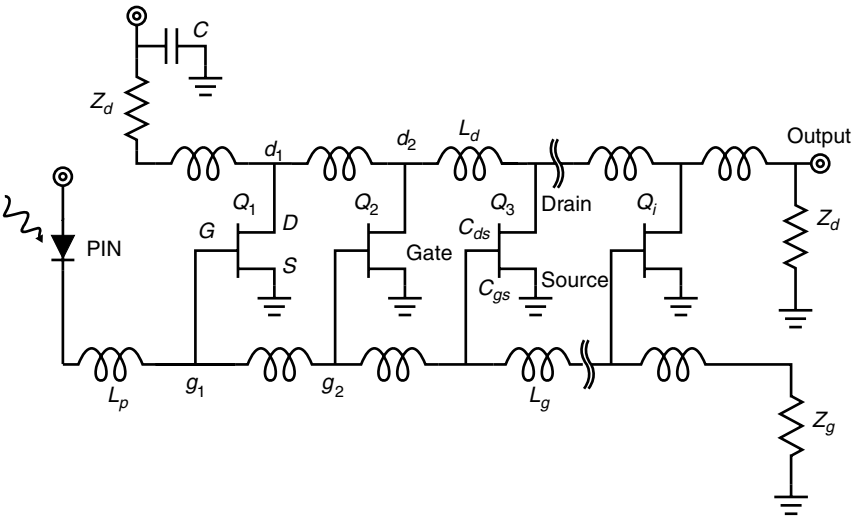


Figure 16.17 Diagram of distributed preamplifiers. (After A. P. Freundorfer and P. Lionais [25].)

of $g_1 - g_2 - d_2$ and $g_1 - d_1 - d_2$ and the two components add. The forward wave is enhanced.

The transmission line being artificial, it behaves like a low-pass filter and there is a cutoff frequency of

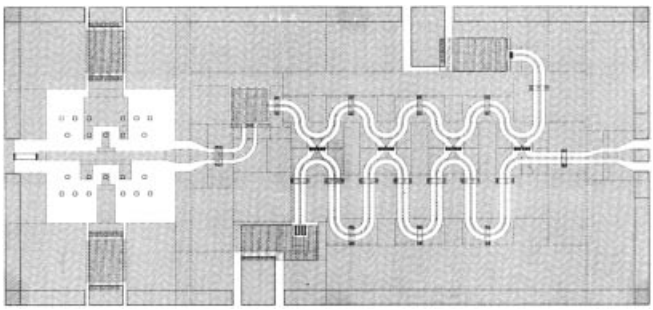
$$f_{cg} = \frac{1}{\pi\sqrt{L_g C_{gs}}} = \frac{1}{\pi Z_g C_{gs}} \tag{16.25}$$

and f_{cg} limits the frequency bandwidth of the receiver. The cutoff frequency f_{cg} can be increased by reducing Z_g but the thermal noise i_{th} (to be explained in Section 16.5.2) also increases with a reduction of Z_g. The thermal noise generated in the system is predominantly determined by Z_g and is approximately

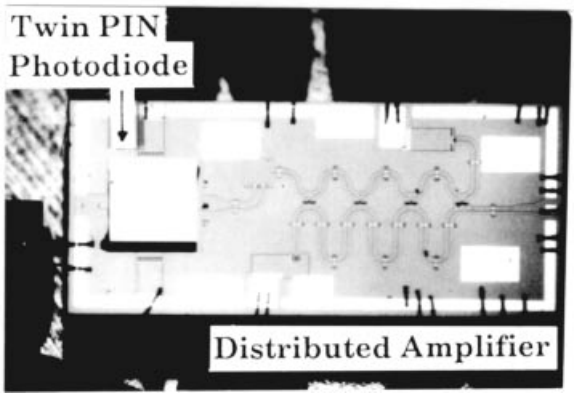
$$i_{\text{th}}^2 = 4kTB/Z_g$$

Consequently, Z_g has to be set at a compromised value.

Figure 16.18 shows a photograph of a receiver consisting of a PIN photodiode and the distributed FET amplifiers. Such a distributed amplifier has a frequency bandwidth of 23 GHz. The gain of the amplifier is expressed in terms of the transimpedance



(a)



(b)

Figure 16.18 Balanced optical receiver. (a) Preamplifier pattern layout. (b) Realized optical receiver. (Courtesy of N. Takachio et al. [27].)

$Z_T = V_O/i$, where V_O is the output voltage and i is the photocurrent from the photodiode. The transimpedance is 46 dB · Ω (or approximately 40,000 Ω) [27].

16.5 NOISE IN THE DETECTOR SYSTEM

The minimum detectable power of the receiver is determined from noise considerations. The two major sources of noise are shot noise and thermal noise.

16.5.1 Shot Noise

The flow of electrons across the p-n junction is like pouring a bucket of beans into another bucket. Each bean arrives discretely and randomly. Each discrete electron is equivalent to a δ function of current whose Fourier transform has a constant spectrum in the frequency domain, giving a white noise characteristic. The quantity used to characterize the shot noise is the mean square value of the noise current

$$\langle i_{\text{shot}}^2 \rangle = 2eIB \tag{16.26}$$

where B is the bandwidth of the receiver and I is the average current through the detector. The shot noise current is $\sqrt{2eIB}$ and increases with $\sqrt{B}$. The shot noise current per $\sqrt{\text{Hz}}$ is $\sqrt{2eI}$.



Flow of electron ''beans'' generates shot noise.

Not only the discreteness but also the randomness of the arrival time is important. The number of electrons in I amperes of flow for 1 second is $N = I/e$. If these electrons were precisely equally spaced, then a periodic flow of current would be generated with period (e/I) s that can be expressed as $\Pi(t/(e/I))$ (see Section 1.4.6). The fundamental frequency component of such a hypothetical current would be $f_0 = I/e$. The fundamental frequency f_0 is already a big number. For instance, if $I = 10^{-6}$ A, then $e = 1.59 \times 10^{-19}$ coulomb gives $f_0 = 6300$ GHz. There would not be any spectrum between dc and 6300 GHz!

16.5.2 Thermal Noise

Any heated substance emits radiation. Even at room temperature, there will be some thermal radiation, and hence an associated thermal noise. The only way to shut off thermal noise completely is to cool the substance to absolute zero on the Kelvin scale.

The thermal noise power per unit spectrum is kT, and the mean square of the noise current generated inside the resistor is

$$\langle i_{th}^2 \rangle = 4kTB/R \tag{16.27}$$

where R is the resistance of the resistor considered as the noise source, k is Boltzmann's constant 1.38×10^{-23} J·K, T is the temperature of the resistor on the Kelvin scale, and B is the bandwidth. We will sidetrack briefly to explain how Eq. (16.27) was derived.

When a piece of resistor is heated, it becomes an electromagnetic wave radiator. For that matter, any material becomes a good radiator when heated. The radiating power spectrum contained in the frequency interval between ν and $\nu + \Delta\nu$ is given by Planck's factor

$$p(\nu) = \frac{h\nu}{e^{h\nu/kT} - 1} \tag{16.28}$$

whose maximum value shifts with temperature.

Now the heated resistor is connected to a load resistor, which extracts energy from the heated source resistor. When the resistor is piping hot, the mechanism of energy transfer is primarily due to blackbody radiation in free space, which radiates and also heats up the load resistor. The amount of energy transfer does not depend on the value of the resistance of the load; it primarily depends on the effective cross-sectional area and the emitted infared wavelength of the heated resistor. Even though the amount of energy transfer to the load can be optimized by shaping and "coloring" the load resistor, the one invariant parameter that cannot be changed is the maximum available power from the source. This is set by Planck's factor $p(\nu)\,d\nu$.

The main difference between the cooled down and piping hot case is the means of energy transfer to the load resistor. Next, the case when the source resistor is cooled down to room temperature is considered. The resistor still continues to radiate but the amount of energy radiated is reduced and its radiated spectrum is shifted toward a lower frequency. The shifted frequency spectrum is now too low to radiate effectively. The noise power, however, can conduct very effectively through the connecting copper wire to the load as a noise current. In this case, the amount of transfer of energy depends on the resistance of the load while the maximum available power is again determined by Planck's factor, Eq. (16.28).

Next, a mathematical model for an equivalent circuit of the resistor as a noise source is developed. The model is a resistor with the same resistance but paralleled with a constant noise current source as encircled in Fig. 16.19. The value of the noise generator current i_N is determined by setting the maximum available power to a fictitious load R_1 equal to the invariant value of $p(\nu)\,d\nu$ as follows.

With the equivalent circuit in Fig. 16.19, the transferred power P_1 to the load is

$$P_1 = \left\langle \left(i_N \frac{R}{R + R_1} \right)^2 \right\rangle R_1 \tag{16.29}$$

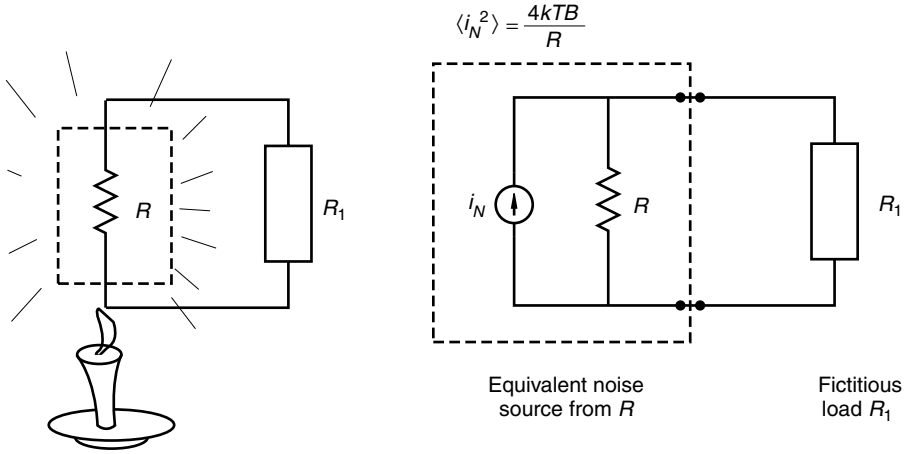


Figure 16.19 The equivalent circuit of noise from resistor R. $\langle i_N^2 \rangle$ is determined such that the noise source delivers a noise power of kTB when R_1 is selected to be R.

The maximum available power is found from $dP_1/dR_1 = 0$ and occurs when $R_1 = R$. Therefore, the maximum available power becomes

$$\frac{\langle i_N^2 \rangle}{4} R = p(\nu)\, d\nu \tag{16.30}$$

When

$$\frac{h\nu}{kT} \ll 1$$

Eq. (16.28) can then be approximated as

$$p(\nu)\, d\nu = kT\, d\nu \tag{16.31}$$

At room temperature, this condition is satisfied for

$$\nu \ll 6.2 \text{ THz} \tag{16.32}$$

Finally, inserting Eq. (16.31) into (16.30) gives the noise current generated from the resistor R in bandwidth B expressed by Eq. (16.27).

16.5.3 Signal to Noise Ratio

The signal to noise ratio S/N is used to quantitatively represent the quality of the detecting system. Initially, the incident light is assumed to be unmodulated. The detection system is a direct detection system such as shown in Fig. 16.13a with its equivalent circuit in Fig. 16.13b.

With incident light power P_s of the PIN diode, the output electrical signal current i_s is, from Eq. (12.8),

$$i_s = \frac{\eta e}{h\nu} P_s \tag{16.33}$$

The electrical signal power S delivered to the load R_L is $S = i_s^2 R_L$ and

$$S = \left(\frac{\eta e}{h\nu}\right)^2 P_s^2 R_L \tag{16.34}$$

The shot noise power N_{shot} is, from Eq. (16.26),

$$N_{\text{shot}} = 2e\left(\frac{\eta e}{h\nu} P_s + I_d\right) R_L B \tag{16.35}$$

where I_d is the dark current and is essentially equal to I_{so} in Eq. (16.10). The thermal noise power is generated in the load resistor. The noise power N_{th} generated from resistance R_L is

$$N_{\text{th}} = \frac{4kTB}{R_L} R_L = 4kTB \tag{16.36}$$

The signal to noise ratio of the PIN diode taking both shot and thermal noise into consideration, therefore, is

$$\frac{\text{S}}{\text{N}} = \frac{\left(\frac{\eta e}{h\nu}\right)^2 P_s^2}{2e\left(\frac{\eta e}{h\nu} P_s + I_d\right) B + \frac{4kTB}{R_L}} \tag{16.37}$$

As seen from Eq. (16.37), the behavior of the S/N with regard to P_s depends on whether shot noise or thermal noise is predominant. When the shot noise contribution is predominant, the S/N is said to be *shot noise limited or quantum limited.* On the other hand, when the thermal noise is predominant, it is said to be *thermal noise limited*.

For most PIN diodes, $\eta e/h\nu = 0.5$ A/W and $I_d = 2$ nA, and as long as P_s is larger than 4 nW, the contribution of I_d in the parentheses in the denominator of Eq. (16.37) is not important. When P_s is smaller than 4 nW, a more expensive low dark current photodiode is necessary to detect the light.

In the quantum-limited case, Eq. (16.37) becomes

$$\frac{\text{S}}{\text{N}} = \frac{\eta}{2h\nu B} P_s \tag{16.38}$$

where the assumption was made that

$$P_s \gg 4 \text{ nW} \tag{16.39}$$

Note that the S/N increases linearly with P_s and decreases linearly with B. In order to maintain the same S/N for a wider frequency of operation, P_s must be raised accordingly.

Next, in the thermal-noise-limited case, Eq. (16.37) becomes

$$\frac{\text{S}}{\text{N}} = \frac{\left(\frac{\eta e}{h\nu}\right)^2}{\frac{4kTB}{R_L}} P_s^2 \tag{16.40}$$

This time the signal to noise ratio increases quadratically with P_s and is more sensitive to P_s than in the quantum-limited case, and it increases with R_L while maintaining the same dependence on B. In the thermal-noise-limited case, a low dark current photodiode is completely unnecessary.

16.5.4 Excess Noise in the APD

As shown in Fig. 12.5b, the APD consists of a photoelectric converter layer and a multiplier layer. The noise from the APD is normally characterized by a combination of the noise in these two layers.

When the APD is back-biased, the internal resistance is on the order of megohms. The shot noise current generated in the photoelectric converter layer, $i_{\text{shot}} = \sqrt{2eIB}$, is also multiplied by the multiplication factor M in the multiplier layer; thus, the mean square value of the shot noise from the APD becomes

$$i_{\text{shot}}^2 = 2eB\left(\frac{\eta e}{h\nu}P_s + I_d\right)M^{2+x} \tag{16.41}$$

Note that Eq. (16.41) contains an extra factor M^x, which is called the excess noise (x is called the excess noise index). The excess noise is due primarily to the noise generated during the avalanche effect. The value of x is determined by the constituent materials of the APD. For example, $x \approx 0.3$–0.5 for silicon and $x \approx 0.9$–1.0 for germanium. Sometimes, a high-quality PIN diode is more desirable than an APD in the presence of a high excess noise index.

The signal to noise ratio of an APD is

$$\frac{\text{S}}{\text{N}} = \frac{\left(\frac{\eta e}{h\nu}P_s M\right)^2}{2eB(I_s + I_d)M^{(2+x)} + 4kTB/R_L} \tag{16.42}$$

where $I_s = \eta e/h\nu P_s$. Note that I_s is the average signal current before avalanche multiplication. From this equation we see that although the output power increases with M^2, the shot noise increases more rapidly with M^{2+x}; thus, the S/N starts to decrease beyond a certain value of M as shown in Fig. 16.20. A larger M is not necessarily better, and the back-biased voltage should be adjusted for the proper value of M, which is normally a few hundred.

16.5.5 Noise Equivalent Power (NEP)

The minimum detectable light power is a quantity that is often used for designing optical communication systems. The minimum detectable light power $P_{s\,\min}$ is defined as the unmodulated light level for which the output electrical signal can barely be detected. This value can be obtained by setting the S/N in Eq. (16.42) equal to unity, giving

$$P_{s\,\min} = \frac{h\nu}{\eta e}\sqrt{2eI_d M^x + \frac{4kT}{M^2 R_L}}\sqrt{B} \tag{16.43}$$

where

$$I_s \ll I_d$$

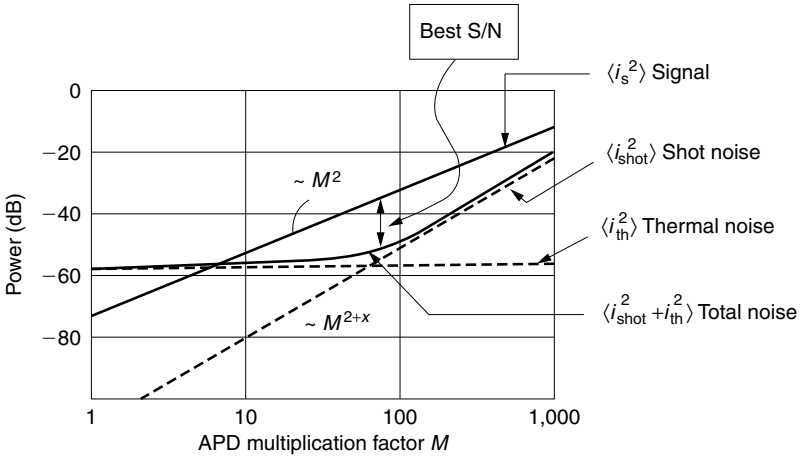


Figure 16.20 Dependence of signal and noise powers of an APD on the multiplication factor M.

was assumed. The NEP is defined as

$$P_{s\,\min} = \text{NEP} \cdot \sqrt{B} \tag{16.44}$$

hence,

$$\text{NEP} = \frac{h\nu}{\eta e}\sqrt{2eI_d M^x + \frac{4kT}{M^2 R_L}} \quad \text{W}/\sqrt{\text{Hz}} \tag{16.45}$$

The NEP is used to quantify the sensitivity of the detector and is the minimum detectable power for a 1-Hz signal bandwidth. The strange unit of $\text{W}/\sqrt{\text{Hz}}$ is used because the minimum detectable power is proportional to the square root of the frequency band.

Example 16.1 A detection system employs a silicon avalanche photodiode Type S2383 listed in the Hamamatsu Photonics catalogue. The characteristics of the APD are given in Fig. 16.21. The operational parameters are as follows. The input light power P_s (unmodulated) is 10 nW, and the wavelength λ is 0.83 μm. The cutoff frequency f_c is 50 MHz, and the temperature is 20°C.

Assume that the APD characteristics given in Fig. 16.21 for 25°C are the same for 20°C. Answer the following questions.

(a) What is the responsivity R? Responsivity was defined in Section 12.4.2 as

$$R = \frac{\eta e}{h\nu} M$$

(b) What is the relevant loaded resistance R_L of the APD?

(c) What is the electrical signal power, the shot noise power, and the thermal noise power?

(d) What is the S/N?

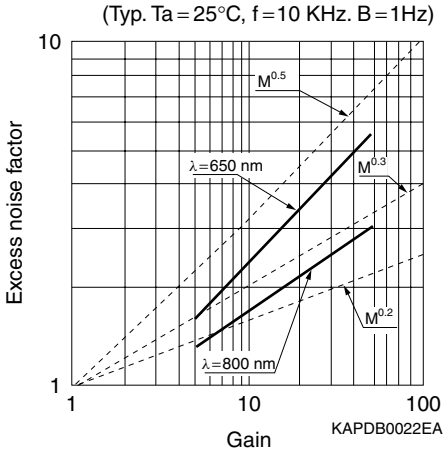
(e) Is the S/N quantum limited or thermal noise limited?

(f) What is the value of the NEP?

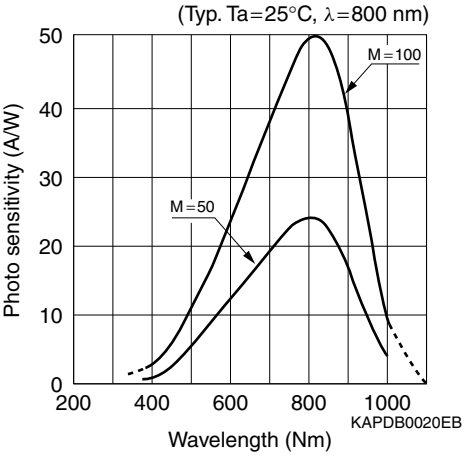
Si APDs (Low-bias operation types, for 800 nm range)

Type No.	Spectral response range λ (nm)	Peak sensitivity wavelength λp (nm)	Photo *4 sensitivity M = 1 λ = 800 nm (A/W)	Quantum Efficiency M = 1 λ = 800 nm (%)	Dark current *4 I_D (nA) Typ.	Dark current *4 I_D (nA) Max.	Cut-off *4 frequency f_c R_L = 50 Ω (MHz)	Terminal *4 capacitance Ct (pF)	Excess *4 noice figure x (pF)	Gain M λ = 800nm	Type No.
S2381	400 to 1000	800	0.5	75	0.05	0.5	100	1.5	0.3	100	S2381
S2382					0.1	1	900	3			S2382
S5139											S5139
S2383					0.2	2	600	6			S2383
S2383-10 *2											S2383-10
S3884					0.5	5	400	10			S3884
S2384					1	10	120	40		60	S2384
S2385					3	30	40	95		40	S2385

● Excess noise factor vs. gain



● Spectral response



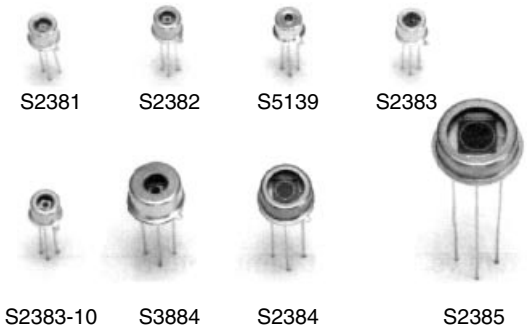


Figure 16.21 APD characteristics. (From Hamamatsu Photonics catalogue [19].)

Solution The relevant parameters in Fig. 16.21 are:

$$P_s = 10 \text{ nW}$$
$$\eta = 0.75$$
$$I_d = 0.2 \text{ nA}$$
$$C_f = 6 \text{ pF}$$

$x = 0.3$
$M = 100$
$T = 293$ K

(a) The responsivity of the APD is

$$R = \frac{\eta e}{h\nu} M = \frac{(0.75)(1.6 \times 10^{-19})(100)}{(6.63 \times 10^{-34})(3.6 \times 10^{14})} = 50.2 \text{ A/W}$$

(b) The load impedance for $f_c = 50$ MHz is, from Eq. (16.15),

$$R_L = \frac{1}{2\pi f_c C_f} = \frac{1}{2\pi(5 \times 10^7)(6 \times 10^{-12})} = 531\ \Omega$$

(c) The electrical signal power is

$$\begin{aligned} S &= \left(\frac{\eta e}{h\nu} M P_s\right)^2 R_L = (RP_s)^2 R_L \\ &= [(50.2)(10^{-8})]^2(531) = 1.34 \times 10^{-10}\ W \end{aligned}$$

From Eq. (16.41), the shot noise power is

$$\begin{aligned} N_{\text{shot}} &= 2eB\left(\frac{\eta e}{h\nu} P_s + I_d\right) M^{2+x} R_L \\ &= 2(1.6 \times 10^{-19})(5 \times 10^7)\{(0.502)(10^{-8}) + 0.2 \times 10^{-9}\}(100^{(2+0.3)})(531) \\ &= 1.77 \times 10^{-12}\ W \end{aligned}$$

From Eq. (16.36), the thermal noise power is

$$\begin{aligned} N_{\text{th}} &= 4kTB \\ &= 4(1.38 \times 10^{-23})(293)(5 \times 10^7) \\ &= 8.09 \times 10^{-13}\ W \end{aligned}$$

(d) $\text{S/N} = S/(N_{\text{shot}} + N_{\text{th}}) = 1.34 \times 10^{-10}/1.77 \times 10^{-12} + 8.09 \times 10^{-13} = 52.0$
(e) It is quantum limited.
(f) From Eq. (16.45), the NEP is

$$\begin{aligned} \text{NEP} &= \frac{h\nu}{\eta e}\sqrt{2eI_d M^x + \frac{4kT}{M^2 R_L}} \\ &= \frac{1}{0.502}\sqrt{2(1.6 \times 10^{-19})(0.2 \times 10^{-9})(100)^{0.3} + \frac{4(1.38 \times 10^{-23})(293)}{100^2(531)}} \\ &= 1.15 \times 10^{-13}\ \text{W}/\sqrt{Hz} \end{aligned}$$

□

Example 16.2

(a) The same system as given in Example 16.1 is to be used at much lower input light power. What is the minimum detectable light power?

(b) The cutoff frequency of the system described in Example 16.1 was lowered to 5 MHz and the load impedance R_L was raised accordingly. What is the new value of S/N?

Solution

(a) From Eq. (16.44), the minimum detectable light power is

$$P_{\min} = \text{NEP}\sqrt{B} = (1.15 \times 10^{-13})(\sqrt{5 \times 10^7}) = 0.81 \text{ nW}$$

(b) By reducing $f_c(= B)$ to $f_c/10$, not only is the noise bandwidth reduced to $1/10$, but R_L is also increased by 10 times. From result (d) of Example 16.1,

$$\frac{\text{S}}{\text{N}} = \frac{1.34 \times 10^{-10}(10)}{1.77 \times 10^{-12}\left(\frac{1}{10}\right)(10) + 8.09 \times 10^{-13}\left(\frac{1}{10}\right)} = 724$$

and the S/N is increased about 24 times. □

16.5.6 Signal to Noise Ratio for ASK Modulation

What modifications should be made if the intensity of the incident light is modulated? The effect of modulation on signal to noise ratio is examined for amplitude shift keying (ASK) mentioned in Section 16.2.7. The current from the detector Eq. (12.8) has to be modified as

$$i_s(t) = \frac{\eta e}{h\nu} M P_s A(t) \tag{16.46}$$

where $A(t)$ is the ASK modulation and is either 1 or 0.

The average square signal current, therefore, is

$$\langle i_s^2 \rangle = \frac{1}{2}\left(\frac{\eta e}{h\nu} M P_s\right)^2 \tag{16.47}$$

where use was made of the relationship

$$[A(t)]^2 = A(t) \tag{16.48}$$

and the average of the "0"s and the "1"s of $A(t)$ is $1/2$.

It is important to remember that the signal electrical power $\langle i_s^2 \rangle R_L$ is proportional to the square of the signal optical power P_s, where P_s is the power for the "1" bit.

The signal to noise ratio of an APD is

$$\frac{\text{S}}{\text{N}} = \frac{\frac{1}{2}\left(\frac{\eta e}{h\nu} P_s M\right)^2}{2eB(I_s + I_d)M^{(2+x)} + 4kTB/R_L} \tag{16.49}$$

where $I_s = \frac{1}{2}\eta e/h\nu P_s$ and I_d is the amplitude of the dark current.

Another point is that I_s is the average signal current before avalanche multiplication. For the ASK modulation it has a factor of $\frac{1}{2}$ because the light is off half the time.

The minimum light power to establish S/N = 1 for the ASK modulated light is

$$P_{s\,\min}^{\text{ASK}} = \sqrt{2}\frac{h\nu}{\eta e}\sqrt{2eI_d M^x + \frac{4kT}{M^2 R_L}}\sqrt{B} \tag{16.50}$$

which is $\sqrt{2}$ times that of the unmodulated incident light given by Eq. (16.43). It might be added that the NEP is always defined with unmodulated light.

16.5.7 Signal to Noise Ratio of Homodyne Detection

The output signal current from homodyne detection ($f_{\text{IF}} = 0$) without modulation is given by Eq. (12.22). If the signal is ASK modulated, the output electrical signal power is

$$S = \left(2\frac{\eta e}{h\nu}\right)^2 P_S P_L M^2 \overline{A(t)^2} \tag{16.51}$$

and the signal to noise ratio of homodyne detection for ASK coded signal light is

$$\frac{\text{S}}{\text{N}} = \frac{\left(\dfrac{2\eta e}{h\nu}\right)^2 P_s P_L M^2 \overline{A(t)^2}}{2eB_{IF}\left(\dfrac{\eta e}{h\nu}P_L + I_d\right)M^{2+x} + \dfrac{4kTB_{IF}}{R_L} + N_{\text{LO}}} \tag{16.52}$$

N_{LO} is the local oscillator intensity noise (LOIN) power, which is the noise due to the fluctuation of the local oscillator intensity. This, however, can be removed by using the balanced mixer, as mentioned in Section 12.7. For a large local oscillator laser power P_L, the thermal noise power can be ignored and the detector becomes quantum limited, almost regardless of R_L, and can provide the ultimate limit in sensitivity. Equation (16.52) can be approximated as

$$\frac{\text{S}}{\text{N}} = \frac{\eta}{M^x}\left(\frac{P_s}{h\nu}\right)\frac{1}{B_{\text{IF}}} \tag{16.53}$$

where use was made of $\overline{A^2(t)} = \overline{A(t)} = \frac{1}{2}$. Note that the IF bandwidth B_{IF} is twice the bandwidth B of the signal, as mentioned in Section 16.3.2.

Comparing the signal to noise ratios of direct detection by an APD of Eq. (16.49) to that of homodyne detection of Eq. (16.53), one realizes that the S/N of the direct detection is proportional to P_s^2 in the lower limit that the I_s term in the denominator of Eq. (16.49) is small, whereas the signal to noise ratio of the homodyne detection is proportional to P_s. This difference is significant for the commonly small values of P_s.

Let us examine Eq. (16.53) more closely. The factor in the parentheses is the number of photons coming in per unit time. As mentioned above, $B_{\text{IF}} = 2B$. From Section 16.6.5, the bit rate B_t for the NRZ code is $B = B_t/2$ and $B_{\text{IF}} = B_t$. The number of electrons contained in one pulse is $(P_s/h\nu)(1/B_t)$. If η/M^x is assumed to be a typical

value such as 0.4, then we can conclude that only 2.2 photons in one pulse is good enough to achieve a signal to noise ratio of 1!

16.5.8 Borderline Between the Quantum-Limited and Thermal-Noise-Limited S/N

It is useful to know the borderline between the regions of the quantum-limited and thermal-noise-limited S/N to facilitate the calculation of the required light power for a given S/N. The borderline is calculated from the denominator of Eq. (16.42) by setting $N_{\text{shot}} = N_{\text{th}}$:

$$2e\frac{\eta e}{h\nu}M^2P_s^a = \frac{4kT}{R_L} \tag{16.54}$$

where the contributions of the excess noise and dark current are ignored.

Note that in Eq. (16.42), I_s is the time average current. A time average is necessary because the encoded light is off some of the time, and during the off time, no shot noise is generated. Thus, P_s^a (which is the average light power received by the PIN diode or APD) rather than P_s (which is the peak light power or power for the "on" state) has to be used.

Inserting the physical constants and $\eta e/h\nu = 0.5$, $x = 0$, and $T = 293$ K, we can say that the separation between the quantum-limited and thermal-noise-limited regions is roughly as follows:

$$\begin{aligned} P_s^a M^2 &> \frac{0.1}{R_L} \quad \text{quantum limited} \\ P_s^a M^2 &< \frac{0.1}{R_L} \quad \text{thermal noise limited} \end{aligned} \tag{16.55}$$

Thus, the regions are determined by the relative values of P_s^a and $1/R_L$. Generally, analog systems are more often quantum limited and digital systems are usually thermal noise limited. The reason for this is that analog systems need a large P_a^s to meet the requirement of a large S/N, whereas digital systems do not need a large S/N or P_a^s, but rather have small R_L for a higher bit rate. In Fig. 16.22, the borderline given by Eq. (16.55) is plotted with average received power P_s^a as a function of load resistance R_L. The solid lines represent a PIN diode with $M = 1$, and the dashed line represents an APD with $M = 10$, $x = 0$ and with $M = 100$, $x = 0$. The region of lower R_L with lower P_s^a is the thermal-noise-limited S/N and that of higher R_L with higher P_s^a is the quantum-limited S/N. The quantum-limited region expands with increasing M.

The choice of the load resistance R_L is made according to the cutoff frequency f_c of the detection system. When the system needs a wide frequency bandwidth B of operation, the R_LC time constant given by Eq. (16.15) has to be reduced so that the cutoff frequency $f_c = 1/2\pi R_LC$ is larger than B. But with a smaller R_L, the gain of the system suffers, as shown by Eq. (16.13). The value of $f_c = 1/2\pi R_LC$ for $C = 1.6$ pF is included in the bottom horizontal axis in Fig. 16.22.

16.5.9 Relationship Between Bit Error Rate (BER) and Signal to Noise Ratio

In the field of digital signal communication, the quality of the digital signal is evaluated by the number of error bits over the total number of bits transmitted. This relationship

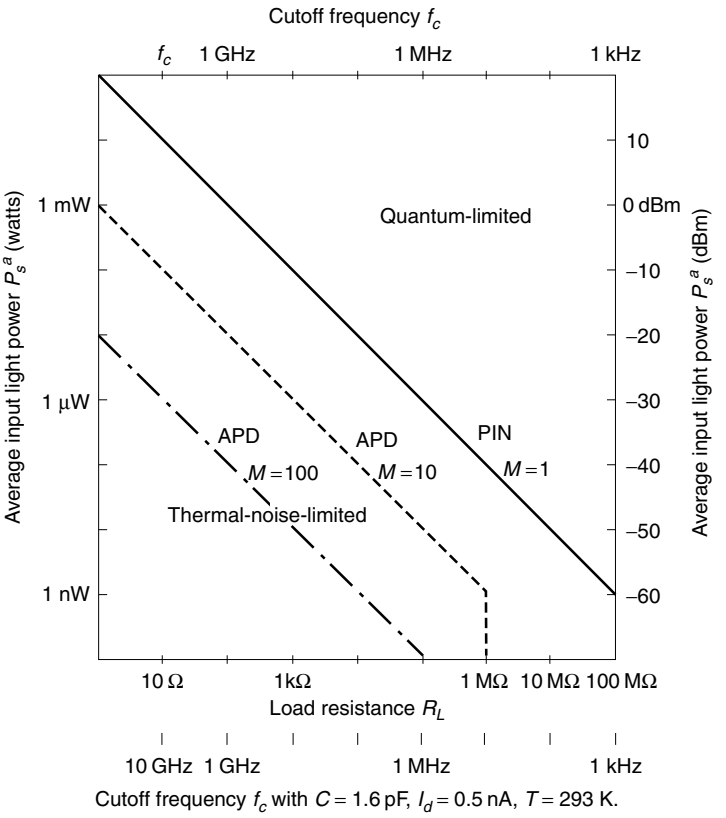


Figure 16.22 Borderline between quantum-limited and thermal-noise-limited cases of PIN and APD photodiodes.

is called the bit error rate (BER). The noisier the received signal is, the more the detection is subject to error, increasing the BER of the system. The digital system uses BER while the analog system uses S/N. The relationship that exists between BER and S/N will now be presented [28,29].

A pulse that contains noise in the "1" state is shown in Fig. 16.23. (Noise in the "0" state is not drawn for clarity.) If the probability density function $p(v)$ that the voltage is between v and $v + \mathrm{d}v$ is Gaussian, then

$$p(v) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_1} e^{-1/2((v-s)/\sigma_1)^2} \tag{16.56}$$

where s is the voltage representing "1," for example, 5 V for TTL (transistor–transistor logic).

The probability P_{10} of mistaking "1" as "0" is represented by the area of the shaded portion in Fig. 16.23 and is equal to

$$P_{10} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_1} \int_{-\infty}^{V_{\text{th}}} e^{-1/2((v-s)/\sigma_1)^2} dv \tag{16.57}$$

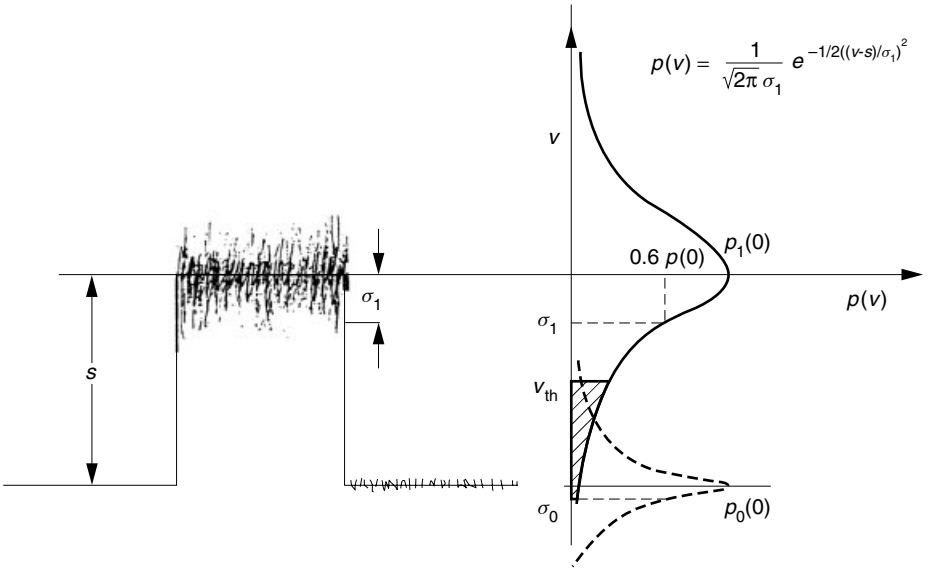


Figure 16.23 Pulse with Gaussian noise in ''1.'' The noise in ''0'' is suppressed for clarity.

Setting

$$\lambda = \frac{s - v}{\sqrt{2}\sigma_1} \tag{16.58}$$

Eq. (16.57) is rewritten as

$$P_{10} = \frac{1}{\sqrt{\pi}} \int_{(s-v_{\text{th}})/\sqrt{2}\sigma_1}^{\infty} e^{-\lambda^2}\, d\lambda \tag{16.59}$$

Similarly, the probability P_{01} of mistaking "0" as "1" is represented by

$$P_{01} = \frac{1}{\sqrt{\pi}} \int_{v_{\text{th}}/\sqrt{2}\sigma_0}^{\infty} e^{-\lambda^2}\, d\lambda \tag{16.60}$$

where σ_0 is the rms of the noise superimposed on "0." If one-half of the signal is "1" and the other half is "0," then

$$\text{BER} = \tfrac{1}{2}P_{10} + \tfrac{1}{2}P_{01} \tag{16.61}$$

The skew rate is defined as the number of errors caused by reading "1" as "0" divided by the total number of errors. By changing the threshold value v_{th}, the skew rate changes. When v_{th} is set too high, the skew rate becomes larger than 0.5. A skew rate of 0.5 approximately provides the condition for the minimum error rate. For the condition that provides exactly the minimum BER, see Problem 16.8. If the skew rate is set to 0.5, then

$$P_{10} = P_{01} \tag{16.62}$$

It follows that

$$\frac{s - v_{\text{th}}}{\sigma_1} = \frac{v_{\text{th}}}{\sigma_0} \tag{16.63}$$

and

$$v_{\text{th}} = \frac{\sigma_0}{\sigma_1 + \sigma_0} s \tag{16.64}$$

and from Eqs. (16.60)–(16.64)

$$\text{BER} = \frac{1}{\sqrt{\pi}} \int_{s/\sqrt{2}(\sigma_1+\sigma_0)}^{\infty} e^{-\lambda^2}\, d\lambda \tag{16.65}$$

Equation (16.65) can be rewritten using the definition of the complementary error function

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-\lambda^2}\, d\lambda \tag{16.66}$$

Thus,

$$\text{BER} = \tfrac{1}{2}\text{erfc}(Q) \tag{16.67}$$

where

$$Q = \frac{s}{\sqrt{2}(\sigma_0 + \sigma_1)} \tag{16.68}$$

There is no closed form of integration of Eq. (16.66), but the calculated curve is available in Fig. 16.24.

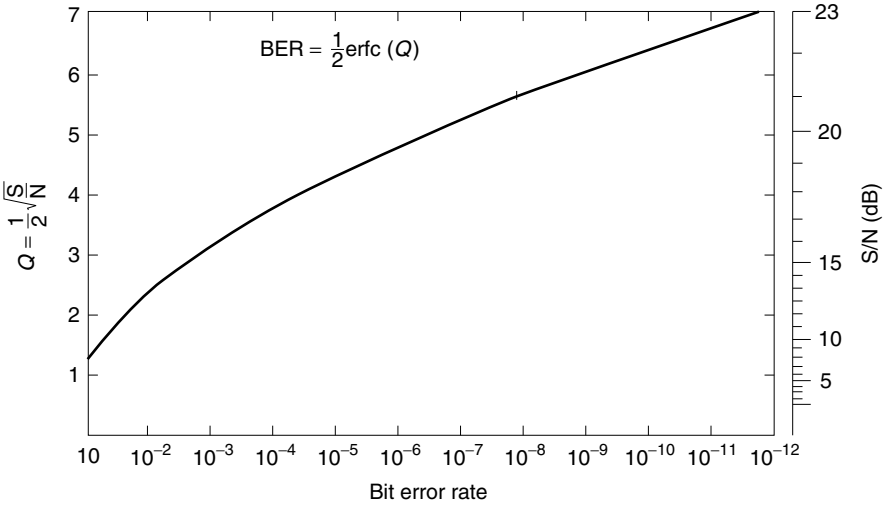


Figure 16.24 Bit error rate versus *Q* and S/N in dB.

An approximate expression for erfc(x), which is good for the entire range of x, is [30]

$$\tfrac{1}{2}\ \mathrm{erfc}(x) = \frac{1}{(1-a)x + a\sqrt{x^2+b}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{16.69}$$

where

$$a = 1/\pi$$
$$b = 2\pi$$
$$0 < x < \infty$$

For $x^2 \gg 2\pi$, Eq. (16.69) reduces to

$$\tfrac{1}{2}\ \mathrm{erfc}(x) = \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}$$

It should be noted that σ_0 and σ_1 are not necessarily the same. For instance, in the case of direct detection, shot noise due to the signal current is present for "1" but absent during "0."

With some approximation, the value of Q will be represented in terms of the S/N. Let us assume that

$$\sigma_0 = \sigma_1 = \sigma \tag{16.70}$$

and that the amplitude of the ASK signal for "1" is s volt. For a 50% duty cycle, the signal power S is

$$S = \tfrac{1}{2}s^2/R_L \tag{16.71}$$

σ^2 represents the average value of $\overline{(v-s)^2}$ and the noise power is

$$N = \sigma^2/R_L \tag{16.72}$$

From Eqs. (16.68) to (16.72) this means that

$$Q = \tfrac{1}{2}\sqrt{\frac{\mathrm{S}}{\mathrm{N}}} \tag{16.73}$$

For instance, from Fig. 16.24, in order to achieve BER $= 10^{-9}$, $Q = 6$ is required. The corresponding S/N from Eq. (16.73) is S/N $= 144$ or 21.6 dB.

The required BER depends on the kind of information being sent. Roughly speaking, BER $= 10^{-2}$ is sufficient for a voice channel. BER $= 10^{-4}$ is required for data transmission and BER $= 10^{-9}$ is necessary for computer communication.

Even a relatively noisy transmission system of S/N $= 144$ (S/N $= 5000$ is a typical requirement of an analog signal transmission) can transmit the signal at a BER of 10^{-9} if the signal is first digitized and transmitted. Thus, by using digital modulation, even a poor performance transmission system can be used. The drawback of the digital system is the demand for wider frequency bandwidth, which is, of course, the allure of the

large information-carrying capacity of the optical fiber. Another advantage of digital systems is that the effects of not only the noise but also the pulse shape distortion are not as critical as for analog systems.

As a final note, it should be recognized that the expression for the BER depends on the discriminator mechanism that differentiates between "0" and "1" signals. The ASK discriminator uses the threshold voltage of the receiver output, while the FSK and PSK discriminators use the outputs between the two frequencies and the outputs between the two states of phases of the light, respectively. Rigorously speaking, the expressions are slightly different.

Example 16.3 A PIN diode whose NEP is 3.9×10^{-13} W/$\sqrt{\text{Hz}}$ and whose sensitivity is thermal noise limited is used as a receiver. Find the required incident light power to the PIN diode to maintain the BER $= 10^{-9}$. The frequency bandwidth of the PIN diode is 300 MHz.

Solution

Method I. From the BER curve in Fig. 16.24, the required S/N for BER $= 10^{-9}$ is S/N $= 144$.

For unmodulated light, the light power $P_{s\,\min}$ needed for S/N $= 1$ is

$$\begin{aligned} P_{s\,\min} &= NEP\sqrt{B} \\ &= (3.9 \times 10^{-13})(\sqrt{3 \times 10^{8}}) \\ &= 6.76 \times 10^{-9} \text{ W} \\ &= -51.7 \text{ dBm} \end{aligned}$$

For the ASK modulated light, the corresponding necessary light power is $\sqrt{2}$ times larger:

$$\begin{aligned} P_{s\,\min}^{\text{ASK}} = \sqrt{2}P_{s\,\min} &= 9.56 \times 10^{-9} \text{ W} \\ &= -50.2 \text{ dBm} \end{aligned}$$

The output electrical signal power S is proportional to the square of the optical light power P_s^{ASK}. In order to increase the electrical power S by 144 times, the light power P_s^{ASK} has to be increased by $\sqrt{144}$.

$$\begin{aligned} P_s^{\text{ASK}} &= \sqrt{144} \times 9.56 \times 10^{-9} \\ &= 115 \text{ nW} \\ &= -39.4 \text{ dBm} \end{aligned}$$

Method II. From the definition of NEP

$$\left(\frac{\mathrm{S}}{\mathrm{N}}\right)_1 = \frac{(RP_{s\,\min})^2 R_L}{N_1} = 1$$

$$\left(\frac{\mathrm{S}}{\mathrm{N}}\right)_2 = \frac{\frac{1}{2}(RP_s^{\mathrm{ASK}})^2 R_L}{N_2} = 144$$

where N_1 is the noise power at the minimum detectable power and N_2 is the noise power at 144 times the minimum detectable power.

$$P_{s\,\min} = \mathrm{NEP}\sqrt{B}$$

If the diode is thermal noise limited, then

$$N_1 = N_2 = N$$

Solving these simultaneous equations gives

$$\begin{aligned} P_s^{ASK} &= \sqrt{2}\sqrt{144}\; NEP\sqrt{B} \\ &= \sqrt{2}\sqrt{144}(3.9 \times 10^{-13})\sqrt{3 \times 10^8} \\ &= 115\mathrm{nW} = -39.4\ \mathrm{dBm} \end{aligned}$$

Method III. The electrical noise current i_N for a given NEP and responsivity R is

$$i_N = R\ \mathrm{NEP}\sqrt{B}$$

$$N = \langle i_N^2 \rangle R_L = R^2(\mathrm{NEP})^2 B R_L$$

The required electrical signal power to establish (S/N) = 144 is

$$S = \left(\frac{\mathrm{S}}{\mathrm{N}}\right) N = 144\ R^2(\mathrm{NEP})^2 B R_L$$

whereas the light power P_s^{ASK} needed to generate this electrical signal is obtained by setting

$$S = \tfrac{1}{2}(RP_s^{\mathrm{ASK}})^2 R_L$$

Hence,

$$P_s^{\mathrm{ASK}} = \sqrt{2}\sqrt{144}\ \mathrm{NEP}\sqrt{B} \qquad \square$$

16.6 DESIGNING FIBER-OPTIC COMMUNICATION SYSTEMS

When designing a fiber-optic communication system, one must keep two major considerations in mind. The first consideration is power. There must be sufficient

light power to maintain the desired levels of S/N or BER. The second consideration is the response time of the system, which must be fast enough to respond to the variation of the signal. Achieving adequate response time ensures that information can be transported with the required accuracy.

16.6.1 System Loss

A fiber-optic communication system consists of three components: an optical fiber, a transmitter, and a receiver. The optical fiber is terminated by the transmitter at one end and the receiver at the other. While there is a minimum light power requirement for a receiver to interrogate information, there is also a limit to the light power that the laser diode can put out. The allowed loss of light power is calculated by the difference between the light power of the transmitter and the receiver. If decibel scales are used, the formula for the power requirement is

$$P_S = P_T - P_A \tag{16.74}$$

where P_S is the power received by the PIN or APD detector in dBm, P_T is the transmitter light power from the LD or LED in dBm, and P_A is the attenuation of the light incurred in the path from the transmitter to the receiver in dB.

Let us first deal with the attenuation P_A. The transmission loss of light in the fiber is one of several mechanisms for attenuation in the system. Fiber loss ranges from 0.2 dB/km to a few dB/km depending on the quality of the fiber and the wavelength of operation.

The next item contributing to loss is the light power lost in coupling the output light from the LD or LED into the end face of the fiber. Coupling loss is several dB, but quite often the LD and LED are pigtailed to the fiber by the manufacturers. A short piece of optical fiber prealigned and glued to the LD or LED by the manufacturer is called a pigtailed LD, and the coupling loss of such an LD is already accounted for in the light output power rating.

Splicing losses are due to fiber connections that are needed during installation for joining spools of fiber or making repairs at the time of accidental breakages. A splice is intended to be a permanent connection, whereas a connector is designed so that it can easily be disconnected and reconnected. The splice loss is less than 0.1 dB per splice.

Optical fiber connectors are needed at both transmitter and receiver ends. Connector loss is less than 1 dB each. Time degradation of the output power of the LD or LED should also be accounted for by adding a few dB to the net loss.

After taking these power losses into account, generally a margin of safety is also included. A sample calculation is shown in Table 16.4 for a 50-km fiber with a pigtailed source.

The procedures for calculating the power requirement for analog and digital modulated signals are different and they will be treated separately.

16.6.2 Power Requirement for Analog Modulation

Calculating the power restriction normally starts from the receiver side. First, the light power P_s, which is required for a given frequency bandwidth B to maintain a

Table 16.4 Loss budget for a 50-km span

Item	Unit Loss	Quantity	Total
Fiber loss	0.3 dB/km	50	15.0
Connector loss	1 dB	2	2.0
Splicing loss	0.1 dB	24	2.4
Degradation	3 dB	1	3.0
Margin of safety	5 dB	1	5.0
			$P_A = 27.4$ dB

specified S/N, will be calculated. An increase in the frequency bandwidth B significantly increases noise, which raises the required light power. The influence of B on noise is twofold. First, shot as well as thermal noise contributions are proportional to B. Second, in order to increase B, R_L in the detector circuit must be decreased. Reducing R_L, however, further raises the thermal noise $(4kT/R_L)B$ contribution.

The case of direct detection is considered. If both shot noise as well as thermal noise are taken into consideration, the S/N is a quadratic function of P_s as shown by Eq. (16.42). Solving for P_s for a given S/N is somewhat involved. The following trial and error method may be simpler.

Let us first make the assumption that the S/N is thermal noise limited in order to calculate the required power P_s for a given S/N, and then reexamine the validity of this assumption from the obtained P_s.

Since the expression for the signal power for ASK modulation has already been derived, for a new challenge, let us now find the signal power S for a sinusoidally modulated analog signal where the light power is

$$P = P_s^a(1 + m\cos\omega_{st}) \tag{16.75}$$

A modulation index $m = 1$ is assumed. The signal power $S' = \langle i_s^2 \rangle R_L$ from the APD is, from Eq. (16.34),

$$S = \tfrac{1}{2}\left(\frac{\eta e}{h\nu} P_s^a M\right)^2 R_L \tag{16.76}$$

This turns out to be identical to that of ASK modulation. However, note that P_s^a with the superscript a in Eq. (16.75) is the average light power, while P_s for ASK is the peak power or power for the "on" state.

In the thermal-noise-limited case, Eq. (16.49) for amplitude modulation reduces to

$$\left(\frac{\mathrm{S}}{\mathrm{N}}\right) = \frac{\tfrac{1}{2}\left(\dfrac{\eta e}{h\nu} M\right)^2 P_s^{a^2}}{\dfrac{4kTB}{R_L}} \tag{16.77}$$

The value of R_L is determined by the cutoff frequency f_c of the detector to provide the bandwidth B. From Eq. (16.15) with $f_c = B$, R_L is

$$R_L = \frac{1}{2\pi CB} \tag{16.78}$$

From Eqs. (16.77) and (16.78), the average light power P_s^a in terms of S/N and B is

$$P_s^a = 2\frac{1}{M}\frac{h\nu}{\eta e}\sqrt{4kT\pi C\left(\frac{\mathrm{S}}{\mathrm{N}}\right)B} \tag{16.79}$$

Inserting typical values $\eta e/h\nu = 0.5$ A/W, $T = 293$ K, and $C = 1.6$ pF, we get

$$P_s^a = 1.14 \times 10^{-15}\frac{1}{M}\sqrt{\frac{\mathrm{S}}{\mathrm{N}}B} \tag{16.80}$$

A few values calculated from Eq. (16.80) for $\mathrm{S/N} = 10^5$, which is the largest value of S/N considered, are tabulated below.

	P_s^a (dBm)	
B	PIN Diode ($M = 1$)	APD ($M = 100$)
1 MHz	−34.4	−54.4
10 MHz	−24.4	−44.4
1 GHz	−4.4	−24.4

By examining where the above values fall in relation to Fig. 16.22, we can verify that all the points for the PIN diode fall into the assumed thermal-limited region, but those for the APD do not with the given value of $\mathrm{S/N} = 10^5$. When the detector is an APD, the assumption of the thermal-noise-limited S/N does not hold for $\mathrm{S/N} = 10^5$, and the quantum-limited S/N and Eq. (16.88), which are discussed later on, have to be used. Therefore, Eq. (16.80) can be used safely only for the PIN photodiode. The final results are plotted in Fig. 16.25.

In summary, when a PIN diode is used as the detector, find P_s^a either by Fig. 16.25 or Eq. (16.80) for a given S/N and B, and then adjust the values of P_T and P_A to be within the constraint of Eq. (16.74).

It should be mentioned that the designed frequency bandwidth of the receiver should be estimated to be about 5% wider than the required bandwidth so as to allow for the contribution of the rise times of the fiber and the transmitter $(1.05)B$ rather than B is used in Eq. (16.80) as explained in Section 16.6.4.

16.6.3 Rise-Time Requirement for Analog Modulation

The frequency response of the overall fiber-optic communication system has to account for the frequency responses of the transmitter, fiber, and receiver. There is an empirical formula for the overall frequency response of the entire system. This formula is

$$t = \sqrt{t_1^2 + t_2^2 + t_3^2 + \cdots + t_i^2 + \cdots} \tag{16.81}$$

where t is the rise time of the overall system and t_i is the rise time for an individual component. This is a very handy formula but the performance of each component has to be characterized by the rise time rather than the frequency response. The factor that converts the frequency bandwidth to the rise time will be obtained next.

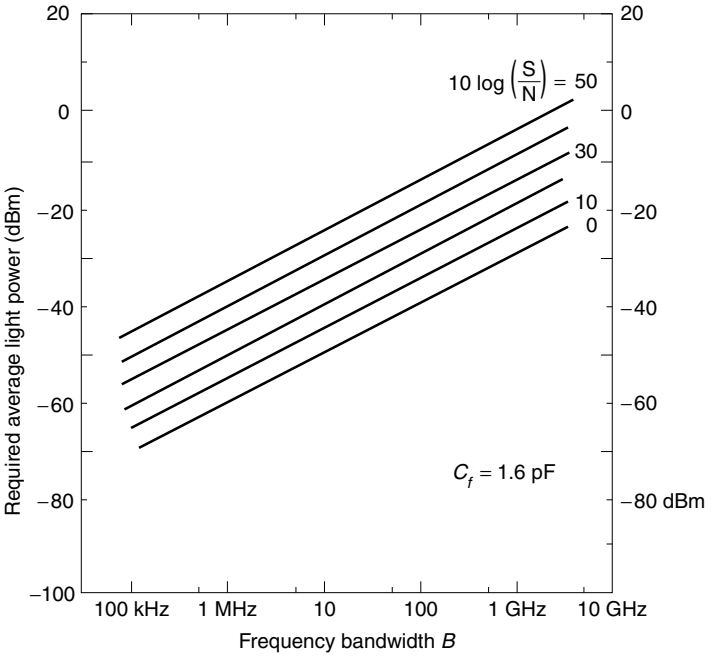


Figure 16.25 Required average light power of a PIN photodiode as a function of frequency bandwidth B with S/N as a parameter. The modulation is analog and the S/N is thermal noise limited. $C_f = 1.6$ pF.

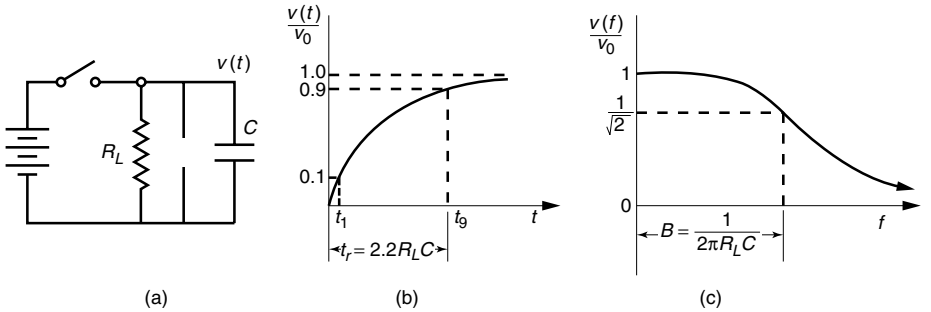


Figure 16.26 Relationship between rise time and frequency response of an RC circuit to show $Bt_r = 0.35$ by eliminating R_LC from the expressions. (a) RC circuit. (b) Rise time. (c) Frequency response.

When a step function generator is applied to an RC circuit such as shown in Fig. 16.26a, the voltage $V(t)$ starts to rise with time as follows:

$$V(t) = V_0(1 - e^{-t/RC}) \tag{16.82}$$

The curve of Eq. (16.82) is drawn in Fig. 16.26b. The rise time, which is defined as the time taken to rise from 0.1 V_0 to 0.9 V_0, is calculated from Eq. (16.82) as

$$\begin{aligned} 0.9 &= 1 - e^{-t_9/RC} \\ 0.1 &= 1 - e^{-t_1/RC} \end{aligned} \tag{16.83}$$

and $t_r = t_9 - t_1$ is

$$t_r = 2.2\ CR \tag{16.84}$$

With the same circuit, this time, a swept frequency generator is applied. The response would be something like the one shown in Fig. 16.26c. The 3-dB cutoff frequency bandwidth B is given by Eq. (16.78). Thus, the frequency bandwidth and rise time are related from Eqs. (16.78) and (16.84) as

$$Bt_r = 0.35 \tag{16.85}$$

The transmitters and detectors are all characterized by the modulation frequency bandwidth, but their rise times can be obtained from Eq. (16.85) if we assume that their frequency characteristic behaves like an RC circuit.

As for the fiber, the frequency response is characterized by the spread of the group delay. In a multimode fiber, mode dispersion is dominant and the spread of the group delay of the step-index fiber is given by Eq. (11.7), which is repeated here:

$$\Delta\tau = \frac{1}{2cn_1}(\text{NA})^2 \tag{11.7}$$

and that of the graded-index fiber is given by Eq. (11.242), which is repeated here:

$$\Delta\tau = \frac{1}{8}\frac{(\text{NA})^4}{cn_c^3} \tag{11.242}$$

In a single-mode fiber, the dispersion constant is given by the graph in Fig. 15.14. We can assume that the group delay spread is essentially equal to the rise time of the fiber,

$$t_r = \Delta\tau \tag{16.86}$$

where $\Delta\tau$ is the group delay.

In short, the rise time is calculated by converting the frequency response or group delay spread of the system into the rise time and summing their squares to obtain the rise time of the overall system.

16.6.4 Example of an Analog System Design

As a specific example, let us say that a 1-GHz bandwidth analog signal is to be transmitted over a 50-km distance with a signal to noise ratio greater than 1000, as shown in Fig. 16.27.

The transmitter is a double-channel planar buried heterostructure (DC-PBH) laser diode whose modulation frequency characteristics are shown in Fig. 14.12. The parameters of the other elements are as follows. The fiber transmission loss is 0.3 dB/km. The fiber is sold in spools, each wound with 2 km of fiber. The spools are spliced together to make the required system length of fiber. A PIN photodiode with responsivity $R = 0.5$ is used in the detector circuit. The wavelength of operation is $\lambda = 1.55\ \mu\text{m}$.

First, the required average light power P_s^a to the detector is determined. The designed frequency bandwidth of the receiver is estimated to be 5% wider than required to allow

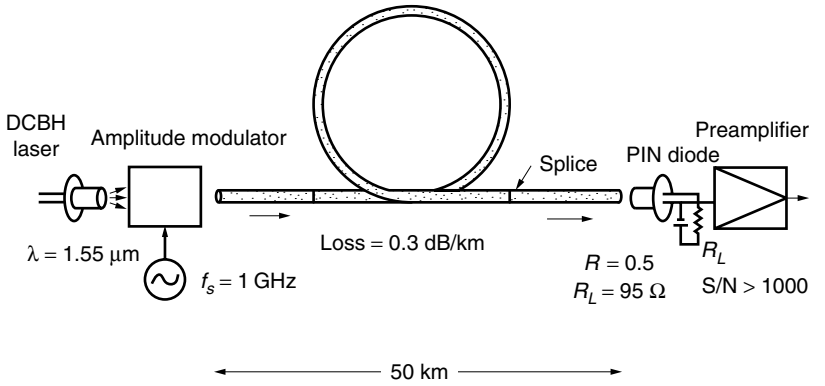


Figure 16.27 Another example of an analog system design.

later for the fiber and transmitter rise times. If 5% is insufficient, this margin can be increased and the design procedure would then be reiterated. The S/N is first assumed to be thermal noise limited. From Fig. 16.25 (or directly from Eq. (16.80)), the required P_s^a for S/N = 1000 and $B = 1.05$ GHz is $P_s^a = -14$ dBm or in terms of the peak power $P_s = -11$ dBm.

The load impedance for 1.05 GHz is, from Eq. (16.15),

$$R_L = \frac{1}{2\pi BC} = \frac{1}{2\pi(1.05 \times 10^9)(1.6 \times 10^{-12})} = 95\ \Omega$$

Next, the total attenuation loss P_A must be found. The relevant numbers are found in Table 16.4 and the result is

$$P_A = 27.4 \text{ dB}$$

Thus, the required transmitter power is, from Eq. (16.74),

$$P_T = P_S + P_A = -11 + 27.4 = 16.4 \text{ dBm}$$

An LD whose peak output is 16.4 dBm or 43.7 mW is needed. Such a power level is somewhat demanding. Later, an alternative detector circuit will be considered to see if the demand on transmitter power can be reduced.

Next, the rise time is calculated. The rise time t_S of the system is

$$t_S^2 = t_D^2 + t_F^2 + t_T^2 \tag{16.87}$$

where t_S is the rise time for the whole system, t_D is the rise time for the detector, t_F is the rise time for the fiber, and t_T is the rise time for the transmitter. The rise time t_S of Eq. (16.85) of the desired 1-GHz system is

$$t_S = \frac{0.35}{B} = \frac{0.35}{10^9} = 350 \text{ ps}$$

The rise time t_D of the detector circuit whose rise time is 5% shorter than desired to accommodate for t_F and t_T is

$$t_D = \frac{0.35}{B} = \frac{0.35}{1.05 \times 10^9} = 333 \text{ ps}$$

Next, the rise time t_F of a single-mode fiber will be calculated. The spread of the group delay $\Delta\tau$ is calculated from the wavelength bandwidth of the signal. The wavelength bandwidth $\Delta\lambda$ due to the 1-GHz modulation is

$$\Delta\lambda = 2\left(\frac{c}{f_c} - \frac{c}{f_c + 10^9}\right) = 0.016 \text{ nm}$$

where f_c is the unmodulated 1.55-μm wavelength light. From Fig. 15.14 the dispersion constant D at $\lambda = 1.55$ μm is

$$D = 17 \text{ ps/(km} \cdot \text{nm)}$$

and

$$\Delta\tau = (17)(50)(0.016) = 13.6 \text{ ps}$$

Therefore, from Eq. (16.86), $t_F = 13.6$ ps.

The modulation frequency bandwidth of the laser is shown in Fig. 14.12 as 8 GHz for $i = 30$ mA, and hence,

$$t_T = \frac{0.35}{8 \times 10^9} = 43.8 \text{ ps}$$

The overall rise time therefore is

$$\sqrt{t_D^2 + t_F^2 + t_T^2} = \sqrt{333^2 + 13.6^2 + 43.8^2}$$
$$= 336 \text{ ps}$$

which is shorter than the required 350-ps system rise time and satisfies the budgeted rise-time requirement.

Lastly, let us see if the overstressed requirement of the transmitter light power can be relieved by installing the transimpedance circuit shown in Fig. 16.16 in the detector circuit. Earlier, Eq. (16.78) was used to obtain Eq. (16.80) relating B and R_L, but in this case, R in Fig. 16.16 has to be used and the results in Fig. 16.25 can no longer be applied. The frequency band B of the transimpedance circuit is given by

$$B = \frac{1}{2\pi(R_f/G)C} \tag{16.24}$$

and the two conditions of Eq. (16.22) for this to be true are $G \gg 1$ and $RG \gg R_f$. The impedance R of the PIN diode, being of the order of megohms, easily satisfies the second condition. In addition, the thermal noise contribution $4kTB/R$ becomes

negligibly small compared to the shot noise, and the quantum-limited approximation has to be used. The quantum-limited S/N for analog modulation using a PIN diode is

$$\frac{\mathrm{S}}{\mathrm{N}} = \frac{\frac{1}{2}\left(\frac{\eta e}{h\nu} P_s^a\right)^2}{2eB\frac{\eta e}{h\nu} P_s^a} \tag{16.88}$$

and

$$P_s^a = 4e\frac{h\nu}{\eta e}\left(\frac{\mathrm{S}}{\mathrm{N}}\right) B \tag{16.89}$$

Inserting parameters and typical values of $\eta e/h\nu = 0.5$, Eq. (16.89) becomes

$$P_s^a = 1.28 \times 10^{-18}\left(\frac{\mathrm{S}}{\mathrm{N}}\right) B \tag{16.90}$$

Using Eq. (16.90), $P_s^a = -28.9$ dBm or $P_s = -25.9$ dBm is obtained and P_T becomes

$$P_T = -25.9 + 27.4 = 1.5 \text{ dBm}$$

or 1.4-mW peak power. A laser diode with such an output power is readily available.

Example 16.4 A 6-MHz wide analog TV signal is to be sent over a 10-km optical fiber link between a TV station in the city and a TV tower on a hill for microwave broadcasting as shown in Fig. 16.28. Spools wound with 2 km of multimode graded-index fiber with NA = 0.2 and $n_c = 1.55$ are intended to be used. The fiber loss is 1 dB/km. The wavelength is 0.82 μm. A PIN diode with a terminal capacitance of 7 pF and zero dark current is used as the detector and a laser diode with 10-ns rise time is used as the transmitter. The laser diode is not pigtailed. The required signal to noise ratio is 40 dB.

(a) What is the required transmitter peak power of the laser diode?
(b) Is it possible to use a step-index multimode fiber?

Solution A 5% extra bandwidth is allotted for the rise-time budget and $B = 6.3$ MHz. If the system is assumed thermal noise limited, the average light power P_s^a needed for the detector is -32 dBm $= 0.63$ μW from Fig. 16.25, which verifies the thermal-noise-limited S/N from Fig. 16.22. However, according to the specification, this PIN diode has a terminal capacitance of 7 pF while a capacitance of 1.6 pF was assumed when Figs. 16.22 and 16.25 were prepared. Note from Eq. (16.79) that P_s^a is proportional to $\sqrt{C}$, and hence the required power P_s^a is

$$P_s^a = (0.63)\sqrt{\frac{7}{1.6}} = 1.3 \text{ μW}$$

Point $P_s^a = 1.3$ μW and $R_L = (2\pi BC)^{-1} = 3609\ \Omega$ is again located in the thermal-noise-limited region in Fig. 16.22. Alternatively, from Eq. (16.35), the shot noise power is $N_{\text{shot}} = 4.73 \times 10^{-15}$ W. From Eq. (16.36), the thermal noise power is

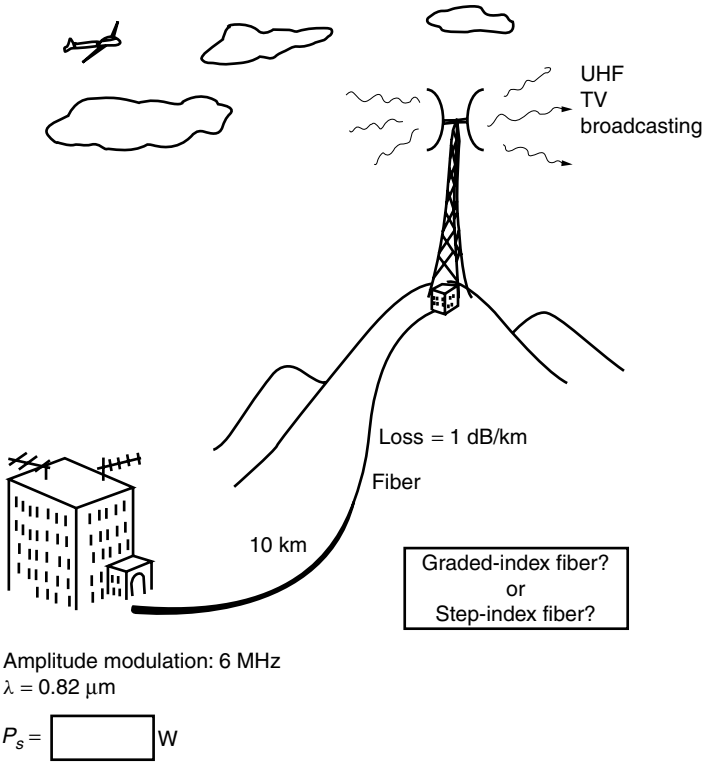


Figure 16.28 Another example of an analog system design.

$N_{\text{th}} = 1 \times 10^{-13}$ W. Thus, the assumption of the thermal-noise-limited system is verified for $C = 7$ pF. The peak power P_s is $2P_s^a$ and

$$P_s = 2.6\ \mu\text{W} = -25.9\ \text{dBm}$$

The power attenuation list is:

Coupling loss from the source to fiber	5 dB
Fiber loss (1) (10)	10 dB
Connector loss	2 dB
Splice loss (0.1) (5)	0.5 dB
Degradation	3 dB
Margin of safety	5 dB
	25.5 dB

Hence, the required transmitter power P_T is

$$\begin{aligned} P_T &= P_S + P_A = -25.9 + 25.5 = -0.4\ \text{dBm} \\ &= 0.91\ \text{mW} \end{aligned}$$

(b) The system rise time t_S is

$$t_S = \frac{0.35}{B} = \frac{0.35}{6 \times 10^6} = 58 \text{ ns}$$

The rise time for the detector is

$$t_D = \frac{0.35}{(1.05)B} = \frac{0.35}{6.3 \times 10^6} = 56 \text{ ns}$$

The spread of the arrival time of the pulse in the graded-index fiber is calculated from Eq. (11.242):

$$\Delta\tau = \frac{(\text{NA})^4}{8cn_c^3}L = \frac{(0.2)^4}{8(3 \times 10^5)(1.55^3)}(10) = 1.8 \text{ ns}$$

Thus, the total budget is

$$\sqrt{56^2 + 1.8^2 + 10^2} = 56.9 \text{ ns}$$

Next, the possibility of using a step index multimode fiber is considered. The group delay spread is given by Eq. (11.7) as

$$\Delta\tau = \frac{(NA)^2}{2cn_1}L = \frac{(0.2)^2}{2(3 \times 10^5)1.55}(10) = 430 \text{ ns}$$

The fiber alone exceeds the system rise time of 58 ns and thus this fiber cannot be used. In other words, from Eq. (16.85), the bandwidth of the fiber is

$$B = \frac{0.35}{t_F} = \frac{0.35}{4.3 \times 10^{-7}} = 814 \text{ kHz}$$

which is too narrow for the 6-MHz system. □

16.6.5 Required Frequency Bandwidth for Amplifying Digital Signals

The design procedure for the digital system is more or less the same as that of the analog system. The differences are the units used. In digital systems, BER is used instead of S/N, and bit rate B_t is used instead of bandwidth B. Since the conversion between BER and S/N has already been done (see Fig. 16.24), the conversion between B and B_t will be explained before going on to the next design procedure.

The baseband electrical pulse, such as the one shown in Fig. 16.29a, is represented by the rectangle function from Section 1.4.1 as

$$f(t) = \Pi\left(\frac{t}{T}\right) \tag{16.91}$$

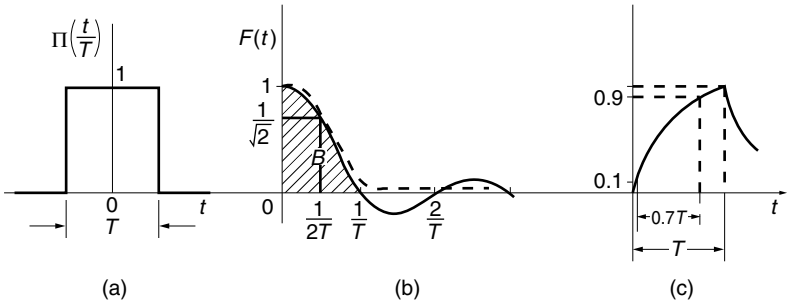


Figure 16.29 Relationship between NRZ pulse, its Fourier transform, and the output of an amplifier with $B = 1/2T$. (a) NRZ pulse in the time domain. (b) NRZ pulse in the frequency domain. (c) Output from an amplifier with $B = 1/2T$.

where T is the period of the NRZ pulse. The frequency spectrum $F(f)$ of this pulse is obtained from its Fourier transform as

$$F(f) = T \text{ sinc } Tf \tag{16.92}$$

As shown in Fig. 16.29b, the first zero crossing point appears at $f = 1/T$ and the 3-dB down (one-half of the power and $1/\sqrt{2}$ of the voltage) point appears at approximately the midpoint between the origin and the first zero crossing. Suppose this pulse is applied to an amplifier whose 3-dB bandwidth B is set at $1/2T$. The frequency characteristic of the amplifier is shown by the dotted lines in Fig. 16.29b. As seen from the figure, the spectral components beyond the first zero crossing cannot be amplified by this amplifier, even though the major frequency components are retained in the passband.

If it is approximated that this amount of retained spectrum will do, then the required frequency bandwidth of the amplifier is

$$B = \frac{1}{2T} \tag{16.93}$$

The pulse of the NRZ code in Table 16.2 takes up one entire period of coding and the bit rate B_t and period are related by

$$B_t = \frac{1}{T} \tag{16.94}$$

Thus, the required frequency bandwidth to amplify the NRZ pulses at the bit rate B_t is, from Eqs. (16.93) and (16.94),

$$B = \tfrac{1}{2}B_t \tag{16.95}$$

The conclusion is that the required 3-dB frequency bandwidth of the amplifier is one-half of the bit rate of the NRZ pulses.

Let us see how the output of the amplifier looks if the bandwidth of the amplifier is set according to Eq. (16.93). The rise time t_r when a step function is applied to an

amplifier with a 3-dB cutoff at B is expressed by Eq. (16.85). Inserting Eq. (16.93) into (16.85) gives

$$t_r = 0.7\ T \tag{16.96}$$

The rise time of the output is 70% of the period as shown in Fig. 16.29c. For most designs, an amplified signal such as this is satisfactory. The rise time required for the RZ code is $0.35T$. If an amplifier is used with $B = N/2T$, which is N times wider than that specified by Eq. (16.93), then the rise time of the output becomes N times faster and the output more closely resembles the input. However, for the purpose of digital communication, a faithful reconstruction of the pulse is a luxury because all that is really needed in binary coding is whether or not the pulse is present, and the bandwidth of the amplifier specified by Eq. (16.95) is sufficient for the NRZ pulses.

In the case of the RZ coding, the pulsewidth becomes one-half of the pulse of NRZ coding, and the rise time has to be twice as fast to retain the same quality of the output pulse. For RZ coding, B has to be twice as wide as set by Eq. (16.93) and

$$B = \frac{1}{T} \tag{16.97}$$

Comparing Eq. (16.97) with (16.94),

$$B = B_t \tag{16.98}$$

Thus, for the RZ coding, the frequency bandwidth of the amplifier must be identical with that of the bit rate of the digital signal. This concludes the explanation of the conversion between B and B_t. Now let us return to the design procedure for digital modulation.

16.6.6 Digital System Design

When explaining the procedure of calculating power for digital systems, some of the calculated results for analog modulation can be modified and utilized. The required BER can be converted into the required S/N using the graph in Fig. 16.24. The bit rate B_t can be converted into the required frequency bandwidth B of the detector circuit using Eq. (16.95) for the NRZ code and Eq. (16.98) for the RZ code. Hence, the methods of calculating power are similar for both analog and digital modulations.

The case of the NRZ code will provide an example. The peak light power P_s, which is the light power for the "1" bit, is used in the expressions.

First, the peak light power P_s needed for a given BER will be calculated assuming a thermal-noise-limited S/N. If half the bits are "0"s and the other half are "1"s, then the signal to noise ratio is

$$\frac{\mathrm{S}}{\mathrm{N}} = \frac{\frac{1}{2}\left(\frac{\eta e}{h\nu} M P_s\right)^2}{\frac{4kTB}{R_L}} \tag{16.99}$$

The frequency bandwidth requirement of the digital case is only one-half of the bit rate as given by Eq. (16.95). The value of the load impedance, therefore, is double

that of the analog case; explicitly,

$$R_L = \frac{1}{\pi C B_t} \tag{16.100}$$

Inserting Eqs. (16.95) and (16.100) into Eq. (16.99) gives

$$P_s = \frac{1}{M}\frac{h\nu}{\eta e}\sqrt{4kT\pi C\left(\frac{\mathrm{S}}{\mathrm{N}}\right)B_t} \tag{16.101}$$

Inserting $\eta e/h\nu = 0.5$ A/W, $T = 293$ K, and $C = 1.6$ pF,

$$P_s = 5.70 \times 10^{-16}\frac{1}{M}\sqrt{\frac{\mathrm{S}}{\mathrm{N}}B_t} \tag{16.102}$$

The S/N for a given BER is found from the conversion curve in Fig. 16.24. The required peak powers P_s with $M = 1$ and $B_t = 1$ Gb/s for the three cases BER $= 10^{-2}$, 10^{-4}, and 10^{-9} are tabulated for comparison.

PEAK POWER P_s REQUIRED FOR VARIOUS BER

BER	S/N	$\sqrt{\mathrm{S/N}}$	P_s (dBm)
10^{-2}	21.9	4.68	−25.7
10^{-4}	54.8	7.4	−23.8
10^{-9}	144	12.0	−21.7

As seen from the table, a change in BER does not profoundly influence the value of P_s. A rule of thumb is that the value of P_s for 10 times the BER can be found by subtracting 0.5 dBm from the original value of P_s. For instance, the value of P_s for BER $= 10^{-8}$ is obtained from that for BER $= 10^{-9}$ as $-21.7 - 0.5 = -22.2$ dBm.

$P_s^a (= \frac{1}{2}P_s)$ is plotted against B_t for a PIN photodiode in Fig. 16.30 for BER $= 10^{-9}$ and $C = 1.6$ pF. As in the case of analog modulation, this curve is reexamined to see whether or not the thermal-noise-limited assumption was correct. Referring to Fig. 16.22, the curves of not only the PIN photodiode but also the APD in Fig. 16.30 do indeed satisfy the thermal-noise-limited assumption. This is simply because the digital system does not need a large value of S/N so that P_s is smaller and the lower equation in Eq. (16.55) holds. Unless such elaborate circuits as mentioned in Section 16.4.3 are used, the PIN and APD detectors are usually thermal noise limited for digital signals.

The peak value P_s has been used for the NRZ code instead of the average P_s^a. The average value, however, is much easier to measure. The vender's specification of the output power from the laser diode or LED is according to the peak power. The designer has to convert between the two values as necessary. Equation (16.102) can be expressed in terms of P_s^a using

$$P_s^a = \tfrac{1}{2}P_s \tag{16.103}$$

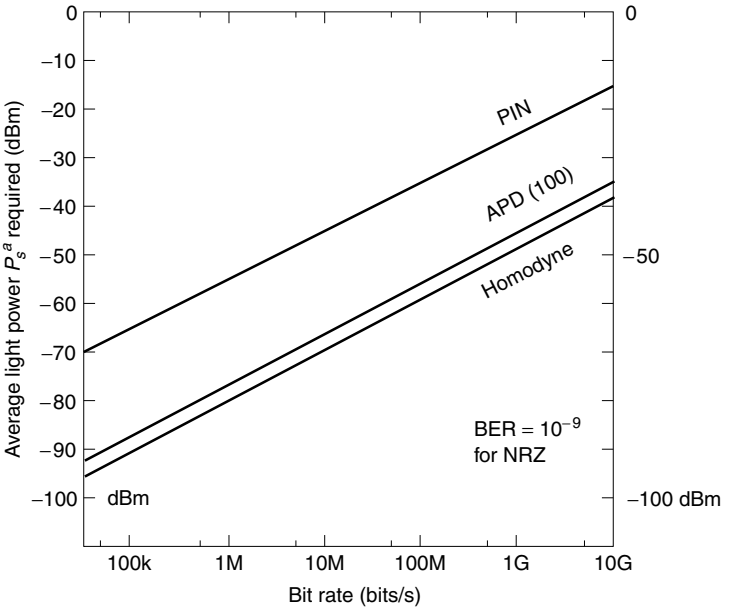


Figure 16.30 Input average light power P_s^a versus bit rate (bits/s) for NRZ code. For the value of the RZ code add 3 dB. The PIN photodiode and APD are thermal noise limited but the homodyne is quantum limited.

For the thermal-noise-limited case for the PIN photodiode, Eq. (16.102) becomes

$$P_s^a = 2.85 \times 10^{-16} \frac{1}{M} \sqrt{\frac{\mathrm{S}}{\mathrm{N}}} B_t \tag{16.104}$$

Lastly, homodyne detection is considered. With

$$2P_s^a = P_s, \quad B_{\mathrm{IF}} = 2B, \quad B = \frac{B_t}{2}, \quad M^x = 1, \text{ and } \frac{\eta e}{h\nu} = 1 \tag{16.105}$$

Eq. (16.53) for homodyne detection becomes

$$\begin{aligned} P_s^a &= \tfrac{1}{2} e \left(\frac{\mathrm{S}}{\mathrm{N}}\right) B_t \\ &= 0.8 \times 10^{-19} \left(\frac{\mathrm{S}}{\mathrm{N}}\right) B_t \end{aligned} \tag{16.106}$$

These curves for $BER = 10^{-9}$ or S/N = 144 are plotted in Fig. 16.30.

If a peak light power P_s is needed in the process of selecting a laser diode for the NRZ code, then 3 dB is added to the value obtained by the graph in Fig. 16.30. For the RZ code, $P_s^a = \frac{1}{4} P_s$ and 6 dB is added to the value obtained by the graph. This concludes the power calculations.

The rise-time calculations are quite similar to the case of analog modulation and will be explained in the following example.

16.6.7 Example of Digital System Design

Let us construct a 500-Mb/s fiber-optic link with RZ code and BER $= 10^{-9}$. The components used are a 3-mW peak power nonpigtailed laser diode with a rise time of 200 ps as the transmitter and an APD with $\eta e/h\nu = 0.5$ and $x = 0.3$, and $M = 100$ as the receiver. The wavelength used is $\lambda = 1.55$ μm.

Find the maximum distance of the transmission if a single-mode fiber with 0.3-dB/km loss is used. The fiber is available in lengths of 2 km. The problem is outlined in Fig. 16.31.

16.6.7.1 Power Requirement

The required average light power is found from Fig. 16.30 and is equal to $P_s^a = -48.7$ dBm. The point for $P_a^s = -48.7$ dBm and $B = B_t = 500$ Mb/s is indeed in the thermal-noise-limited region in Fig. 16.22. The required peak power for the RZ code is found by adding 6 dB:

$$-48.7 + 6 = -42.7 \text{ dBm}$$

The transmitter's 3-mW peak power in dBm is 4.8 dBm.

The affordable attenuation P_A between the transmitter and receiver is

$$P_A = P_T - P_S = 4.8 - (-42.7) = 47.5 \text{ dB}$$

The essential losses for the fiber link are as follows:

Coupling loss from the source to fiber	5 dB
Connector loss	2 dB
Degradation	3 dB
Margin of safety	5 dB
	15 dB

After considering the attenuation losses listed above, the rest of the loss allowed is

$$47.5 - 15 = 32.5 \text{ dB}$$

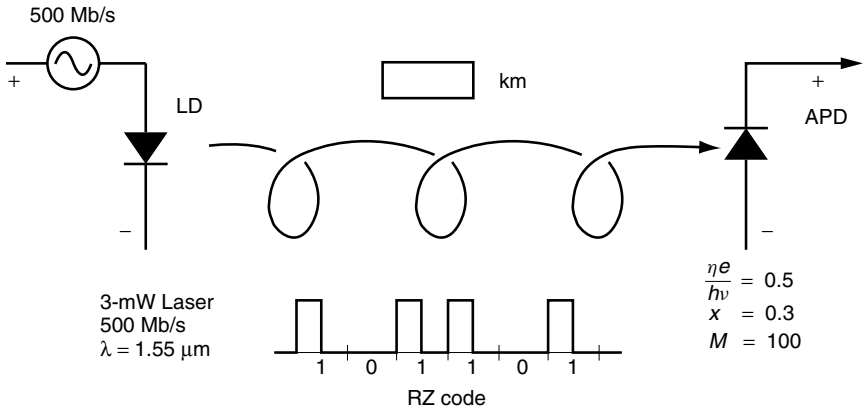


Figure 16.31 Layout of a digital system design.

The fiber has to be spliced every 2 km with 0.1-dB loss, which is equivalent to 0.05-dB/km splicing loss. The total length x of the fiber is

$$32.5 = (0.3 + 0.05)x$$

$$x = 92.9 \text{ km}$$

16.6.7.2 Rise-Time Requirement

An extra 5% is added to the bit rate to allow for the fiber and transmitter. The required rise time for the RZ code must be half as short as that of the NRZ code given by Eq. (16.96):

$$t_S = 0.35\ T = \frac{0.35}{5 \times 10^8} = 700 \text{ ps}$$

The rise time of the detector is

$$t_D = \frac{0.35}{5.25 \times 10^8} = 667 \text{ ps}$$

Next, the spread of the delay time of the single-mode fiber is calculated. For the RZ code, the necessary frequency bandwidth B from Eq. (16.98) is $B = 500$ MHz. The wavelength bandwidth $\Delta\lambda$ due to the modulation is

$$\Delta\lambda = 0.008 \text{ nm}$$

The dispersion constant D from Fig. 15.14 at $\lambda = 1.55$ μm is $D = 17$ ps/(km·nm) and

$$t_F = \Delta\tau = (17)(92.9)(0.008) = 12.6 \text{ ps}$$

Thus,

$$\sqrt{t_D^2 + t_F^2 + t_T^2} = \sqrt{667^2 + 12.6^2 + 200^2} = 696 \text{ ps}$$

which shows that the rise-time requirement is also satisfied.

Example 16.5 A fiber-optic communication system that can transmit the U.S. standard rating DS-4 PCM with BER $= 10^{-10}$ will be installed on a 100-km trunk line. Homodyne detection with the NRZ code is used. The wavelength is $\lambda = 1.55$ μm. A single-mode fiber with transmission loss of 0.4 dB/km in 2-km spools is used.

(a) What is the required transmitter peak power of the pigtailed laser diode?
(b) What is the rise-time requirement?

Solution The question is summarized in Fig. 16.32.

(a) From Table 16.3, DS-4 has a bit rate of $B_t = 274.175$ Mb/s. With 5% margin, $B_t = 288$ Mb/s. However, the curve in Fig. 16.30 was calculated for BER $= 10^{-9}$. From the rule of thumb of Section 16.6.6, 0.5 dBm should be added to this value to

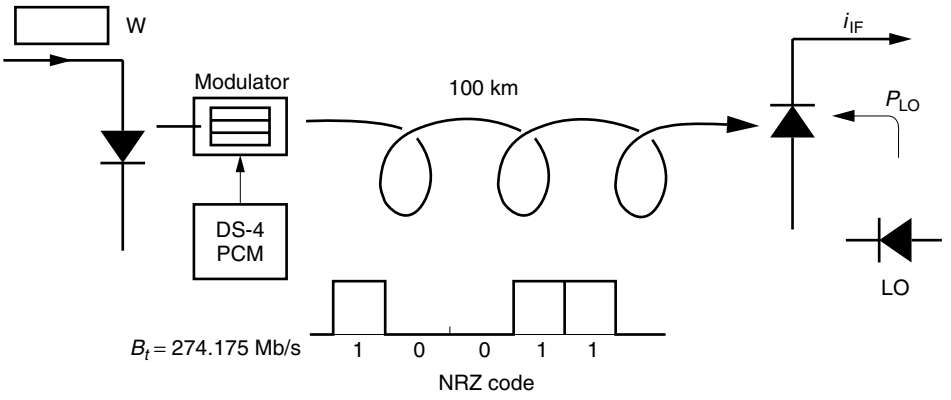


Figure 16.32 Layout of Example 16.5.

obtain $P_s^a = -55.6$ dBm, which corresponds to $-55.6 + 3 = -52.6$-dBm peak power. The tabulated attenuations of the components are as follows:

Fiber loss (0.4)(100)	40 dB
Connector loss	2 dB
Splice loss (0.1)(49)	5 dB
Degradation	3 dB
Margin of safety	5 dB
	55 dB

Finally, the required transmitter peak power is

$$P_T = P_R + P_A = -52.6 + 55 = 2.4 \text{ dBm}$$
$$= 1.7 \text{ mW}$$

(b) The system rise time is

$$t_S = \frac{0.35}{B} = \frac{0.35}{274.175 \times 10^6} = 1.28 \text{ ns}$$

The rise time for the detector is

$$t_D = \frac{0.35}{(1.05)B} == 1.22 \text{ ns}$$

In order to calculate the dispersion of the fiber, the wavelength spread $\Delta\lambda$ due to the modulation is calculated as

$$\Delta\lambda = 0.0044 \text{ nm}$$

Thus, the spread of the group delay is, from Fig. 15.14,

$$t_F = \Delta\tau = (17)(100)(0.0044) = 7.48 \text{ ps}$$

Thus, the resultant rise time is

$$\sqrt{(1.22)^2 + (7.48 \times 10^{-3})^2 + 0.18^2} = 1.23 \text{ ns}$$

and is shorter than the desired system rise time. □

PROBLEMS

16.1 The parameters of a receiver system are $\eta e/h\nu = 0.5$, $M = 1$, $C = 1.6$ pF, $I_d = 0$, and $T = 293$ K. The average input is $P_s^a = -23$ dBm and $B = 2$ kHz. No special frequency band compensation circuit is installed in the preamplifier. Is the S/N of the receiver system thermal noise limited or quantum limited?

16.2 A fiber-optic communication link will be made using a multimode step-index fiber with a core refractive index of $n_1 = 1.55$ and a numerical aperture NA $= 0.2$. The transmitter laser diode has a rise time of $t_T = 20$ ns and the detector has a rise time of $t_D = 50$ ns. The signal to be transmitted is a 4-MHz analog signal. Considering only the rise-time requirement, what is the maximum distance of transmission?

16.3 What is the required peak power to a PIN photodiode in the following system. A transimpedance circuit is used in the preamplifier and the detector can be considered quantum limited. The system operates at 100 Mb/s in RZ mode with 10^{-10} BER rate. The responsivity of the PIN diode is $R = 0.5$.

16.4 A fiber-optic communication link is to be established between two locations separated by 20 km with the following specifications.

(1) The frequency bandwidth of the signal is 300 MHz. Assume also that a 300-MHz low-pass filter has been installed.
(2) ASK modulation with NRZ code is employed.
(3) The receiver is a silicon APD, type 2382 of Hamamatsu Photonics.
(4) The transmission loss of the optical fiber is 2 dB/km.
(5) The sum of all other losses including splicing and connector loss is 2 dB. Both long-term degradation and safety margin are disregarded.
(6) The light output power from the laser diode without modulation is 20 mW, and the light wavelength is $\lambda = 830$ nm.
(7) All the characteristics are at 25°C.

(a) Find the peak current into the load.
(b) Determine the received electrical signal power.
(c) Find the shot noise power.
(d) Determine the thermal noise power.

(e) Find the signal to noise ratio.

(f) Is the system quantum limited or thermal noise limited?

(g) If $M = 1$ is chosen, is the answer to (f) still the same?

(h) What is the NEP of this APD?

(i) What is the minimum detectable light power of the APD?

(j) In this example, the span distance was 20 km. What is the theoretical limit of the span distance if the same light source, fiber, detector, and ASK modulation scheme are used? Disregard dispersion considerations.

(k) If only the power requirement is considered, what is the span distance if the same system is used for 300-Mb/s NRZ modulation with BER $= 10^{-9}$?

16.5 Answer the same questions as in Problem 16.4 but use the type S2384 APD from Fig. 16.21 with $M = 60$ and with the signal bandwidth of 100 MHz. This time, the APD is connected directly to a 50-Ω input impedance preamplifier, as in Fig. 16.14.

16.6 Find the optimum value of the multiplication factor M of an APD analytically from the viewpoint of the signal to noise ratio.

16.7 What happens to the cutoff frequency in a frequency region where the stray current $j\omega C_s V_f$ in Fig. 16.16 can no longer be neglected, where C_s is the stray capacitance in the feedback circuit?

16.8 A digital signal whose probability density functions are shown in Fig. P16.8 is applied to a discriminator. The signal level of "1" is s_1 volts and that of "0" is s_0 volts. The variance for "1" is σ_1 and that for "0" is σ_0. Find an expression that enables one to calculate the threshold voltage v_{th} that minimizes the bit error rate. Verify also that the optimum value of the threshold voltage for $\sigma_0 = \sigma_1 = \sigma$ is $v_{\text{th}} = \frac{1}{2}(s_1 + s_0)$.

16.9 Figure 16.8 shows an example of a multiplexing system designed to deliver 10 TV channels to each of 15 households. Each TV channel occupies 4-MHz bandwidth, and the TV channels are spaced by 8 GHz. The distance between the TV station and a household is 10 km on average, and a S/N of 30 dB is required.

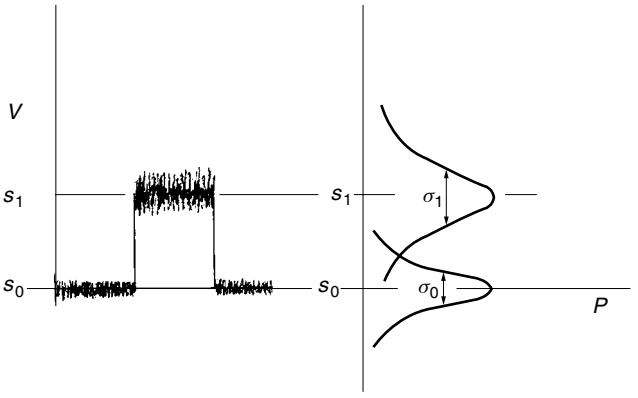


Figure P16.8 A pulse with different probability density function for ''0'' and ''1'' states.

The fiber is single-mode fiber with a loss of 1 dB/km at $\lambda = 1.55$ μm and 3 dB/km at $\lambda = 1.3$ μm and the fiber is spooled in lengths of 2 km. Laser diodes operating at $\lambda = 1.55$ μm are being proposed as the transmitters. Heterodyne detection systems with a rise time of 10^{-12} s and $\eta e/h\nu = 0.5$ are used as the receiver. The scrambler has 0.5-dB loss to each TV channel. For the rise-time budget, t_T is ignored. AM modulation is used.

(a) What is the required peak light output power?

(b) Is the rise-time requirement satisfied for the proposed transmitter?

(c) If a source wavelength of $\lambda = 1.3$ μm is used, is the rise-time requirement satisfied?

(d) What is the modification to the power requirement for $\lambda = 1.3$-μm transmitters?

REFERENCES

1. K. Iizuka (Producer), "Fiber optics labs from around the world," paneled by C. Kao, P. Cochrane, H. Kogelnik, T. Uchida, Y. Suematsu, and T. Nakahara, IEEE video, Product Number VC0022-4-POX, 1992.
2. S. Ryu, *Coherent Lightwave Communication Systems*, Artech House, Boston, 1995.
3. D. J. H. Maclean, *Optical Line Systems*, Wiley, New York, 1996.
4. J. M. Senior, *Optical Fiber Communications, Principles and Practice*, 2nd ed., Prentice Hall, New York, 1992.
5. J. Gowar, *Optical Communication Systems*, 2nd ed., Prentice Hall, New York, 1993.
6. J. C. Palais, *Fiber Optic Communications*, 4th ed., Prentice Hall, New York, 1998.
7. G. P. Agrawal, *Fiber-Optic Communication Systems*, 2nd ed., Wiley, New York, 1997.
8. L. G. Kazovsky, S. Benedetto, and A. E. Willner, *Optical Fiber Communication Systems*, Artech House, Boston, 1996.
9. J. P. Powers, *An Introduction to Fiber Optic Systems*, Aksen Associates, Irwin, Boston, 1993.
10. Special issue on "Multiple Wavelength Technologies and Networks," *J. Lightwave Technol.*, **14**(6), 932–1577 (1996).
11. A. E. Willner, "Mining the optical bandwidth for a terabit per second," *IEEE Spectrum*, 32–41 (Apr. 1997).
12. H. A. Willebrand and B. S. Ghuman, "Fiber optics without fiber," *IEEE Spectrum*, 41–45 (August 2001).
13. J. Pearson, *Basic Communication Theory*, Prentice Hall, New York, 1992.
14. R. E. Ziemer and W. H. Tranter, *Principles of Communications: Systems, Modulation, and Noise*, 3rd ed., Houghton Mifflin Co., Boston, 1990.
15. L. W. Couch II, *Digital and Analog Communication Systems*, 5th ed., Prentice Hall, New York, 1997.
16. T. Matsuda, A. Naka, and S. Saito, "Comparison between NRZ and RZ signal formats for in-line amplifier transmission in the zero-dispersion regime," *J. Lightwave Technol.*, **16**(3), 340–348 (1998).
17. S. Yamazaki, M. Shibutani, N. Shimosaka, S. Murata, T. Ono, M. Kitamura, K. Emura, and M. Shikada, "A coherent optical FDM CATV distribution system," *J. Lightwave Technol.*, **8**(3), 396–405, 1990.

18. T. Miya, "Silica-based lightwave circuits: passive and thermally active devices," *IEEE J. Selected Topics Quantum Electron.*, **6**(1), 38–45 (2000).
19. Catalogue of the Hamamatsu Corporation, Japan.
20. M. S. Park and R. A. Minasian, "Ultra-low-noise and wideband-tuned optical receiver synthesis and design," *J. Lightwave Technol.*, **12**(2), 254–259 (1994).
21. E. M. Kimber, B. L. Patel, I. Hardcastle, and A. Hadjifotiou, "High performance 10 Gb/s PIN-FET optical receiver," *Electron. Lett.* **28**(2), 120–122 (1992).
22. M. Miyashita, K. Maemura, K. Yamamoto, T. Shimura, M. Nogami, K. Motoshima, T. Kitayama, and Y. Mitsui, "An ultra broadband GaAs MESFET preamplifier IC for a 10 Gb/s optical communication system," *IEEE Trans. Microwave Tech.*, **40**(12), 2439–2444 (1992).
23. Y. Akahori, M. Ikeda, A. Kohzen, and Y. Akatsu, "11 GHz ultrawide-bandwidth monolithic photoreceiver using InGaAs PIN PD and InAlAs/InGaAs HEMTs," *Electron. Lett.*, **30**(3), 267–268 (1994).
24. A. P. Freundorfer and T. L. Nguyen, "Noise in distributed MESFET preamplifiers," *IEEE J. Solid-State Circuits*, **31**(8), 1100–1111 (1996).
25. A. P. Freundorfer and P. Lionais, "A low-noise broad-band GaAs MESFET monolithic distributed preamplifier," *IEEE Photonics Technol. Lett.*, **7**(4), 424–426 (1995).
26. K. Takahata, Y. Muramoto, H. Fukano, K. Kato, A. Kozen, S. Kimura, Y. Imai, Y. Miyamoto, O. Nakajima, and Y. Matsuoka, "Ultrafast monolithic receiver OEIC composed of multimode waveguide p-i-n photodiode and HEMT distributed amplifier," *IEEE J. Selected Topics Quantum Electron.*, **6**(1), 31–37 (2000).
27. N. Takachio, K. Iwashita, S. Hata, K. Onodera, K. Katsura, and H. Kikuchi, "A 10 Gb/s optical heterodyne detection experiment using a 23 GHz bandwidth balanced receiver," *IEEE Trans. MTT* **38**(12), 1900–1905 (1990).
28. J. Minkoff, *Signals, Noise, and Active Sensors: Radar, Sonar, Laser Radar*, Wiley, New York, 1992.
29. J. R. Barry and E. A. Lee, "Performance of coherent optical receivers," *Proc. IEEE* **78**(8), 1369–1394 (1990).
30. P. O. Börjesson and C. W. Sunberg, "Simple approximations of the error function $Q(x)$ for communications applications," *IEEE Trans. Commun.*, **27**(3), 639–643 (1979).

APPENDIX A

PIN PHOTODIODE ON AN ATOMIC SCALE

A.1 PIN PHOTODIODE

A PIN photodiode consists of p-type, intrinsic type, and n-type semiconductors. The "I"of PIN stands for intrinsic, and we will begin with an explanation of the intrinsic-type semiconductor. The simplest intrinsic-type semiconductor is made out of highly purified group IV elements such as silicon (Si) or germanium (Ge), which have four valence electrons (four electrons in the outermost orbit). Such elements form a crystalline structure by acquiring an additional four electrons from neighboring atoms, making the total number of electrons in the outer orbit eight, and thus completely filling the allowed eight electron states in the outer valence shell. The most stable shell configuration is that of the completely filled shell. A stable shell means that there are no free electrons at 0 K, and the conductivity is zero at 0 K. The bond picture for the intrinsic semiconductor is shown in Fig. A.1a.

The above situation is very often represented by energy levels. Free electrons are at a higher energy level than those in orbit. The energy levels of the free electrons are in the conduction band and those of the orbital electrons are in the valence band. The lowest energy level of the conduction band is denoted by E_c, and the uppermost energy level of the valence band is denoted by E_v. The energy gap between conduction and valence bands is denoted by the symbol E_g, and an energy greater than or equal to E_g is needed for an electron in orbit to be freed. For an intrinsic semiconductor at 0 K, all electrons are in the valence band, and none are in the conduction band. The Fermi function $f(E)$ gives the probability that a given energy state E is occupied by an electron. The *Fermi energy level*, which marks the border between filled and unfilled states, lies midway between the valence band and the conduction band, as shown in Fig. A.1b. There are no actual electron states right at this particular Fermi level, as it lies in a forbidden region. Rather, the Fermi level serves as a convenient border marker between filled and unfilled states.

Figure A.1c illustrates what happens as the temperature is raised in the intrinsic semiconductor. The increased thermal energy means that some electrons will gain enough energy to leave orbit and become free electrons or carrier electrons. These free

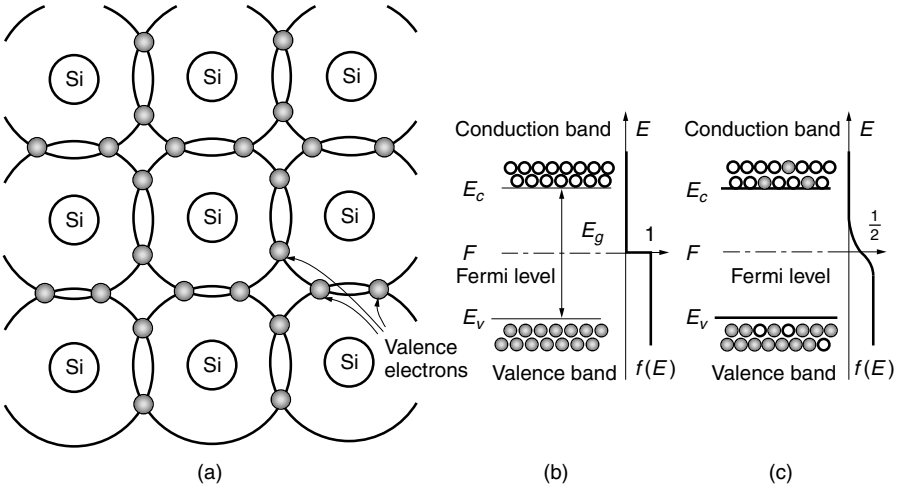


Figure A.1 Silicon intrinsic semiconductor. (a) Bond picture. (b) Energy levels at $T = 0$ K. (c) Energy levels at $T > 0$ K.

electrons are represented by the filled circles in the conduction band in Fig. A.1c. For each electron that leaves orbit to become a free electron, a vacant valence state is left behind, as represented by the open circles in the valence band. The Fermi level is still midway between the valence and conduction levels, and in a broader sense, the Fermi level is defined as that level for which the probability of the state being filled equals the probability of the state being empty. In an intrinsic silicon semiconductor at room temperature, the concentration of free electrons is of the order of 10^{10} cm^{-3} and is much lower than the 10^{22} cm^{-3} concentration characteristic of a metal conductor.

The atomic structure of the n-type layer will be explained next. The n-type layer is made by substituting some of the group IV atoms of the intrinsic semiconductor by donors that are group V atoms like N, P, As, Sb, Bi, or Pa and have five valence electrons. Let us take Si as the intrinsic semiconductor and As as the donor and refer to Fig. A.2a. Wherever the Si atom is replaced by the donor, the total number of available valence electrons becomes nine. At lower temperatures, the As atom uses eight electrons to complete the outermost shell, and the ninth electron is loosely bound in the Coulomb field of the nucleus of the As atom. As the temperature is raised, it only takes a little bit of energy for the ninth electron to break loose and become a free electron. The dissociation of the ninth electron from the As atomic site needs much less energy than that of the other electrons.

In terms of the energy band picture in Fig. A.2b, the energy levels of the ninth electrons of the donor atoms are just slightly below the conduction band, as only a small amount of energy is needed for these electrons to be freed. As a consequence, the Fermi level for the n-type semiconductor is closer to the conduction band, as shown in Fig. A.2b. As the concentration of donor atoms is increased, the Fermi level moves closer and closer to the conduction band, and for very high concentrations of donor atoms, the Fermi level is actually within the conduction band. For a given concentration of donor atoms, raising the temperature doesn't change the Fermi level, but it does cause more electrons to move to the conduction band, and as they do so, vacant states are left behind as shown in Fig. A.2c.

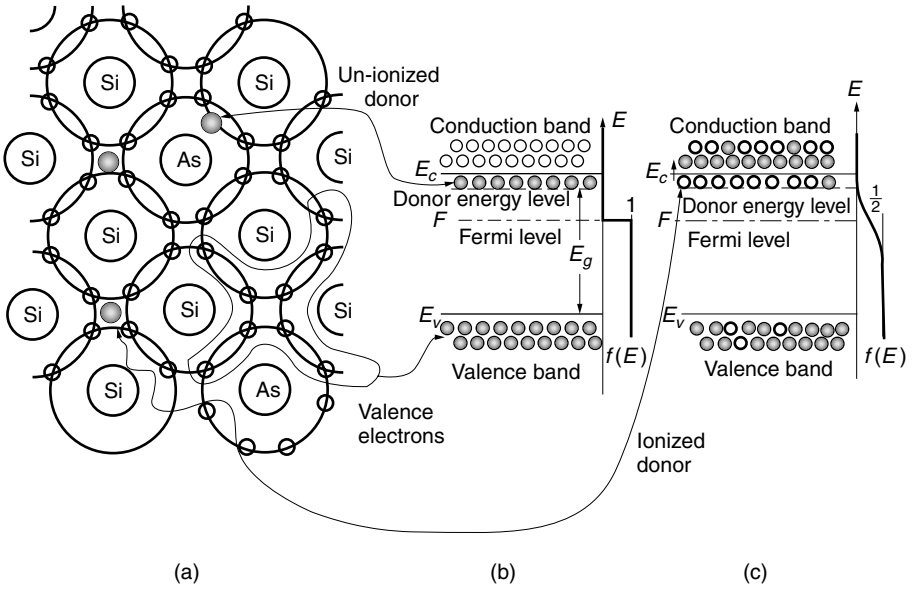


Figure A.2 n-Type semiconductor. (a) Bond picture. (b) Energy levels at $T = 0$ K. (c) Energy levels at $T > 0$ K.

Finally, the p-type semiconductor will be explained. The p-type semiconductor is made by substituting group III atoms for some of the group IV atoms in the intrinsic semiconductor. The group III atoms have three valence electrons and are called acceptor atoms. Examples are B, Al, Ga, In, Tl, or Ac. For the sake of illustration, let us take a specific p-type semiconductor consisting of Si with Ga as the acceptor, as shown in Fig. A.3a. Wherever the Si atom is replaced by the acceptor, the total number of available valence electrons is only seven, which is one short of forming a complete shell. In its eagerness to form a complete shell, the acceptor site may "accept" an electron from a nearby Si atom. If this happens, Ga will have a complete shell, but the Si atom that gave up the electron will now be incomplete. This process continues. The Si atom with the missing electron will capture an electron from another Si atom, thereby completing its shell but leaving an incomplete shell elsewhere. The movement of the missing site is electrically equivalent to the movement of a positive charge. This fictitious mobile charge is called a *hole*. Another term that is used to describe the hole is positive carrier.

It is important to realize that n-type and p-type semiconductors are electrically neutral overall. For the n-type semiconductor, for every electron that is freed from the As site to produce a negative carrier, an immobile positive charge is left behind. Likewise, for the p-type semiconductor, for every positive carrier (hole), there is a corresponding immobile negative charge at the Ga site.

The energy bands for the p-type semiconductor at $T = 0$ are shown in Fig. A.3b. The "holes"of Si are represented by the open circles in the valence band, while the electrons are represented by the filled circles. The unfilled eighth electron state of the acceptor atom is represented by an energy level just slightly above the silicon valence band, as it only takes a little bit of energy for an electron to leave the valence band and fill a state at the acceptor energy level. As a consequence, the Fermi level for

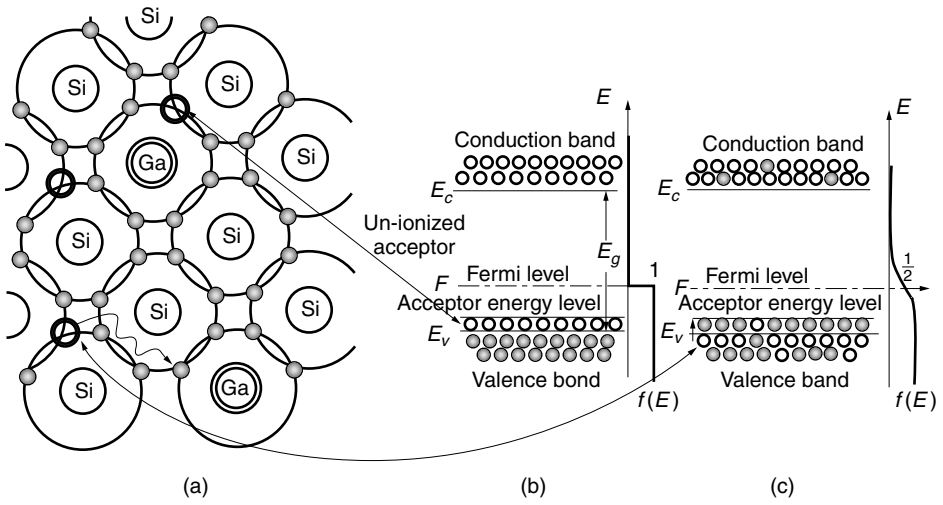


Figure A.3 p-Type semiconductor. (a) Bond picture. (b) Energy levels at $T = 0$ K. (c) Energy levels at $T > 0$ K.

the p-type semiconductor is closer to the valence band, and for high concentrations of acceptor atoms, the Fermi level is within the valence band. As the temperature is raised, some valence electrons will gain enough energy to become free electrons, leaving holes behind, as is the case for Fig. A.3c. Thus, at room temperature, a p-type semiconductor can have some free electrons. A point to keep in mind is that the density of holes will be much greater in a p-type semiconductor than the density of free electrons.

So far, only the simplest forms of semiconductors have been discussed. Before starting the topic of p-n junctions, it is worth mentioning that an intrinsic semiconductor can also be made by using compounds of group III and group V atoms in equal portions. Examples are gallium arsenide (GaAs) and indium phosphide (InP). Using group II atoms to replace some of the group III atoms in the III–V compound produces a p-type semiconductor. Likewise, using group VI atoms to replace some of the group V atoms in the III–V compound results in an n-type semiconductor.

So far, n- and p-type semiconductors have been considered separately, as indicated in Fig. A.4a. They will now be joined together as indicated in Fig. A.4b to form a p-n junction. The electron carrier density in the n-type layer is about 10^5 times larger than the p-type layer. The free electrons in the n-type layer start to diffuse into the p-type layer. This diffusion, however, does not continue forever. If it did, the result would be a perpetual current generator!

The neighborhood of the As atom was originally electrically neutral before the connection. As the electrons start leaving the site of As due to diffusion, the neighborhood of the As atom effectively becomes electrically positive. This positive charge $\oplus$ is, however, immobile because the As atom is held firm in the crystal. Immobile charges are encircled in Fig. A.4 and free charges are not. In order to emphasize this lack of mobility, the pattern of $\oplus$ and $\ominus$ in Fig. A.4a is preserved in Fig. A.4b.

The center of the junction region is normally an intrinsic region because when a free electron meets a hole (or empty site of the orbit), there is a good chance that

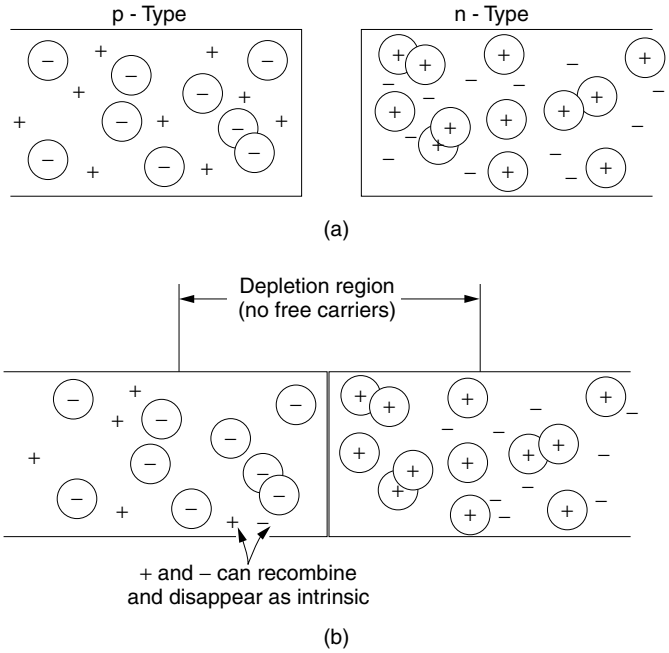


Figure A.4 Flow of charges after junction is made. (a) Before junction is made. (b) After junction is made.

the free electron will get back into orbit. One way of describing this process is to say that the free electron and hole have recombined, and the result of the recombination is a filled valence state. In analogy with particle physics, an alternate description is to view the process as the *annihilation* of a free electron–hole pair. The reverse process is called *pair production*.

It should be mentioned that recombination is most likely to take place between the positive and negative carriers. Recombination of a free electron with an immobile $\oplus$ charge is far less common. For instance, $\oplus$ of As has already disposed of the ninth electron, and because it is reasonably content with its completed external shell of eight electrons, it is almost indifferent to gaining back the ninth electron. On the other hand, the positive carrier means that a Si atom somewhere has an incomplete shell of seven valence electrons and this atom is very eager to grab a free electron to complete its outershell to a full complement of eight.

Getting back to the diffusion in the junction region, as the diffusion of free electrons progresses, the amount of accumulated immobile charges $\oplus$ increases. The Coulomb force between the immobile positive charge $\oplus$ and the negatively charged electrons slows down the diffusion. At equilibrium, there is no diffusion and the n-type layer is positively charged. With precisely the same reasoning but with opposite polarity, the equilibrium distribution of the holes is established. Holes diffuse into the n-layer, but at equilibrium, the diffusion is halted by the accumulation of negative $\ominus$ immobile charges left behind in the p-type layer. Thus, the p-type layer is negatively charged. The potential difference caused by the immobile charges is the contact potential of the junction.

A.2 I–V CHARACTERISTICS

The movement of electrons across the p-n junction will be calculated using energy bands. The contact potential V_0 causes a downward translation of the energy levels of the n-type layer by eV_0. In equilibrium, the Fermi level is the same throughout the p and n regions as indicated in Fig. A.5a. The case of lightly doped n and p regions at room temperature is considered, although the analysis is applicable to high dopant concentrations as well. Without an external bias voltage, the lower edges of the conduction bands of the p and n layers are higher than the Fermi level by A and B electron volts, respectively, as indicated in Fig. A.5a, and

$$A = eV_0 + B \tag{A.1}$$

With a forward bias V (positive potential) applied to the p-type layer and negative potential to the n-type layer, the whole energy band structure of the n-type layer translates upward by eV as indicated in Fig. A.5b. In order to cut down on the number of notations, the spacing between the levels rather than the levels themselves will be used here. When the term electron density is used without qualification, the density of free electrons is implied.

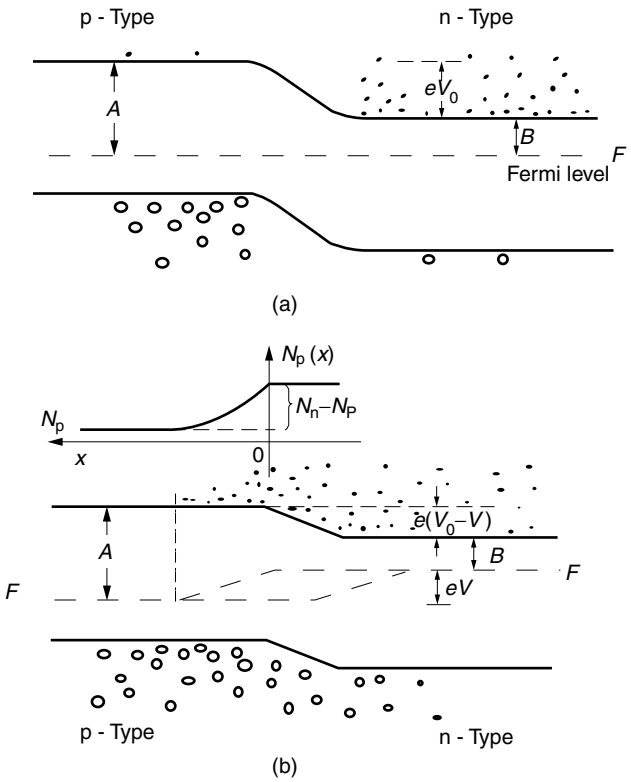


Figure A.5 Energy band structure of a p-n junction. (a) Energy levels without bias. (b) Energy levels with bias.

In the present explanation, simplicity will override a rigorous derivation of the expression and the Boltzmann distribution function rather than the Fermi distribution function will be used, even though this assumption is true only when the levels considered are away from the Fermi level. According to the Boltzmann distribution, the electron density at an energy level higher than the Fermi level by B electron volts is simply expressed by

$$N'_{\mathrm{n}} = N_0 e^{-B/kT} \tag{A.2}$$

where N_0 is the effective electron density extrapolated to the Fermi level. N'_{n} is the electron density at the bottom of the conduction band of the n-type layer.

Now, the electron density when a forward bias is applied is calculated. The electron density in the n layer at the level corresponding to the bottom of the conduction band of the p-type layer is, from Fig. A.5b,

$$N_{\mathrm{n}} = \left(N_0 e^{-B/kT}\right) e^{-e(V_0 - V)/kT} \tag{A.3}$$

Using Eq. (A.1) in (A.3),

$$N_{\mathrm{n}} = N_0 e^{-A/kT} e^{eV/kT} \tag{A.4}$$

The electron density N_{p} at the bottom of the conduction band of the p-type layer can be calculated in a similar manner and

$$N_{\mathrm{p}} = N_0 e^{-A/kT} \tag{A.5}$$

Note that Eq. (A.5) is identical with the first factor of Eq. (A.4), and

$$\begin{aligned} N_{\mathrm{p}} &< N_{\mathrm{n}} \quad (V > 0, \text{ forward bias}) \\ N_{\mathrm{p}} &< N_{\mathrm{n}} \quad (V < 0, \text{ backward bias}) \end{aligned} \tag{A.6}$$

A step of the electron density $(N_{\mathrm{n}} - N_{\mathrm{p}})$ is created across the junction. With a forward bias, electrons in the n-type layer diffuse into the p-type layer passing through the intrinsic layer without much loss of electrons. Inside the p-type layer, however, there is a sea of holes to annihilate the electrons (recombine with electrons). The electron density decays as the electrons go deeper into the p layer. Without a supply of electrons from an external source, the prescribed difference $(N_{\mathrm{n}} - N_{\mathrm{p}})$ could not be maintained because of the annihilation inside the p layer.

The amount of necessary supply current is calculated from the diffusion equation. The spatial distribution of the electrons inside the p-type layer is

$$N_{\mathrm{p}}(x) = N_{\mathrm{p}}(0) + (N_{\mathrm{n}} - N_{\mathrm{p}}) e^{-x/L_{\mathrm{n}}} \tag{A.7}$$

where x is the distance away from the junction. The origin is taken at the beginning of the p-type layer and not the center of the p-n junction. The electric current due to the gradient of the electron density is

$$J_e(x) = -eD_{\mathrm{n}} \frac{d}{dx} N_{\mathrm{p}}(x) \tag{A.8}$$

where D_n is the diffusion constant for the electrons. Insertion of Eq. (A.7) into (A.8) with $x = 0$ gives the current at the junction as

$$J_e(0) = \frac{eD_n}{L_n}(N_n - N_p) \tag{A.9}$$

By inserting Eqs. (A.4) and (A.5), one finally obtains the current for the p-n junction:

$$J_e = I_{so}\left(e^{eV/kT} - 1\right) \tag{A.10}$$

with

$$I_{so} = \frac{eD_n}{L_n}N_p + \frac{eD_p}{L_p}N_n \tag{A.11}$$

The second term in Eq. (A.11) accounts for the contribution of the holes. As seen from Eq. (A.10), I_{so} is the current when the p-n junction is deeply back-biased. It is called the saturated back-bias current.

The expressions so far are for the case without incident light. When light whose quantum hv is larger than the bandgap is incident on the junction, there is a chance

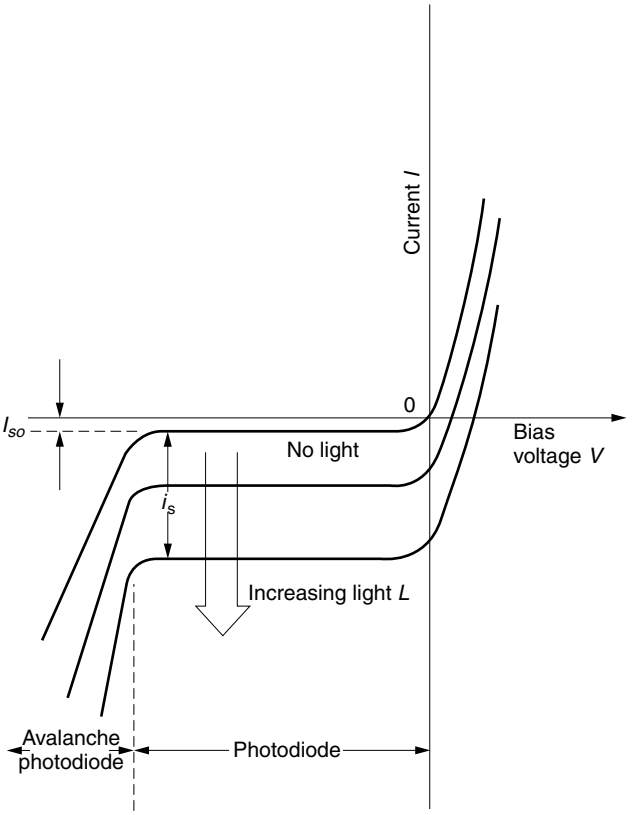


Figure A.6 Current–voltage characteristic of a p-n junction.

that an electron in the valence band will absorb the energy of the photon and move into the conduction band, leaving a hole behind. This process is called *photon-induced pair production*. The current due to the incident photons or photocurrent has to be added to Eq. (A.10). The characteristic curve including the photocurrent is shown in Fig. A.6. Increasing the number of incident photons causes a downward translation of the I–V curve. In the negative bias region, also referred to as the back-biased region, the n layer is positive and the p layer is negative. The negative bias region is used as a photodetector because the dark current (current in absence of the incident light) is at its minimum, having reached the saturated back-bias current. The output current is then primarily due to the photocurrent. In this region, the incremental impedance of the p-n junction is practically independent of the back-bias voltage.

The region of extreme negative bias is that of the avalanche effect. With a large bias voltage, the accelerated electrons start creating additional electron–hole pairs by colliding with the orbital electrons. The photocurrent can be multiplied up to a thousand times. This is the avalanche photodiode (APD).

APPENDIX B

MODE DENSITY

It will be proved that the density $m(\nu)$ of the modes per unit frequency per unit volume of a blackbody radiator is

$$m(\nu) = \frac{8\pi n_1^3}{c^3}\nu^2$$

The general expression for a sinusoidally oscillating wave is

$$\Psi(x, y, z, t) = \psi(x, y, z)e^{-j2\pi\nu t} \tag{B.1}$$

Equation (B.1) has to satisfy the wave equation:

$$\nabla^2\Psi - \frac{1}{v^2}\frac{\partial^2\Psi}{\partial t^2} = 0 \tag{B.2}$$

Using the method of separation of variables, Ψ will be found. Assume the solution is expressed by

$$\psi(x, y, z) = X(x)Y(y)Z(z) \tag{B.3}$$

Inserting Eq. (B.3) into (B.2) and then dividing the equation by Ψ gives

$$\frac{X''(x)}{X(x)} + \frac{Y''(y)}{Y(y)} + \frac{Z''(z)}{Z(z)} - \left(\frac{2\pi\nu}{v}\right)^2 = 0 \tag{B.4}$$

Each term of Eq. (B.4) is a function of x, y, and z, respectively. For Eq. (B.4) to be satisfied at any location in space, each individual term has to be constant.

$$\frac{X''(x)}{X(x)} = k_x, \qquad \frac{Y''(y)}{Y(y)} = k_y, \qquad \frac{Z''(z)}{Z(z)} = k_z \tag{B.5}$$

with

$$k_x^2 + k_y^2 + k_z^2 = \left(\frac{2\pi\nu}{v}\right)^2 \tag{B.6}$$

The general solution for $X(x)$ is

$$X(x) = A \sin k_x X + B \cos k_x X \tag{B.7}$$

$X(x)$ has to satisfy the *boundary conditions*.

Two ways of setting up the boundary conditions will be presented: the periodic boundary condition that was not mentioned in the text and the standing-wave boundary condition that has already been mentioned in the text.

Can we really take the size of a crystal as the boundary of a gigantic size potential well in the y and z directions? Strictly speaking, if indeed they are the boundaries of a big potential well, the properties of the crystal would change if the crystal is broken in half. The surface shape or surface roughness of the crystal would change the shape of the potential well. Would the properties of the material then depend on the shape?

The periodic boundary was proposed as an alternative boundary. The periodic boundary is applicable to a crystal of infinite size. The assumption is made that there is a certain distance L at which the wavefunction repeats itself in a gigantic size crystal. The L is taken definitely much larger than 100 Å but much smaller than the physical dimensions of the crystal. In most cases, L drops out from the final answer by expressing the quantity concerned per volume or per area. In the present case as well, the same value of the density of states can be obtained whether the physical size or periodic boundary is used as a boundary as explained later. With the periodic boundary condition,

$$X(0) = X(L_x) \tag{B.8}$$

and

$$k_x L_x = 2\pi n_x \tag{B.9}$$

where n_x is an integer including zero for both forward and backward waves.

$$n_x = 0, \pm 1, \pm 2, \pm 3, \ldots \tag{B.10}$$

Similarly, for k_y and k_z

$$k_y = 2\frac{n_y}{L_y}\pi \tag{B.11}$$

$$k_z = 2\frac{n_z}{L_z}\pi \tag{B.12}$$

where

$$\begin{aligned} n_y &= 0, \pm 1, \pm 2, \pm 3, \ldots \\ n_z &= 0, \pm 1, \pm 2, \pm 3, \ldots \end{aligned} \tag{B.13}$$

The k_x, k_y, k_z set of coordinates shown in Fig. B.1a is useful for counting the number of modes. The values of k_x, k_y, and k_z are discrete as specified by the integers, and the values form lattice points. Every lattice point represents one mode. The wavenumber k_i with a particular frequency ν_i is $k_i = 2\pi\nu_i/v$ and is represented by the length connecting the origin to the lattice point. Modes whose resonance frequency is less than ν are inside a sphere of radius $k = 2\pi\nu/v$.

The total number of modes whose resonance frequency is less than ν is now calculated. The total number M of lattice points inside the sphere with radius $2\pi\nu/v$ is the volume of the sphere divided by the density of the lattice points $(2\pi/L)^3$ with the assumption $L_x = L_y = L_z = L$.

$$M = \frac{4}{3}\pi\left(\frac{2\pi\nu}{v}\right)^3 \Big/ \left(\frac{2\pi}{L}\right)^3 = \frac{4}{3}\pi\left(\frac{L\nu}{v}\right)^3 \tag{B.14}$$

There are two orthogonal directions of polarization and the total number M_t of modes taking polarization into account is

$$M_t = 2M \tag{B.15}$$

Thus, the increase $m(\nu)\,d\nu$ due to an increase in frequency from ν to $\nu + d\nu$ per unit volume is

$$m(\nu)d\nu = \frac{1}{L^3}\frac{dM_t}{d\nu}d\nu \tag{B.16}$$

From Eqs. (B.14) to (B.16), the mode density is

$$m(\nu) = \frac{8\pi\nu^2}{v^3} \tag{B.17}$$

where $v = c/n$ and where n is the refractive index of the blackbody radiator.

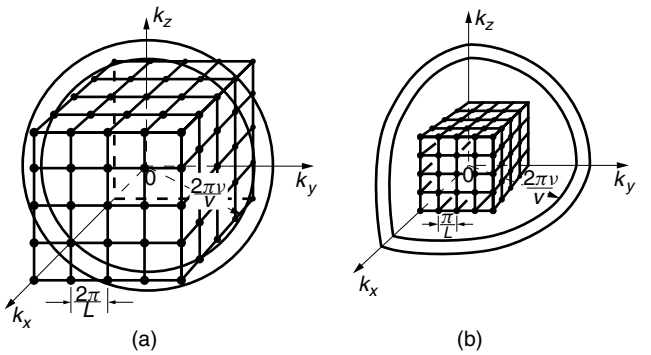


Figure B.1 Lattice points representing the modes whose frequencies are lower than ν. (a) When the periodic boundary condition is used. (b) When the standing-wave boundary condition is used.

Lastly, it will be shown that the standing-wave boundary condition provides the same result as the periodic boundary condition does. The condition for a standing wave to exist in a medium bordered by $x = 0$ and $x = L$ is that Eq. (B.7) satisfy

$$\begin{aligned} B &= 0 \\ k_x &= \frac{n_x}{L}\pi \quad \text{with } n_x = 1, 2, 3, 4, \ldots \end{aligned} \tag{B.18}$$

where n_x is called the mode number. Because

$$\sin(-k_x x) = -\sin k_x x$$

the difference between the standing waves with k_x and $-k_x$ is a difference in the sign of the amplitude. After all, the sine function oscillates between 1 and -1 with time. Standing waves whose amplitudes differ only in sign are considered as the same modes.

Consequently, only positive numbers are used to designate the mode number.

The spacing between adjacent modes when using the standing-wave boundary condition is

$$\frac{\pi}{L}$$

whereas the spacing between modes when using the periodic boundary condition is

$$2\frac{\pi}{L}$$

In one dimension, the mode density for the standing-wave boundary condition is twice as large as for the periodic boundary condition, and in three dimensions, the density is eight times as large.

Next, the total number M of modes that corresponds to Eq. (B.14) will be calculated for the case of the standing-wave boundary condition.

This time, only the positive numbers of k_x, k_y, k_z are allowed, as shown in Fig. B.1b. The total number M of modes whose resonance frequency is less than ν is contained only inside the positive octant of the sphere, but the density of the lattice points is 2^3 times larger than the case of the periodic boundary. The result is

$$M = \frac{1}{8}\frac{4}{3}\pi\left(\frac{2\pi\nu}{v}\right)^3 \Big/ \left(\frac{\pi}{L}\right)^3 \tag{B.19}$$

which is identical to Eq. (B.14).

APPENDIX C

PERTURBATION THEORY

We will deal only with the case when F is expressed in a form such as

$$F(x, y, z, t) = F(x, y)e^{-j\omega t + j\beta z}$$

Perturbation theory allows us to calculate the change in the propagation constant β due to a small change in the index of refraction of the medium [1].

A scalar wave equation before it is perturbed is designated by the subscript 1 and that after the perturbation due to Δk resulting in the change $\Delta\beta$ in β is designated by the subscript 2.

$$\nabla_t^2 F_1 + (k_1^2 - \beta_1^2)F_1 = 0 \tag{C.1}$$

$$\nabla_t^2 F_2^* + (k_2^2 - \beta_2^2)F_2^* = 0 \tag{C.2}$$

were the operator ∇_t is two dimensional. The subscript t will be suppressed.

Introduction of the divergence identity

$$\boldsymbol{\nabla}\cdot(F_2^*\boldsymbol{\nabla}F_1 - F_1\boldsymbol{\nabla}F_2^*) = F_2^*\boldsymbol{\nabla}^2 F_1 - F_1\boldsymbol{\nabla}^2 F_2^* \tag{C.3}$$

suggests that Eq. (C.1) be multiplied by F_2^* and Eq. (C.2) by F_1, and the results be subtracted from each other.

$$\boldsymbol{\nabla}\cdot(F_2^*\boldsymbol{\nabla}F_1 - F_1\boldsymbol{\nabla}F_2^*) = F_1F_2^*(k_1^2 - k_2^2 + \beta_2^2 - \beta_1^2) \tag{C.4}$$

The surface integral is performed on both sides of Eq. (C.4). The left-hand side is converted into a line integral by the two-dimensional Green's theorem [2].

$$\oint_c (F_2^*\nabla F_1 - F_1\nabla F_2^*)\cdot d\mathbf{l} = \int_S F_1F_2^*[k_1^2 - k_2^2 + \beta_2^2 - \beta_1^2]\,dS \tag{C.5}$$

S for the surface integral is taken in a circle in the cross-sectional plane (perpendicular to the fiber axis) centered at the fiber axis with radius much larger than the core radius of the fiber, and $\mathbf{l}$ for the line integral is taken in the circumference of this circle

as shown in Fig. C.1. If the field on the boundary of the fiber core decays and can be approximated as zero, the left-hand side of Eq. (C.5) becomes zero, and

$$\int (\beta_2^2 - \beta_1^2) F_1 F_2^* \, dS = \int (k_2^2 - k_1^2) F_1 F_2^* \, dS \tag{C.6}$$

The change in β on the left-hand side is related to that in k on the right-hand side.

Now, the propagation constants before and after the nonlinear effect are

$$k_1^2 = \omega^2 \mu \epsilon_0 \epsilon = k^2 \epsilon \tag{C.7}$$

$$k_2^2 = k^2 (\epsilon + \delta\epsilon) \tag{C.8}$$

where ϵ is the relative dielectric constant of the medium. Hence

$$k_2^2 - k_1^2 = k^2 \delta\epsilon \tag{C.9}$$

$$\beta_2^2 - \beta_1^2 \doteq 2\beta \Delta\beta \tag{C.10}$$

Inserting Eqs. (C.9) and (C.10) into (C.6) gives

$$\Delta\beta = \frac{k^2 \int \delta\epsilon F_1 F_2^* \, dS}{2\beta \int F_1 F_2^* \, dS} \tag{C.11}$$

Equation (C.11) can be expressed in terms of the refractive index by noting

$$\epsilon + \delta\epsilon = (n_1 + \delta_n)^2 = n_1^2 + 2 n_1 \delta_n \tag{C.12}$$

and

$$\delta\epsilon = 2 n_1 \delta n \tag{C.13}$$

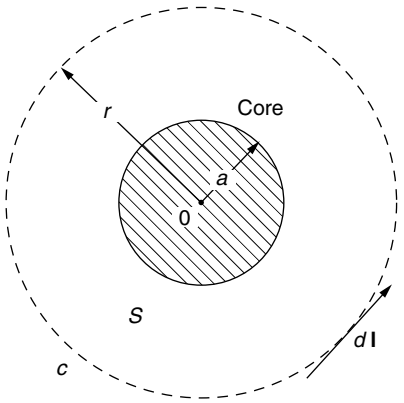


Figure C.1 Geometry of surface and line integrals.

Insertion of Eq. (C.13) in (C.11) finally leads to

$$\Delta\beta = \frac{k^2 \iint n_1 \delta n |F(x, y)|^2 \, dx \, dy}{\beta \iint |F(x, y)|^2 \, dx \, dy} \tag{C.14}$$

where

$$F_1 = F_2 = F(x, y) \tag{C.15}$$

we assumed.

Equation (C.14) is useful for calculating the amount of change $\Delta\beta$ in the propagation constant due to the change δn in the refractive index of the fiber core. The change δn may be attributed to strain (fiber strain sensor), bending (fiber displacement sensor) temperature (fiber temperature sensor) or nonlinearity due to high intensity of light (self phase modulation of the soliton wave).

1. H. A. Haus, *Waves and Fields in Optoelectronics*, Prentice-Hall Inc., Englewood Cliffs, NJ, 1984.
2. W. Hauser, *Introduction to the Principles of Electromagnetism*, Addison-Wesley, Reading, Massachusetts, 1971.

ANSWERS TO PROBLEMS

Chapter 9

9.1 The γ values in the lower half-plane are negative, which means a growing wave and that does not make physical sense. Both Kd tan Kd and $-Kd$ cot Kd are even functions of Kd, and only one side is needed.

9.2 From Fig. 9.2 when V is an integral multiple of $\pi/2$,

$$V = Kd \tag{1}$$

Hence,

$$(kn_1)^2 - (kn_2)^2 = K^2 \tag{2}$$

From Fig. 9.5, we have

$$\cos\theta = \frac{K}{n_1 k} \tag{3}$$

Inserting Eq. (3) into (2) gives

$$\sin\theta = \frac{n_2}{n_1} \tag{4}$$

Recall such θ is at the critical angle associated with the core–cladding boundary,

$$\theta = \theta_c$$

9.3 From Fig. 9.2, the value of V that supports five modes is

$$2\pi < V < \tfrac{5}{2}\pi$$

$$\frac{4\pi}{k\sqrt{n_1^2 - n_2^2}} < 2d < \frac{5\pi}{k\sqrt{n_1^2 - n_2^2}}$$

$$9.67\ \mu\text{m} < 2d < 12.09\ \mu\text{m}$$

9.4 The correct word decrease or increase is underlined.

If the thickness of the slab guide is *decreased* with all other physical constants fixed, the normalized thickness V(*increases,* <u>*decreases*</u>) and $K_2 d$ of the TM_2 mode (*increases,* <u>*decreases*</u>) and the value of K_2 (<u>*increases*</u>, *decreases*), and this means that the value of β_2 (*increases,* <u>*decreases*</u>). Thus, in order to obtain the region of a larger effective index of refraction $N = \beta_2/k$, the thickness 2d has to be (<u>*increased*</u>, *decreased*).

9.5 From Figs. 9.6a and 9.6b,

$$K_{\max} = k\sqrt{n_1^2 - n_2^2}$$

$$n_1^2 = \left(\frac{K_{\max}}{k}\right)^2 + n_2^2$$

$$n_1 = \sqrt{\left(\frac{1.3 \times 1}{2\pi}\right)^2 + 1^2} = 1.02$$

9.6 The angle θ that the incident ray at $x = x_0$ makes with the boundary layer of the guide is (see Fig. A9.6)

$$\sin\theta = \frac{r + x_0}{r + d}$$

The angle θ is smallest for the ray entering at $x_0 = -d$. This has to be larger than the critical angle for the light to remain in the core. At the critical angle we have

$$\sin\theta_c = \frac{n_2}{n_1}$$

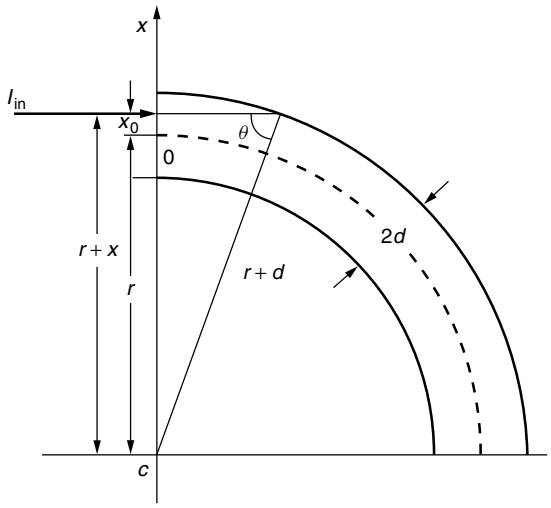


Figure A9.6 Allowed angle of bend.

and hence

$$\frac{n_2}{n_1} = \frac{r-d}{r+d}$$

$$\therefore r = \frac{n_1+n_2}{n_1-n_2}d = \left(\frac{2}{\Delta} - 1\right)d$$

For $d = 50$ μm and $\Delta = 0.055$, $r = 1.77$ mm.

9.7 **(a)** Taking the origin of the x axis in the center of the core layer,

$$\begin{matrix} H_y(x=d) \\ H_y(x=-d) \\ E_z(x=d) \\ E_z(x=-d) \end{matrix} \begin{bmatrix} \cos Kd & \sin Kd & -e^{-\gamma_0 d} & 0 \\ \cos Kd & -\sin Kd & 0 & -e^{\gamma_2 d} \\ -K'\sin Kd & K'\cos Kd & \gamma_0' e^{-\gamma_0 d} & 0 \\ K'\sin Kd & K'\cos Kd & 0 & -\gamma_2' e^{-\gamma_2 d} \end{bmatrix} = \Delta$$

with

$$K' = \frac{K}{n_1^2}, \qquad \gamma_0' = \frac{\gamma_0}{n_0^2}, \qquad \gamma_2' = \frac{\gamma_2}{n_2^2}$$

The condition of $\Delta = 0$ is imposed. After some manipulation of the determinant, the characteristic equation becomes

$$\frac{\left(\dfrac{\gamma_0}{n_0'^2} + \dfrac{\gamma_2}{n_2'^2}\right)K}{K^2 + \dfrac{\gamma_0\gamma_2}{n_0'^2 n_2'^2}} = \tan 2Kd$$

where

$$n_0' = \frac{n_0}{n_1}, \quad n_2' = \frac{n_2}{n_1}$$

(b) Equation (9.100) for the given configuration is

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \cos 2Kd & (1/Z_1)\sin 2Kd \\ -Z_1 \sin 2Kd & \cos 2Kd \end{bmatrix}$$

The characteristic equation (9.108) is

$$Z_2 \cos 2Kd + \frac{Z_0 Z_2}{Z_1}\sin 2Kd - Z_1 \sin 2Kd + Z_0 \cos 2Kd = 0$$

which is rewritten as

$$(Z_0 + Z_2)\cos 2Kd - \left(Z_1 - \frac{Z_0 Z_2}{Z_1}\right)\sin 2Kd = 0$$

Inserting the values of Z_0, Z_2, and Z_1,

$$Z_0 = \frac{j\gamma_0}{\omega\epsilon_0}, \quad Z_2 = \frac{j\gamma_2}{\omega\epsilon_0\epsilon_{r2}}, \quad Z_1 = \frac{jK}{\omega\epsilon_0\epsilon_{r1}}$$

gives

$$\frac{K\left(\frac{\gamma_0}{n_0'^2} + \frac{\gamma_2}{n_2'^2}\right)}{K^2 - \frac{\gamma_0\gamma_2}{n_0'^2 n_2'^2}} = \tan 2Kd$$

Both answers are identical.

9.8 From Eqs. (9.119) and (9.200),

$$\frac{E_{z0}(x)}{H_{y0}(x)} = \frac{Z_1 \sin 2K_1 d - Z_2 \cos 2K_1 d \tanh \gamma_2 s}{\cos 2K_1 d + (Z_2/Z_1) \sin 2K_1 d \tanh \gamma_2 s}$$

Dividing both the denominator and numerator by $\cos 2K_1 d$ gives

$$\frac{E_{z0}(x)}{H_{y0}(x)} = \frac{Z_1 \tan 2K_1 d - Z_2 \tanh \gamma_2 s}{1 + (Z_2/Z_1) \tan 2K_1 d \tanh \gamma_2 s}$$

From T_- in Eqs. (9.170) and the even modes characteristic Eq. (9.172), the right-hand side of the above equation is Z_0.

Chapter 10

10.1 Note that expressions for the ridge guide can be obtained immediately by flipping it upside down and interchanging n_0 and n_2 of the imbedded guide shown in Fig. 10.4 that was treated in Example 10.1 in the text.

The normalized width V_w of the present problem, however, is different from that of Example 10.1.

$$V_w = k\sqrt{n_1^2 - n_0^2}w = 5.47\pi \text{ rad}$$

In every $\pi/2$ radians of V_w, one mode is generated. The total number of modes including the zeroth mode is 11 modes.

For the asymmetric modes, exactly the same results as those of Example 10.1 are applicable. There are two possible modes in the x direction. Hence, the combination of the two modes creates

$$E^x_{1,1}, \quad E^x_{1,2}, \quad E^x_{1,3}, \ldots, E^x_{1,11}$$
$$E^x_{2,1}, \quad E^x_{2,2}, \quad E^x_{2,3}, \ldots, E^x_{2,11}$$

There are 22 TM-like modes altogether.

10.2 The light beam will deflect toward the base of the cone, and this device is used as a prism.

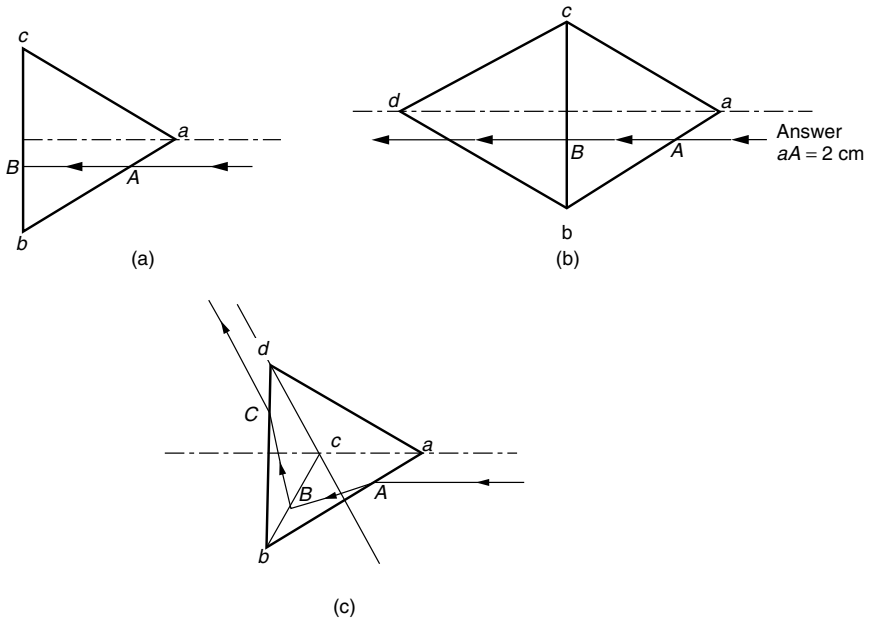


Figure A10.4 Solution of Problem 10.4. (a) Trace *abc*. (b) Traces *abc* and *bcd*. (c) Top view.

10.3 An incident parallel beam is split into two beams, both deflected toward the base of their respective cones. This configuration is used as a beamsplitter.

10.4 The answer is in Fig. A10.4.

10.5 Since the light is Y-directed propagation, the "cross-sectional ellipse" is obtained with $Y = 0$ in Eq. (5.4).

From Eq. (5.10) with $\varepsilon = \varepsilon_z$,

$$a_{11} = \frac{1}{n_o^2} + r_{13}\varepsilon_z$$

$$a_{33} = \frac{1}{n_e^2} + r_{33}\varepsilon_z$$

In a manner similar to Example 5.1, the cross-sectional ellipse is obtained as

$$\frac{x_2}{\left(n_o - \frac{1}{2}r_{13}n_o^3\varepsilon_z\right)^2} + \frac{z^2}{\left(n_e - \frac{1}{2}r_{33}n_e^3\varepsilon_z\right)^2} = 1$$

Since the TE mode is excited, the direction of the polarization is in the Z direction.

Hence, the change $\Delta\phi$ in phase is

$$\Delta\phi = \tfrac{1}{2}k_0 r_{33} n_e^3 \varepsilon_z l$$

with

$$r_{33} = 32.2 \times 10^{-12}$$

$$n_e = 2.2$$

$$\varepsilon_z = 3 \times 10^6 \text{ V/m}$$
$$l = 1 \text{ mm}$$
$$\Delta\phi = 2.48 \text{ rad}$$

Chapter 11

11.1 From the viewpoint of the numerical aperture of the fiber,

$$\sin\phi = \sqrt{n_1^2 - n_2^2} = 0.20$$
$$\phi = 11.5^{\circ}$$
$$f = \frac{0.4}{\tan 11.50^{\circ}} = 1.96 \text{ mm}$$

From the viewpoint of the diffraction limit (see Section 1.4.4), the radius r_i of the first null of the Airy diffraction pattern on the end surface of the fiber is obtained.

First zero of Eq. (1.99) appears when

$$a\rho = 0.61$$

or

$$a\frac{r_i}{\lambda z_i} = 0.61$$

with

$$z_i = f$$
$$f = \frac{1}{0.61}\frac{a}{\lambda}r_i$$
$$= \frac{1}{0.61}\frac{400}{1.55}(4.75)$$
$$= 2\,010 \text{ μm}$$

Thus,

$$1.96 \text{ mm} < f < 2.01 \text{ mm}$$

11.2 From Eq. (11.96) with $\nu = 1$,

$$\frac{J_0(Ka)}{KaJ_1(Ka)} = \frac{K_0(\gamma a)}{\gamma a K_1(\gamma a)} = \frac{1}{\gamma a}\left(1 - \frac{1}{2\gamma a}\right)$$

where Eq. (11.101) with $\nu = 1$ was used. With $\gamma a \to \infty$,

$$J_0(Ka) = 0$$

and the asymtotes of the $HE_{1\mu}$ are

$$Ka = 2.4, 5.5, 8.7, \ldots$$

11.3 TM_{02}.

11.4 **(a)** From Table 11.1 the cutoff values ($\gamma a \to 0$) and far from the cutoff values ($\gamma a \to \infty$) are:

	$HE_{3\mu}$	$EH_{1\mu}$
$\gamma a \to 0$	$J_1(Ka) = 0$, except $Ka = 0$	$J_1(Ka) = 0$, except $Ka = 0$
$\gamma a \to \infty$	$J_2(Ka) = 0$	$J_2(Ka) = 0$

The cutoff values are:

Mode	HE_{31}	HE_{32}	HE_{33}	EH_{11}	EH_{12}	EH_{13}
Cutoffs	3.8	7.0	10	3.8	7.0	10

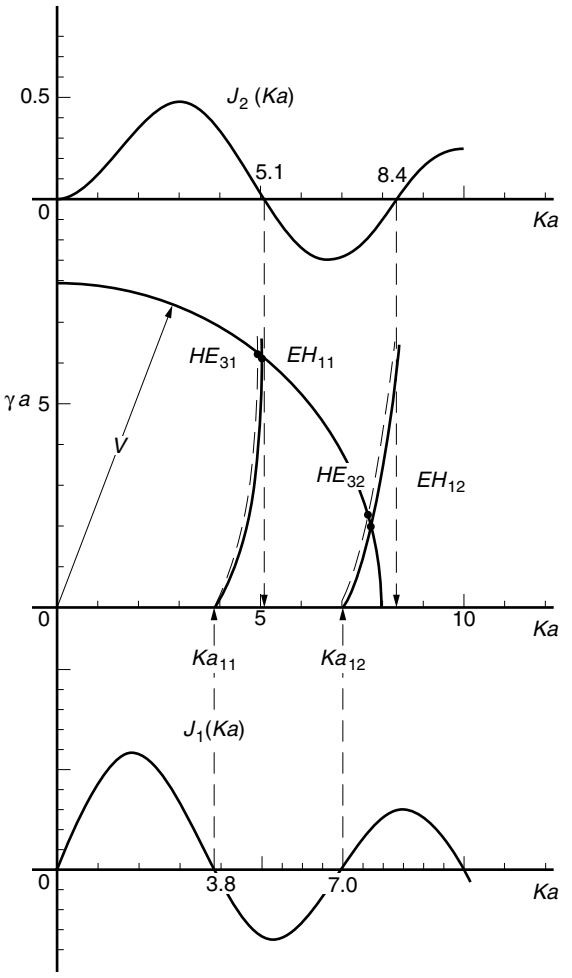


Figure A11.4 Curves of the characteristic equations of $HE_{3\mu}$ and $EH_{1\mu}$ modes.

(b)

	Mode	V_c	Source
(1)	HE_{11}	0	Fig. 11.17
(2)	TE_{01}	2.4	Fig. 11.16
(3)	TM_{01}	2.4	Fig. 11.16
(4)	HE_{21}	2.4	Eq. (11.105)
(5)	HE_{31}	3.8	Fig. A11.4
(6)	EH_{11}	3.8	Fig. A11.4
(7)	HE_{12}	3.8	Fig. 11.17

11.5 Figure A11.5.

11.6 **(a)** LP_{01}, HE_{11}.

(b) LP_{21}, HE_{31}, EH_{11}.

(c) LP_{11}, TE_{01}, TM_{01}, HE_{21}.

(d) LP_{32}, HE_{42}, EH_{22}.

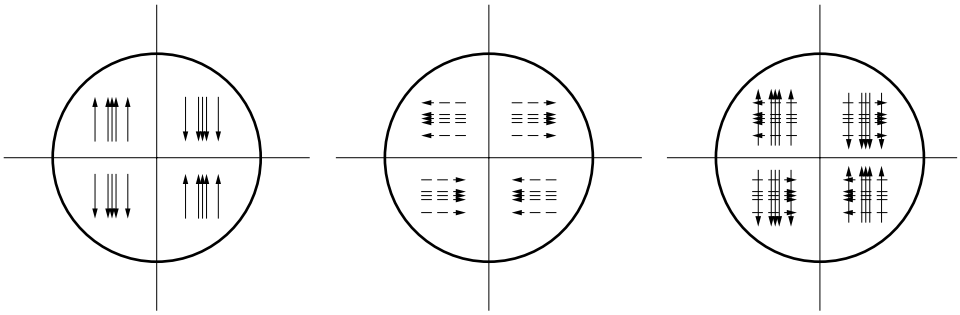


Figure A11.5 Field pattern of the LP_{21} mode.

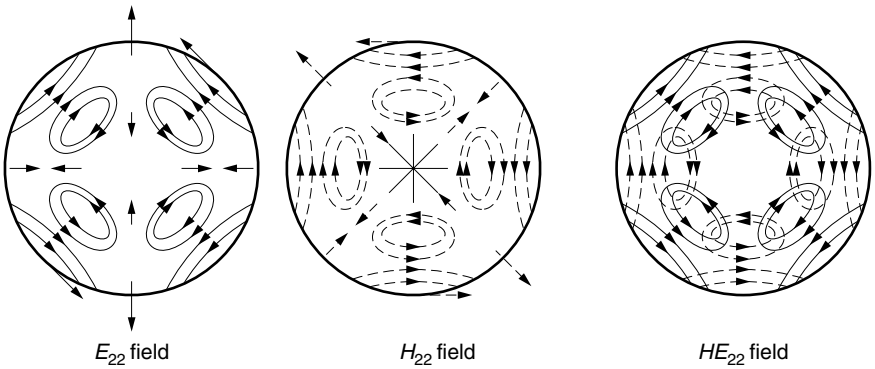


Figure A11.7 Field pattern of the HE_{22} mode.

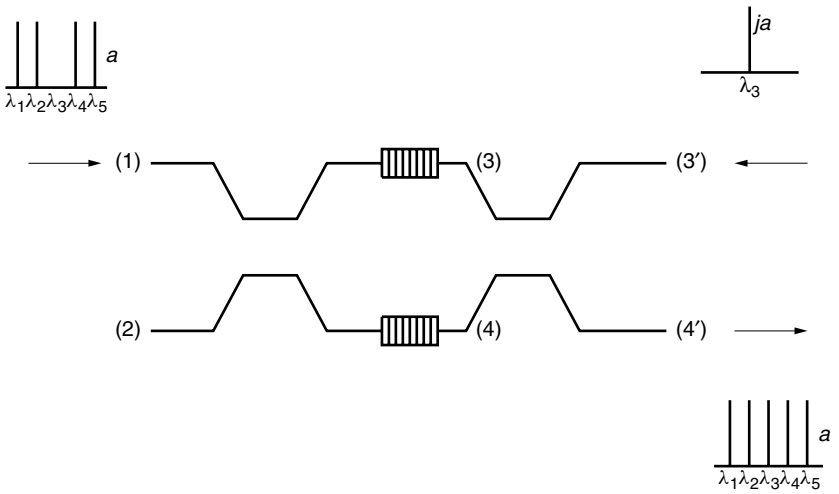


Figure A11.8 Insertion of λ_3 into a WDM transmission stream.

11.7 From Eq. (11.130), the field expressions for HE_{22} are

$$E_r = EJ_1(Kr)\cos 2\phi$$
$$E_\phi = -EJ_1(Kr)\sin 2\phi$$
$$H_r = -\frac{1}{\eta_1}E_\phi$$
$$H_\phi = \frac{1}{\eta_1}E_r$$

The mode pattern is drawn in Fig. A11.7.

11.8 The answer is shown in Fig. A11.8.

Chapter 12

12.1 (a)

$$i = \eta\frac{e}{h\nu}P$$
$$\nu = \frac{3\times 10^{14}}{0.84} = 357 \text{ THz}$$
$$P = 10^{-4.3} = 5\times 10^{-5} \text{ mW} = 50 \text{ nW}$$

In the theoretical limit ($\eta = 100\%$),

$$i = \frac{(1.6\times 10^{-19})(5\times 10^{-8})}{(6.63\times 10^{-34})(3.57\times 10^{14})}$$
$$= 3.4\times 10^{-8} = 33.8 \text{ nA}$$

(b)

$$i = 0.675\eta \text{ A/W}$$
$$\eta = \frac{0.5}{0.675} = 0.74$$

12.2 Let us say that the phase jitter in the transmitting and local oscillators is θ and θ_L, and the phase jitter in the horizontal and vertical waves is α_H and α_V. Then Eq. (12.43) becomes

$$E_H = A(t)\sqrt{\beta}\cos(\omega_c t + \theta + \alpha_H)$$
$$E_V = A(t)\sqrt{1-\beta}\cos(\omega_c t + \theta + \alpha_V)$$
$$E_{LH} = E_{LV} = \frac{1}{\sqrt{2}}E_{\text{LO}}\cos(\omega_L t + \theta_L)$$

The outputs from mixer M_V and M_H are

$$i_H = 2K\left\langle\left[\frac{[A(t)]^2\beta}{2} + \frac{\sqrt{\beta}}{\sqrt{2}}A(t)E_{\text{LO}}\cos(\omega_{\text{IF}}t + \theta - \theta_L + \alpha_H) + \frac{P_L}{2}\right]^2\right\rangle$$
$$i_V = 2K\left\langle\left[\frac{[A(t)]^2(1-\beta)}{2}\right.\right.$$
$$\left.\left. + \frac{\sqrt{1-\beta}}{\sqrt{2}}A(t)E_{\text{LO}}\cos(\omega_{\text{IF}}t + \theta - \theta_L + \alpha_V) + \frac{P_L}{2}\right]^2\right\rangle$$

The signals are band-pass filtered for the IF frequency components. Applying the square operation, and then low-pass filtering gives

$$i_H^D = [KE_{\text{LO}}A(t)]^2\beta$$
$$i_V^D = [KE_{\text{LO}}A(t)]^2(1-\beta)$$

Thus, the phase jitter has disappeared.

12.3 The outputs from the delay-and-detect circuit for an arbitrary $\Delta\omega\,\tau$ are

$$i_0 = K\cos[\Delta\omega t - \Delta\omega\tau - \Phi(t) - \theta_p]$$
$$\times \cos[\Delta\omega t - \Phi(t) - \theta_p] \qquad \text{“0” bit}$$
$$i_1 = K\cos[\Delta\omega t - \Delta\omega\tau - \Phi(t) - \pi - \theta_p]$$
$$\times \cos[\Delta\omega t - \Phi(t) - \theta_p] \qquad \text{“1” bit}$$

The output from the low-pass filter is

$$i_0 = \tfrac{1}{2}K\cos(\Delta\omega\tau) \qquad \text{“0” bit}$$
$$i_1 = \tfrac{1}{2}K\cos(\Delta\omega\tau + \pi) \qquad \text{“1” bit}$$

When $i_0 = i_1$, the function of interrogation is lost. The expression for $i_0 - i_1$ is

$$i_0 - i_1 = K \sin \frac{\pi}{2} \sin \left(\Delta\omega\tau + \frac{\pi}{2} \right)$$

Setting $i_1 - i_0 = 0$ gives the condition

$$\Delta\omega\tau + \frac{\pi}{2} = N\pi$$

which leads to

$$B = \frac{4\Delta f}{2N - 1}$$

with $N = 0, 1, 2, 3, \ldots$.

Chapter 13

13.1 **(a)** Absorption is observed at wavelengths corresponding to each possible transition. The absorption lines are shown in Fig. A13.1b. The depth of each absorption line depends on the atomic structure of Nd YAG as well as the population of the atoms in the levels associated with the transitions.

The task of the spectroscopist is to determine the energy levels conversely from the observed absorption lines such as these.

(b) With the pump light on, the distribution of the population of each level is upset. The atoms in the ground level are pumped to the pump band. Transition rate γ_{32} (the inverse of the lifetime $1/\tau_{32}$) is high and the atoms in level E_3 come down quickly to E_2. The transition rate γ_{10} from E_1 to E_0 is also fast. The transition between E_2 and E_1 is much slower than any of the above. As a result, the population of atoms in E_2 accumulates and that in E_1 diminishes via decay to E_0. These two actions contribute to a population inversion between levels E_2 and E_1. When external light with frequency $\nu_{21} = (E_2 - E_1)/h$ illuminates the rod, stimulated emission is larger than the stimulated absorption and the intensity of the output light becomes larger than that of the input light.

The populations among the levels are controlled by the multilevel rate equations and all absorption lines change to some degree, but the biggest change is a significant increase at $\nu_{21} = (E_2 - E_1)/h$ as indicated by the dotted line in Fig. A13.1b. The deeper absorption at the pump light frequency is due to the filter F.

(c) From

$$\begin{aligned} P_1 &= Ae^{gl} \\ 10^{-2} &= 2 \times 10^{-3} e^{gl} \\ \therefore\ e^{gl} &= 5 \\ P_2 &= Ae^{2gl} = (2 \times 10^{-3})(5^2) = 50 \text{ mW} \end{aligned}$$

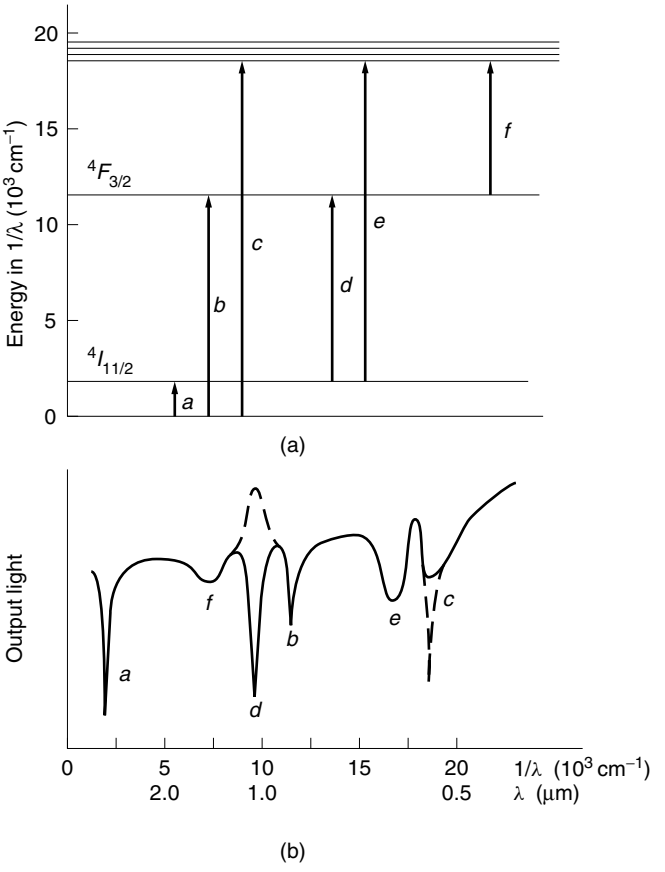


Figure A13.1 Absorption spectrum of Nd YAG.

13.2 (a)

$$G = 10 \log e^{gz}$$

$$z = \frac{G}{10\ g \log e}$$

$$g = \sigma_s(N_2 - N_1)$$

$$= (7.0 \times 10^{-25})[(4.8 - 1.8) \times 10^{23}]$$

$$= 0.21$$

$$z = \frac{35}{10(0.21)(0.434)} = 38.4 \text{ m}$$

(b) The ASE noise $= (G - 1)n_{\text{spon}} m_t h\nu \Delta\nu_f$

$$n_{\text{spon}} = \frac{N_2}{N_2 - N_1} = \frac{4.8 \times 10^{23}}{(4.8 - 1.8) \times 10^{23}} = 1.6$$

G on a linear scale is

$$G = 10^{35/10} = 3162$$

The ASE noise is

$$\begin{aligned}\text{ASE noise} &= (3162-1)(1.6)(1)(6.63\times10^{-34})(1.94\times10^{14})(10^{11})\\ &= 65\ \mu\text{W}\end{aligned}$$

13.3 The threshold pumping rate is $W_p^{\text{th}}\tau$ and τ is expressed as

$$\tau = \frac{1}{(1-\beta)W_p^{\text{th}}}$$

W_p^{th} expressed in terms of the threshold light intensity I_p^{th} is

$$W_p^{\text{th}} = \sigma_p \frac{I_p^{\text{th}}}{h\nu}$$

The value of I_p^{th} has to be determined from the pump power. The pump light actually contributing to pumping is $P_p^{\text{th}}\Gamma$ and the corresponding intensity is

$$I_p^{\text{th}} = \frac{\Gamma P_p^{\text{th}}}{\pi r^2}$$

Thus,

$$\begin{aligned}W_p^{\text{th}} &= \sigma_p \frac{\Gamma P_p^{\text{th}}}{h\nu\pi r^2}\\ &= 0.42\times10^{-25}\frac{(0.4)(20\times10^{-3})}{(6.63\times10^{-34})(2.03\times10^{14})\pi(2\times10^{-6})^2}\\ &= 198.7\ \text{s}^{-1}\\ \tau &= \frac{1}{(1-\beta)W_p^{\text{th}}}\\ &= \frac{1}{(1-0.38)(198.7)} = 8.1\ \text{ms}\end{aligned}$$

13.4 **(a)** The result of Example 13.10 is used:

$$\begin{aligned}a &= 2\eta m_t n_{\text{spon}} = 2(1)(1)(2) = 4\\ b &= \tfrac{1}{2}\eta m_t^2 n_{\text{spon}}^2 h\nu\Delta\nu_f(2 - B/\Delta\nu_f)/P_{\max}\\ &= \tfrac{1}{2}(1)(1^2)(2^2)(6.63\times10^{-34})(1.94\times10^{14})(10^{11})\\ &\quad\times(2-10^{11}/10^{11})/10^{-2} = 2.57\times10^{-6}\end{aligned}$$

$$c = \frac{Dl}{10} = \frac{(1000)(0.2)}{10} = 20$$

$$g_1(k) = \frac{a}{b} \cdot \frac{1}{(c \ln 10 - 2k)} = \frac{7.78 \times 10^5}{(23 - k)}$$

$$g_2(k) = 10^{20/k}$$

k	$g_1(k)$	$g_2(k)$	$g_1(k) - g_2(k)$
1	3.5×10^4	10^{20}	Negative
2	3.7×10^4	10^{10}	Negative
3	3.9×10^4	4.6×10^6	Negative
4	4.1×10^4	10^5	Negative
5	4.3×10^4	10^4	Positive
6	4.6×10^4	2.15×10^3	Positive

The value of k that makes $g_1(k) - g_2(k) = 0$ is between 4 and 5, but much closer to 5 amplifiers.

(b) The gain of each amplifier is

$$\frac{lD}{k} = \frac{(0.2)(1000)}{5} = 40 \text{ dB}$$

(c)

$$P_s = 10^{-2} \times 10^{-4}$$
$$= 10^{-6} \text{ W}$$
$$F_k = \frac{\eta}{2}\left[4 + \frac{km_t n_{\text{spon}} h\nu\Delta\nu_f}{P_s}\left(2 - \frac{B}{\Delta\nu_f}\right)\right] km_t n_{\text{spon}}$$

With the given parameters of $m_t = 1$ and $\eta = 1$, F_k becomes

$$F_k = 2kn_{\text{spon}} + k^2 n_{\text{spon}}^2 h\nu\Delta\nu_f / 2P_s$$
$$F_5 = 2(5)(2) + (5^2)(2^2)(6.63 \times 10^{-34})(1.94 \times 10^{14})(10^{11})/(2 \times 10^{-6})$$
$$= 20 + 0.64 = 20.64$$

In the case of $k = 4$,

$$\frac{lD}{k} = \frac{200}{4} \text{ dB}$$
$$P_s = 10^{-7} \text{ W}$$
$$F_4 = 2(4)(2) + (4^2)(2^2)(6.63 \times 10^{-34})(1.94 \times 10^{14})(10^{11})/10^{-7} = 24.23$$

In the case of $k = 6$, $F_6 = 24.4$ and $k = 5$ is indeed the optimum.

Chapter 14

14.1 From Eq. (14.81), the relaxation time is

$$\frac{1}{\gamma} = \tau_n \frac{J_{\text{th}}}{J} = 1.18 \text{ ns}$$

From Eq. (14.39),

$$\frac{1}{\tau_s} = \frac{c}{n}\left(\frac{1}{L}\ln\frac{1}{R} + \alpha\right)$$

where $R = [(n-1)/(n+1)]^2$ for a cleaved surface.

$$\frac{1}{\tau_s} = \frac{3 \times 10^{10}}{3.5}\left[\frac{1}{0.02}\ln\left(\frac{3.5+1}{3.5-1}\right)^2 + 25\right]$$

$$\tau_s = 1.39 \text{ ps}$$

From Eq. (14.89),

$$\omega_r = \sqrt{\frac{1}{\tau_n \tau_s}\frac{J - J_{\text{th}}}{J_{\text{th}}}} = 15.9 \times 10^9 \text{ rad/s}$$

From Eq. (14.88),

$$f_m^r = \frac{1}{2\pi}\omega_r\sqrt{1 - \frac{1}{2}\left(\frac{\gamma}{\omega_r}\right)^2} = 2.53 \text{ GHz}$$

14.2 From Eq. (14.68),

$$bS_s = \frac{S_s}{\tau_s}\frac{1}{(N_{\text{th}} - N_\alpha)}$$

Using Eqs. (14.48) and (14.52),

$$bS_s = \frac{1}{\tau_n}\frac{J - J_{\text{th}}}{J_{\text{th}}}\frac{N_{\text{th}}}{N_{\text{th}} - N_\alpha}$$

From Eq. (14.72), the decay constant γ is

$$\gamma = \frac{1}{\tau_n}\left(\frac{m-1}{1 - N_\alpha/N_{\text{th}}} + 1\right)$$

where $m = J/J_{\text{th}}$.

14.3 **(a)** Turn-on-delay time. From Eq. (14.61),

$$t_d = \tau_n \ln\left(\frac{J}{J - J_{\text{th}}}\right)$$

$$= 3.7 \ln\left(\frac{1.2}{1.2-1}\right) = 6.63 \text{ ns}$$

(b) Cavity length. From Eq. (14.34), $L = \lambda^2/2n\Delta\lambda$ and from Fig. P14.3b,

$$L = \frac{(1.5337)^2}{(2)(3.56)(0.001787)} = 184.9\ \mu\text{m}$$

(c) Threshold electron density. From Eq. (14.30), $R = [(n-1)/(n+1)]^2$. So

$$R = \left(\frac{3.56-1}{3.56+1}\right)^2 = 0.3152$$

From Eq. (14.29),

$$g \geq \frac{1}{L}\ln\frac{1}{R} + \alpha.$$

$$g_{\text{th}} = \frac{1}{0.01849}\ln\frac{1}{0.3152} + 100 = 162.4\ \text{cm}^{-1}$$

From Fig. 14.6, the electron density corresponding to 162.4 cm^{-1} is

$$N_{\text{th}} = 1.8 \times 10^{18}\ \text{cm}^{-3}$$

(d) Threshold current.

$$I_{\text{th}} = J_{\text{th}} L w$$

From Eq. (14.48),

$$I_{\text{th}} = \frac{(0.5 \times 10^{-4})(1.602 \times 10^{-19})(1.8 \times 10^{18})}{3.7 \times 10^{-9}}(0.01849)(4 \times 10^{-4})$$

$$= 29\ \text{mA}$$

(e) Light power versus current. From Eqs. (14.29) and (14.39), $1/\tau_s = v g_{\text{th}}$.

$$\frac{1}{\tau_s} = \frac{3 \times 10^{10}}{3.56}(162.4) = 1.37 \times 10^{12}\ \text{s}^{-1}$$

$$\tau_s = 0.73\ \text{ps}$$

From Eq. (14.55),

$$\tau_c = 1.9\ \text{ps}$$

Using the approximation $E_g \doteqdot h\nu$, we find

$$E_g(J) = \frac{hc}{e\lambda}\ \text{eV}$$

$$E_g = \frac{1.240}{1.5337} = 0.808\ \text{eV}$$

From Eq. (14.58),

$$P_{\text{out}} = (0.808)\left(\frac{0.73}{1.9}\right)(I - I_{\text{th}})$$
$$= 0.31(I - 0.029)\ \text{W}$$

14.4 Figure A14.4 shows the phasor diagram. The diagonal parallelogram *aobc* satisfies

$$2\tilde{E}_s = \tilde{E}_1 + \tilde{E}_3$$

Triangle *aob* satisfies

$$\tilde{E}_a = \tfrac{1}{2}(\tilde{E}_1 - \tilde{E}_3)$$

These are complex forms of Eq. (14.181).

14.5 From Eq. (14.141),

$$\ln I_{\text{th}} = \ln I_{\text{th0}} + \frac{T}{T_0}$$

$\ln I_{\text{th}}$ is plotted against T and the intersect with the vertical axis is found to determine $\ln I_{\text{th0}}$. The slope determines $1/T_0$. They are

$$I_{\text{th0}} = 0.16\ \text{mA}$$
$$T_0 = 50\ \text{K}$$

Similarly,

$$\eta_0 = 1.3\ \text{W/A}$$
$$T_0' = 100\ \text{K}$$

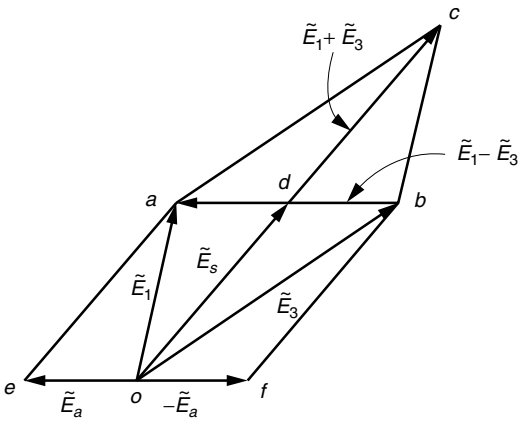


Figure A14.4 Decomposition of the phasor $\tilde{E}_1$ and $\tilde{E}_2$ into $\tilde{E}_s$ and $\tilde{E}_a$.

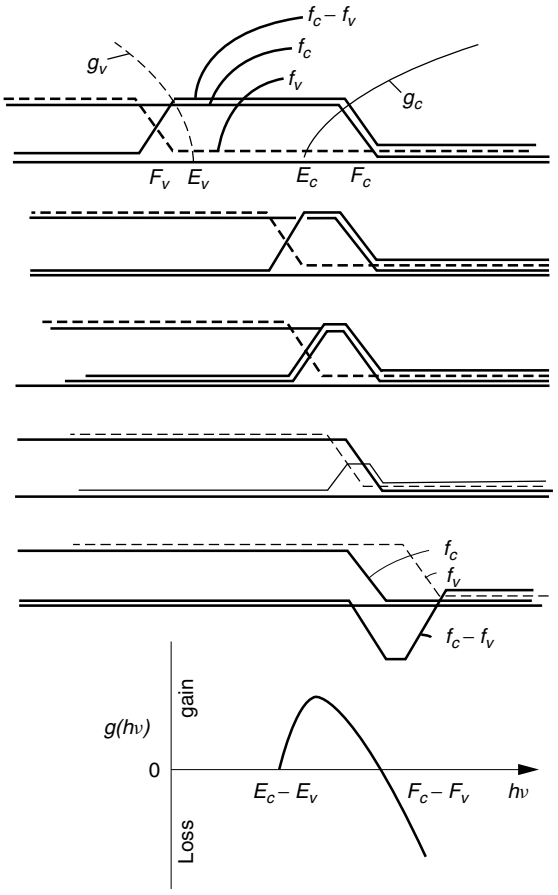


Figure A14.6 Graphical calculation of $g(h\nu)$ (g_v and g_c are suppressed except in the top figure).

Depending on the portions of the experimental curve used, answers may vary. Another way of obtaining η_0 and T_0' is to obtain $\eta = dP/dI$ graphically and then plot it with respect to temperature.

$$\ln\left(\frac{dP}{dI}\right) = \ln\eta_0 - \frac{T}{T_0'}$$

14.6 The Fermi distribution function is represented by a linear ramped curve. The results are shown in Fig. A14.6.

Chapter 15

15.1 Addition of the angles $(K_1 d + \phi)$ and $(-K_1 d + \phi)$ in Eqs. (15.84) and (15.85) gives

$$\tan 2\phi = \frac{K_1(\tilde{\gamma}_0 - \tilde{\gamma}_2)}{K_1^2 + \tilde{\gamma}_0\tilde{\gamma}_2}$$

When $\tilde{\gamma}_0 = \tilde{\gamma}_2$, then

$$\phi = \frac{\pi}{2} q$$
$$q = 0, 1, 2, 3, \ldots$$

In general, $\tilde{\gamma}_0$, $\tilde{\gamma}_2$, and K_1 have to be found first to calculate ϕ.

15.2 Yes, it is possible. Use Fig. 15.12 but with an upside-down pulse shape. You will find that the frequency deviation Δf due to the SPM becomes positive in the leading edge in Fig. 15.12d and the carrier wavelength λ should be chosen at a wavelength in the negative dispersion parameter region ($\lambda < 1.3$ μm).

15.3 From Eq. (15.138),

$$\Delta f_p = -\frac{n_2}{\lambda} L \frac{dI(L, t)}{dt} \tag{1}$$

Inserting

$$I = I_0 e^{-(t/\tau)^2} \tag{2}$$

$$\Delta f_p = \frac{n_2}{\lambda} L I_0 \frac{2t}{\tau^2} e^{-(t/\tau)^2} \tag{3}$$

We need to find the instant at which the value of Δf_p becomes the maximum as well as to rewrite τ in terms of the FWHM τ_s. First, the instant at which Δf_p becomes the maximum is found from

$$\frac{d}{dt}(\Delta f_p) = 0 \tag{4}$$

Inserting Eq. (3) into (4) gives

$$t = \pm \frac{\tau}{\sqrt{2}} \tag{5}$$

Insertion of Eq. (5) with minus sign into (3) gives $\Delta f_{p_{\max}}$ in the left-hand side slope

$$\Delta f_{p_{\max}} = -\frac{n_2}{\lambda} L I_0 \frac{\sqrt{2}}{\tau} e^{-1/2} \tag{6}$$

Next, τ is expressed in terms of the FWHM τ_s.

$$e^{-(\tau_s/2\tau)^2} = \tfrac{1}{2} \tag{7}$$

Solving Eq. (7) for τ gives

$$\tau = \frac{\tau_s}{2\sqrt{\ln 2}} \tag{8}$$

Insertion of Eq. (8) into (6) gives the final result

$$\Delta f_{p_{\max}} = -\frac{n_2}{\lambda} L I_0 \frac{2\sqrt{2\ln 2}}{\tau_s} e^{-1/2} \tag{9}$$

The values of the parameter are inserted

$$\begin{aligned} |\Delta f_{p_{\max}}| &= \frac{3.18 \times 10^{-20}}{1.55 \times 10^{-6}} (200)(1.6 \times 10^{12}) \frac{2\sqrt{(2)(0.693)}}{200 \times 10^{-12}} (0.6065) \\ &= 46.9 \text{ GHz} \end{aligned}$$

The total frequency shifts from both sides of the pulse is

$$2\Delta f_{p_{\max}} = 93.8 \text{ GHz}$$

15.4 Rewriting Eq. (15.232) in practical units gives

$$\begin{aligned} P_s &= \frac{0.775(\lambda \times 10^{-6})^3 (D \times 10^{-12}/[(10^3)(10^{-9})])(A_{\text{eff}} \times 10^{-12})}{\pi^2 (3 \times 10^8)(3.18 \times 10^{-20})(\tau_s \times 10^{-12})^2} \\ &= 0.823 \times 10^{-2} \frac{\lambda^3 AD}{\tau_s^2} \end{aligned}$$

Thus

$$\tau_s \sqrt{P_s} = 0.091 \lambda^{3/2} \sqrt{DA}$$

15.5 **(a)** From Eq. (15.152)

$$|\beta^{(2)}| = \frac{\lambda^2}{2\pi c} |D| \tag{1}$$

In MKS units

$$\begin{aligned} |\beta^{(2)}| &= \frac{(1.55 \times 10^{-6})^2 (17 \times 10^{-12}/10^{3-9})}{2\pi (3 \times 10^8)} \\ &= 2.167 \times 10^{-26} \text{ s}^2/\text{m} \end{aligned} \tag{2}$$

where the dimension of $|\beta^{(2)}|$ is

$$\frac{d^2\beta}{dw^2} = \frac{1/L}{(1/T)^2} = \frac{T^2}{L}$$

Eq. (2) will be converted into practical units

$$\begin{aligned} \beta^{(2)} &= 2.167 \times 10^{-26} \frac{(\text{ps} \times 10^{12})^2}{\text{km} \times 10^{-3}} \\ &= 21.67(\text{ps})^2/\text{km} \end{aligned} \tag{3}$$

(b) From Eq. (15.231)

$$\Gamma = \frac{2\pi n_2}{\lambda A_{\text{eff}}} \tag{4}$$

In MKS units

$$\Gamma = \frac{2\pi(3.18 \times 10^{-20})}{(1.55 \times 10^{-6})(60 \times 10^{-12})} = 2.15 \times 10^{-3}\ \text{W}^{-1}/\text{m}$$

$$\Gamma = 2.15 \times 10^{-3} \frac{1}{10^{-3}}\ \text{W}^{-1}/\text{km}$$

$$= 2.15\ \text{W}^{-1}/\text{km}$$

where the dimension of Γ is

$$\frac{L^2/P}{LL^2} = P^{-1}/L$$

Let us check the value of peak power using Eq. (15.230) rather than Eq. (15.233).

$$P_s = 3.1 \frac{|\beta^{(2)}|}{\Gamma \tau_s^2}$$

In MKS units, P_s is

$$P_s = 3.1 \frac{2.167 \times 10^{-26}}{2.15 \times 10^{-3} \times (20 \times 10^{-12})^2}$$

$$= 78\ \text{mW} \tag{5}$$

The values of Eq. (5) and (15.234) are within the round-off error.

(c) From Eqs. (15.222) and (15.225)

$$z_c = 0.507 \frac{\tau_s^2}{\beta^{(2)}} \tag{6}$$

From (2) and $\tau_s = 20$ ps

$$z_c = 0.507 \frac{(20 \times 10^{-12})^2}{2.167 \times 10^{-26}} = 9.36\ \text{km}$$

It should be pointed out that Eq. (6) can be rewritten using Eq. (1) as

$$z_c = 0.323 \frac{\pi^2 c \tau_s^2}{\lambda^2 |D|} \tag{7}$$

Chapter 16

16.1 The load resistance R_L is

$$R_L = \frac{1}{2\pi CB} = \frac{1}{2\pi(1.6 \times 10^{-12})(2 \times 10^3)}$$
$$= 4.97 \times 10^7 \ \Omega$$

$P_s^a = -23$ dBm $= 5.0 \times 10^{-6}$ W. The receiver system is quantum limited, which can be shown in a variety of ways.

(a) From Eq. (16.55), the receiver is quantum limited if

$$P_s^a > \frac{0.1}{R_L}$$
$$5.0 \times 10^{-6} > \frac{0.1}{4.97 \times 10^7}$$

(b) Being quantum limited can be shown directly from Fig. 16.22.

(c) $N_{\text{shot}} = 2e\dfrac{\eta e}{h\nu} M P_s^a R_L B = 7.95 \times 10^{-14}$ W

$N_{\text{th}} = 4kTB = 3.23 \times 10^{-17}$ W

16.2 The group delay of a step-index fiber is

$$\Delta\tau = \frac{1}{2}\frac{(\text{NA})^2}{cn_1} = \frac{(0.2)^2}{(2)(3 \times 10^5)(1.55)}$$
$$= 4.3 \times 10^{-8} \text{ s/km}$$
$$t_S = \frac{0.35}{B} = \left(\frac{0.35}{4 \times 10^6}\right) = 8.75 \times 10^{-8} \text{ s}$$

The rise-time requirement is

$$t_S^2 = t_T^2 + t_R^2 + t_D^2$$
$$(8.75 \times 10^{-8})^2 = (20 \times 10^{-9})^2 + (4.3 \times 10^{-8}x)^2 + (50 \times 10^{-9})^2$$
$$x = 1.6 \text{ km}$$

16.3 Half the bits are "1" and the other half are "0." The "0" is represented by no light, and "1" is represented by the light being on for only half the bit. Thus, for the RZ code, the light is on an average of one-quarter of the time.

$$\frac{S}{N} = \frac{\frac{1}{4}\left(\dfrac{\eta e}{h\nu} P_s\right)^2}{2e\left(\frac{1}{4}\dfrac{\eta e}{h\nu} P_s\right) B}$$

For the RZ code, $B = B_t$, and P_s is

$$\begin{aligned} P_s &= 2\frac{h\nu}{\eta e}e\frac{\mathrm{S}}{\mathrm{N}}B_t \\ &= 2(\tfrac{1}{0.5})(1.6 \times 10^{-19})(144)(10^8) \\ &= 9.22 \times 10^{-6} \text{ mW} \\ &= -50.4 \text{ dBm} \end{aligned}$$

For BER $= 10^{-10}$, 0.5 dBm has to be added. Finally $P_s = -49.9$ dBm.

16.4 The total system loss is

$$2 \times 20 + 2 = 42 \text{ dB}$$

The light power at the APD in dBm is

$$13 - 42 = -29 \text{ dBm}$$

and in watts it is

$$P_s = 10^{-2.9} \text{ mW} = 1.26 \text{ μW}$$

The peak value of the signal current i_s from the APD is

$$\begin{aligned} i_s &= \left(\frac{\eta e}{h\nu}\right) MP_s \\ &= \left(\frac{(0.75)(1.6 \times 10^{-19})}{(6.63 \times 10^{-34})(3.61 \times 10^{14})}\right)(100)P_s \\ &= 50.1P_s \text{ A} \end{aligned}$$

This implies that the responsivity matches the value given in the spectral response graph at 830 nm in Fig. 16.21.

(a) For $P_s = 1.26$ μW, the peak value of the signal current is

$$i_s = 50.1 \times 1.26 = 63.1 \text{ μA}$$

(b) The load resistance is

$$R_L = \frac{1}{2\pi BC_f} = \frac{1}{2\pi(3 \times 10^8)(3 \times 10^{-12})} = 177 \ \Omega$$

The signal power in the load is

$$\begin{aligned} S &= \overline{[i_s A(t)]^2} R_L \\ &= (63.1 \times 10^{-6})^2 \tfrac{1}{2}(177) \\ &= 3.52 \times 10^{-7} \text{ W} \end{aligned}$$

(c) The shot noise power is

$$N_{\text{shot}} = 2eB(I_s + I_d)M^{2.3}R_L$$

where 0.3 is the typical excess noise index from Fig. 16.21.

$$N_{\text{shot}} = 2(1.6 \times 10^{-19})(3 \times 10^8)(0.631 \times \tfrac{1}{2} + 0.1 \times 10^{-3}) \times 10^{-6}(100^{2.3})(177) = 2.13 \times 10^{-10} \text{ W}$$

Note that if there were no 300-MHz low-pass filter, then 900 MHz from the cutoff frequency of the APD should be used as the value of B.

(d) The thermal noise power is

$$\begin{aligned} N_{\text{th}} &= 4kTB \\ &= 4(1.38 \times 10^{-23})(298)(3 \times 10^8) \\ &= 4.93 \times 10^{-12} \text{ W} \end{aligned}$$

(e) The signal to noise ratio is

$$\begin{aligned} \frac{S}{N} &= \frac{S}{N_{\text{shot}} + N_{\text{th}}} \\ &= \frac{3.52 \times 10^{-7}}{2.13 \times 10^{-10} + 4.9 \times 10^{-12}} = 1615 \end{aligned}$$

The signal to noise ratio in dB is 32.1 dB.

(f) Since $N_{\text{shot}} > N_{\text{th}}$, the system is quantum limited.

(g) If $M = 1$ then $N_{\text{shot}} < N_{\text{th}}$, and the system is thermal noise limited.

(h) The NEP of this APD is

$$\begin{aligned} \text{NEP} &= \frac{h\nu}{\eta e}\left(2eI_d M^x + \frac{4kT}{M^2 R_L}\right)^{1/2} \\ &= \frac{1}{0.501}\left[2(1.6 \times 10^{-19})(0.1 \times 10^{-9})(100^{0.3}) + \frac{4(1.38 \times 10^{-23})(298)}{(100^2)(177)}\right]^{1/2} \\ &= 1.94 \times 10^{-13} \text{ W}/\sqrt{\text{Hz}} \end{aligned}$$

(i) The minimum detectable power of the APD is

$$\begin{aligned} P_{s\,\text{min}} &= \text{NEP} \times \sqrt{B} \\ &= (1.94 \times 10^{-13})\sqrt{300 \times 10^6} \end{aligned}$$

$$= 3.36 \text{ nW}$$
$$= -54.7 \text{ dBm}$$

(j) The minimum detectable light power $P_{s\,\min}^{\text{ASK}}$ for ASK modulation is, from Eqs. (15.43) and (16.50),

$$P_{s\,\min}^{\text{ASK}} = \sqrt{2} P_{s\,\min}$$
$$P_{s\,\min}^{\text{ASK}} = 1.5 - 54.7 = -53.2 \text{ dBm}$$

The allowable system loss between the source and the detector is

$$13 - (-53.2) = 66.2 \text{ dB}$$
$$66.2 = 2x + 2$$
$$x = 32.1 \text{ km}$$

(k) BER $= 10^{-9}$ requires S/N $= 144$. This system is quantum limited and S/N is proportional to P_s as indicated by Eq. (16.38). In order to increase S/N by 144 times that for S/N $= 1$, P_s has to be increased by 144 times:

$$P_s = 144 P_{s\ \min}^{\text{ASK}}$$
$$= (144)(4.75 \times 10^{-9})$$
$$= 6.8 \times 10^{-7} \text{ W}$$
$$= -31.7 \text{ dBm}$$

The allowable system loss between the source and the detector is

$$13 - (-31.7) = 44.7 \text{ dB}$$
$$44.7 = 2x + 2$$
$$x = 21.4 \text{ km}$$

16.5 **(a)**

$$P_s = 1.26 \text{ μW}$$
$$i_s = \left(\frac{\eta e}{h\nu}\right) M P_s$$
$$i_s = \frac{(0.75)(1.6 \times 10^{-19})}{(6.63 \times 10^{-34}) \times (3.61 \times 10^{14})} (60) P_s$$
$$= 30.1 P_s = 38 \text{ μA}$$

(b)

$$S = \overline{[i_s A(t)]^2} R_L$$
$$= (38 \times 10^{-6})^2 \tfrac{1}{2} (50)$$
$$= 36 \text{ nW}$$

(c)

$$\begin{aligned} N_{\text{shot}} &= 2eB(I_s + I_d)M^{2.3}R_L \\ &= 2(1.6 \times 10^{-19})(10^8)[(0.501)(1.26)\tfrac{1}{2} + 1 \times 10^{-3}] \times 10^{-6}(60^{2.3})(50) \\ &= 6.23 \text{ pW} \end{aligned}$$

(d)

$$\begin{aligned} N_{\text{th}} &= 4kTB \\ &= 4(1.38 \times 10^{-23})(298)(10^8) \\ &= 1.65 \text{ pW} \end{aligned}$$

(e)

$$\begin{aligned} \frac{\text{S}}{\text{N}} &= \frac{S}{N_{\text{shot}} + N_{\text{th}}} \\ &= \frac{3.6 \times 10^{-8}}{6.23 \times 10^{-12} + 1.65 \times 10^{-12}} = 4.57 \times 10^3 \end{aligned}$$

The signal to noise ratio in dB is 36.6 dB.

(f) $N_{\text{shot}} > N_{\text{th}}$ and the system is quantum limited.

(g) If $M = 1$, then $N_{\text{shot}} < N_{\text{th}}$, and the system is thermal noise limited.

(h)

$$\begin{aligned} \text{NEP} &= \frac{h\nu}{\eta e}\left(2eI_d M^x + \frac{4kT}{M^2 R_L}\right)^{1/2} \\ &= \frac{1}{0.501}\left[2(1.6 \times 10^{-19})(1 \times 10^{-9})(60^{0.3}) \right. \\ &\quad \left. + \frac{4(1.38 \times 10^{-23})(298)}{(60^2)(50)}\right]^{1/2} \\ &= 2.00\sqrt{1.09 \times 10^{-27} + 9.14 \times 10^{-26}} \\ &= 6.08 \times 10^{-13} \text{ W}/\sqrt{\text{Hz}} \end{aligned}$$

(i)

$$\begin{aligned} P_{s\,\text{min}} &= \text{NEP} \times \sqrt{B} \\ &= 6.08 \times 10^{-13}\sqrt{100 \times 10^6} \\ &= 6.08 \text{ nW} \\ &= -52.2 \text{ dBm} \end{aligned}$$

(j)

$$\begin{aligned} P_{s\,\text{min}}^{\text{ASK}} &= \sqrt{2}P_{s\,\text{min}} \\ &= 1.5 - 52.2 = -50.7 \text{ dBm} \end{aligned}$$

The allowable system loss between source and detector is

$$\begin{aligned} 13 - (-50.7) &= 63.7 \text{ dB} \\ 63.7 &= 2x + 2 \\ x &= 30.9 \text{ km} \end{aligned}$$

(k) Since the system is quantum limited, S/N is proportional to P_s according to Eq. (16.38). The light power P_s needed for S/N = 144 is

$$\begin{aligned} P_s &= 144 P_{s\,\min}^{\text{ASK}} \\ &= 144\sqrt{2} P_{s\,\min} \\ &= 1.24 \times 10^{-6} \text{ W} \\ &= -29.1 \text{ dBm} \end{aligned}$$

The length x of fiber is

$$\begin{aligned} 13 - (-29.1) &= 2x + 2 \\ x &= 20 \text{ km} \end{aligned}$$

16.6

$$\frac{\text{S}}{\text{N}} = \frac{\frac{1}{2}\left(\frac{\eta e}{h\nu}\right)^2 P_s^2 M^2}{2eB(I_s + I_d)M^{2+x} + \dfrac{4kTB}{R_L}}$$

Let

$$\frac{\text{S}}{\text{N}} = \frac{AM^2}{DM^{2+x} + C}$$

$$\frac{d(\text{S/N})}{dM} = \frac{2AM(DM^{2+x} + C) - AM^2(2+x)DM^{x+1}}{(DM^{2+x} + C)^2}$$

For

$$\frac{d(\text{S/N})}{dM} = 0$$

$$M = \left(\frac{2C}{Dx}\right)^{1/(2+x)}$$

Hence,

$$M = \left(\frac{4kTB/R}{eB(I_s + I_d)x}\right)^{1/(2+x)}$$

16.7 Eq. (16.3) can be rewritten as

$$V_O = -\frac{i}{1/R_f + j\omega C/G}$$

The change due to the presence of C_s is

$$R_f \rightarrow \frac{1}{1/R_f + j\omega C_s}$$

$$V_O = \frac{-iR_f}{1 + j\omega(C_s + C/G)R_f}$$

The new cutoff frequency becomes

$$f_{sc} = \frac{1}{2\pi(C_s + C/G)R_f}$$

The ratio of the cutoffs becomes

$$\frac{f_{sc}}{f_c} = \frac{1}{1 + (C_s/C)G}$$

The influence of C_s is amplified by G.

16.8 The probability P_{10} of mistaking "1" as "0" is

$$P_{10} = \frac{1}{\sqrt{2\pi}\sigma_1}\int_{-\infty}^{v_{\text{th}}} e^{-1/2((v-s_1)/\sigma_1)^2}dv$$

Putting

$$-\lambda = \frac{v - s_1}{\sqrt{2}\sigma_1}$$

the integral becomes

$$P_{10} = \frac{1}{\sqrt{\pi}}\int_{\theta_1}^{\infty} e^{-\lambda^2}d\lambda$$

With $\theta_1 = (s_1 - v_{\text{th}})/\sqrt{2}\sigma_1$, P_{10} becomes

$$P_{10} = \tfrac{1}{2}\text{erfc}(\theta_1)$$

Similarly, the probability P_{01} for mistaking "0" as "1" is found as

$$P_{01} = \frac{1}{\sqrt{2\pi}\sigma_0}\int_{v_{\text{th}}}^{\infty} e^{-1/2((v-s_0)/\sigma_0)^2}dv$$

$$= \frac{1}{\sqrt{\pi}}\int_{\theta_0}^{\infty} e^{-\lambda^2}d\lambda$$

$$\theta_0 = \frac{v_{\text{th}} - s_0}{\sqrt{2}\sigma_0}$$

or

$$P_{01} = \tfrac{1}{2}\text{erfc}(\theta_0)$$

For a signal with half "1"s and half "0"s the total probability P of errors is

$$P = \tfrac{1}{2}(P_{10} + P_{01})$$

The value of v_{th} for the minimum P is found from the derivative being zero:

$$\frac{dP}{dv_{\text{th}}} = \frac{dP}{d\theta}\frac{d\theta}{dv_{\text{th}}} = 0$$

Noting

$$\frac{dP}{dv_{\text{th}}} = \frac{1}{2}\left[\frac{dP_{10}}{d\theta_1}\frac{d\theta_1}{dv_{th}} + \frac{dP_{01}}{d\theta_0}\frac{d\theta_0}{dv_{th}}\right]$$

$$\frac{d\theta_1}{dv_{th}} = -\frac{1}{\sqrt{2}\sigma_1}, \quad \frac{d\theta_0}{dv_{th}} = \frac{1}{\sqrt{2}\sigma_0}$$

$$\frac{dP_{10}}{d\theta_1} = -\frac{e^{-\theta_1^2}}{\sqrt{\pi}}, \quad \frac{dP_{01}}{d\theta_0} = -\frac{e^{-\theta_0^2}}{\sqrt{\pi}}$$

we have

$$\frac{dP}{dv_{th}} = \frac{1}{2\sqrt{2\pi}}\left(\frac{1}{\sigma_1}e^{-(s_1 - v_{\text{th}})^2/2\sigma_1^2} - \frac{1}{\sigma_0}e^{-(v_{\text{th}} - s_0)^2/2\sigma_0^2}\right)$$

With $dP/dv_{\text{th}} = 0$, one obtains

$$2\sigma_0^2\sigma_1^2 \ln\left(\frac{\sigma_0}{\sigma_1}\right) = \sigma_0^2(s_1 - v_{\text{th}})^2 - \sigma_1^2(v_{\text{th}} - s_0)^2$$

which is a quadratic equation of v_{th} and the solution can be found.

For the special case of

$$\sigma_0 = \sigma_1 = \sigma$$

the equation becomes simple and $v_{\text{th}} = \frac{1}{2}(s_0 + s_1)$.

16.9 **(a)** As mentioned in Section 16.6.2, the expression for AM is readily obtained by replacing the "on" state power P_s of the ASK by the average power P_s^a in Eq. (16.53). The required power to each heterodyne detector for a given S/N is calculated from Eq. (16.53) with

$$B_{\text{IF}} = 2B, \qquad M^x = 1, \qquad \frac{\eta e}{h\nu} = 0.5$$

$$P_s^a = 4e\left(\frac{\mathrm{S}}{\mathrm{N}}\right)B$$
$$= 6.4 \times 10^{-19}\left(\frac{\mathrm{S}}{\mathrm{N}}\right)B$$

With

$$\frac{\mathrm{S}}{\mathrm{N}} = 1000$$
$$B = 4 \times 10^6$$

P_s is calculated as

$$P_s = 2(4)(1.6 \times 10^{-19})(10^3)(4 \times 10^6)$$
$$= 5.12 \times 10^{-9}\ \mathrm{W}$$
$$= -52.9\ \mathrm{dBm}$$

The total power loss is

	Unit	Quantity	Total
Fiber loss	1 dB/km	10	10
Connector loss	1 dB	2	2
Splicing loss	0.1 dB	5	0.5
Scrambler loss	0.5 dB	1	0.5
Degradation	3 dB	1	3
Margin of safety	5 dB	1	5
		Total	$P_A = 21$ dB

$$P_T = -52.9 + 21 + 10\log 15$$
$$= -20.1\ \mathrm{dBm}$$

where 10 log 15 is due to splitting the power into 15 households.

(b) The fiber rise time is found. The total frequency bandwidth of the signal in the fiber is

$$9(8 \times 10^9) + 2(4 \times 10^6) \doteqdot 7.2 \times 10^{10}\ \mathrm{Hz}$$

The wavelength spread due to the modulation is

$$\Delta\lambda = 1.15\ \mathrm{nm}$$

The spread of the group delay is, from Fig. 15.14,

$$t_F = 17(10 \text{ km})(1.15 \text{ nm}) = 2.0 \times 10^{-10} \text{ s}$$

The system rise time is

$$t_S = \frac{0.35}{7.2 \times 10^{10}} = 4.9 \times 10^{-12} \text{ s}$$

Since $t_F > t_S$, the rise-time requirement cannot be met as long as a $\lambda = 1.55$-μm laser is used.

(c) A laser with $\lambda = 1.3$ μm, where the fiber dispersion is almost zero, has to be used. The fiber rise time is $t_F = 0$, and $t_F^2 + t_D^2 < t_S^2$ is satisfied.

(d) The power requirement must be modified. The fiber at 1.3 μm has a higher loss of 3 dB/km and the total fiber loss has to be increased by 20 dB, and the required peak light power is -6.1 dBm.

INDEX

A (Einstein's *A* coefficient), 842, 900
A mode in a nonlinear layer, 1048
Absorption indicatrix, 400
Acceptor
  action in semiconductors, 1153–1155
  atoms, 1153
  energy level, 1153
Accommodation in eye vision, 93
Acetone $(CH_3)_2CO$, 239
Acoustic wave
  imaging, 95
  surface, 324
Acoustooptic modulator (AOM), 100, 324
  surface acoustic wave (SAW), 324
  used for holographic video displays, 100
Active layer of laser diodes, 904
  gain of, 905
  plasma effect of, 944
Adaptive fiber coupler, 517
ADP, *see* Ammonium dihydrogen phosphate
AFM (atomic force microscope), 160, 161
Ahrens polarizing prisms, 404, 405
Airy pattern, 28
  calculation of, 1172
Alcohol C_2H_5OOH, stimulated Brillouin scattering of, 512
Al_2O_3 (aluminum oxide), 681, 894
$Al_xGa_{1-x}As$ (gallium arsenide doped with aluminum) laser, 895, 949
Alkali metals as cathode materials, 796
Aluminum oxide (Al_2O_3), 687, 894
AM (amplitude modulation), 919, 1086, 1087
Amino acids, optical activity of, 412
Ammonium dihydrogen phosphate (ADP)
  elastooptic properties of, 322
  electrooptic properties of, 305, 316, 1018
Ammonium fluoride (NH_4F), 158
Amplified signal power, 847
Amplified spontaneous emission (ASE) noise, 838, 847, 856, 864, 866, 868, 870
Amplitude distribution
  in coupled slab guides, 646
  in slab optical guides, 615, 616
Amplitude modulation (AM), 919, 1086, 1087
Amplitude modulators, 312, 679, 1018
Amplitude shift keying (ASK), 812, 818, 831, 832, 1089, 1095
Analog modulation, 1083, 1134, 1137
Analytic signal, 526, 547, 1022
  proof of not being applicable to nonlinear cases, 547, 1022
Angle modulation, 1088, 1092
Angular frequency, 1
  convention of, 3, 368
Anisotropic media, 263
Annihilation of negative carriers, 898, 1155
Anomalous dispersion region, 1056, 1066
Antenna radiation pattern, 13, 95
  calculated by Fourier optics, 103, 556
  visualized by microwave holography, 96
Antiglare sheet, 409
  analyzed using Poincaré sphere, 502
Antiglare TV camera, 419
Antireflection (AR) coating, 803, 838, 954
Anti-Stokes radiation, 816
AOM, *see* Acoustooptic modulator
APC (automatic power control), 952
APD detectors, *see* Avalanche photodiode
Aperture functions
  of rectangles, 20
  of triangles, 21
  of circles, 25
  of delta functions, 28
  of shah functions, 30
  of SNOM probes, 114, 158
  numerical (NA), 151, 693, 694, 701, 703, 706, 770
Apodization of radiation patterns, 22
AR (antireflection) coating, 803, 838, 954
Ar_2 (argon) for excimer lasers, 894

Argand diagram, 459
  converted into Poincaré sphere, 469
  custom made, 466
  how to use, 459, 460
  orthogonality between θ and ϵ lines, 465
  ready made, 460
Argon (Ar_2) for excimer lasers, 894
Argon ion (Ar^+) for gas lasers, 894
Array diffraction patterns, 32
  array pattern, 40, 88
  element pattern, 38, 40, 88
  in thin emulsion, 88
  of finite size, 37
  of infinite size, 32
  of irregular spacing, 36
  of regular spacing, 32
Arrayed-waveguide grating (AWG), 673, 674, 1099
Arsenic trisulfide (AsS_3), 330
Artificial ruby for solid state lasers, 894
ASE (amplified spontaneous emission) noise, 838, 847, 856, 864, 866, 868, 870
ASK (amplitude shift keying), 812, 818, 831, 832, 1089, 1095
AsS_3 (arsenic trisulfide), 330
Asymmetric optical guide, 606, 638
Atomic clock, 255
Atomic force microscope (AFM), 160, 161
Atomic polarization, 265
Autocorrelation, 67, 69
  Fourier transform of, 69
Automatic power control (APC), 952
Avalanche effect of APD, 803, 1103, 1159
  multiplication factor, 804
Avalanche photodiode (APD), 796, 801, 803, 1085, 1159
  excess noise of, 1117
  multiplication factor M of, 1118
  noise equivalent power (NEP) of, 1117
  optimizing gain of, 1148
  structure of, 802
AWG (arrayed-waveguide grating), 673, 674, 1099
Axicon, 245
Axis image, 245

b (normalized propagation parameter), 728
B (Einstein's B coefficient), 843, 899
B mode in nonlinear guides, 1048
β matching, 113
Babinet compensator, 380
Back focal plane, 62, 66, 80
Back projection, 469
Backward pumping of EDFA, 864
Balanced mixer, 814, 1122
$Ba_2NaNb_5O_{15}$ (BNN), 1018
Bandgap energy (energy gap), 804, 901, 925, 945, 996, 998
Band-pass optical filter, 747
Band-stop optical filter, 745
Barium titanate ($BaTiO_3$)
  photorefractive effect, 332, 337, 511, 517, 521
  electrooptic properties, 306, 316
Barricade of current by means of a back-biased p-n junction layer, 935
Baseband signal, 1085, 1086
Baseline wander of receivers, 1096
$BaTiO_3$, *see* Barium titanate
BC (buried crescent) laser, 934, 936, 942
BCl_3 (boron chloride), 777
Beam pattern of laser diodes, 946
Beam propagation factor m^2 of Gaussian beam, 220
Beam radius of Guassian beams, 209
Beam splitter
  nonpolarizing, 411
  polarizing, 402, 412
Beam waist of Gaussian beams, 208
Beamwidth of laser beams, 948
Beat noise
  signal-spontaneous, 870, 871
  spontaneous-spontaneous, 870, 871
BeF_2, 708
Bend-induced birefringence in optical fibers, 382
  used for polarizer, 404
  used for retarder, 382
Bends in an optical guide for integrated optics, 131
Benzene (C_6H_6), Kerr constant of, 316, 317, 1017
BER (bit error rate), 1123, 1124, 1134, 1148
Bessel beams, 237, 242, 245, 259
  one dimensional model of, 239
Bessel function
  modified, 714
  of approximate form, 723
  of the first kind of nth order, 17, 27, 243, 713
  of the second kind of nth order, 713
  recurrence formula of, 723
  relationship of, 719
BGO ($Bi_{12}GeO_{20}$) (bismuth germanium oxide), photorefractive effect of, 332
Bias voltage, 899
Biaxial crystal, 267, 300
Bidirectional pumping of EDFA, 864
$Bi_{12}GeO_{20}$ (bismuth germanium oxide) (BGO), 332
$Bi_{12}SiO_{20}$ (bismuth silicon oxide) (BSO), 332
Binary coding, 1095, 1096
Binocular parallax, 92
Binomial expansion, 11, 729
Biometric indicators for identification, 77
Birefringence, 95, 263, 821
  bent-induced in fibers, 382, 404
  circular, 413
  Jones matrix of, 424
  linear, 413
  materials, 365, 412
  of nematic liquid crystals, 822
  of polarization-maintaining fibers, 707, 821

Birefringence fiber polarizer, 404
Birefringence polarizer or polarizing prism, 396, 402
Bismuth germanium oxide (BGO) ($Bi_{12}GeO_{20}$), photorefractive effect of, 332
Bismuth silicon oxide (BSO) ($Bi_{12}SiO_{20}$), photorefractive effect of, 332, 511
Bit error rate (BER), 1123, 1124, 1134, 1148
  versus signal to noise ratio, 1123
Bit rate, 1073, 1081, 1097, 1102, 1139
Blackbody radiation, 843
BNN ($Ba_2NaNb_5O_{15}$), 1018
B_2O_3 (boric oxide) to lower the melting point of glass, 706
Bolometer, 800
Bohr's approximation, 529
Boltzman constant, 816, 876, 926, 1114
Boltzman distribution function, 816, 1157
Boltzman population ratio, 849
Borderline between the quantum-limited and thermal-noise-limited S/N, 1123
Boric oxide (B_2O_3) to lower the melting point of glass, 706
Boron chloride (BCl_3), 777
Boundary, 110
  of anisotropic media, 292
  of isotropic media, 110
  of nonlinear media, 1030
Boundary conditions, 110
  Maxwell's, 112
  periodic, 1161, 1163
  standing-wave, 1161, 1163
  wavelength matching, 113
Bow-tie fiber, 707
Bragg condition, 959
Bragg frequency, 961, 963
Bragg grating method of dispersion compensation, 755
Bragg grating sensor of photoimprinted fibers, 742, 745, 747
Bragg reflection, 89
Bragg reflector, 962, 970, 971, 973
Breakdown voltage of air, 803
Brewster angle, 127, 407, 566
Brewster window, 129, 408
Brillouin scattering, 512, 602, 816, 817
BSO ($Bi_{12}SiO_{20}$) (bithmus silicon oxide), photorefractive effect of, 332, 511
Buffer-covered fiber, 784
Buffered metal guide, 672
Bulge guide, 672
Bulge of field, 1044
Bulk waves, 144, 681
Buried crescent (BC) laser, 934, 936, 942
Burrus-type LED, 1008

Cabling of optical fibers, 783
$CaCO_3$ (calcite), 277, 290, 402
Cadmium (Cd), 936
Cadmium selenide (CdSe), of photocell, 800
Cadmium sulfide (CdS), electrooptic properties of, 306, 800
Cadmium telluride (CdTe), electrooptic properties of, 306
Calcite crystal (Iceland spar) ($CaCO_3$), 277, 290, 402
  physical constants of, 290
Calcium oxide (CaO) to lower the melting point of glass, 706
Calorimetric methods of detecting light, 800
Canada balsam cement, 402, 404
CaO (calcium oxide) to lower melting point of glass, 706
Capacity for information transmission through optical fibers, 703
Capping layer, 803
Carbon dioxide (CO_2) laser, 512, 894
Carbon disulfide (CS_2)
  Kerr constant of, 317
  Kerr effect of, 1017
Carbon dioxide (CO_2)
  laser, 894
  stimulated Brillouin scattering effect of, 512
Carbon tetrachloride (CCl_4), Kerr constant of, 316, 317
Card for encryption, 80, 108, 560
Carrier, 841
  concentration of, 930
  confinement of, 930
  negative and positive, 898
  number of, 841
  population inversion of, 847
  rate of transition of, 841
Carrier electrons, 1151
Carrier frequency, 1085
Cassegrain telescope, 107, 108
Cavity resonance frequencies of laser diodes, 960
Cavity, *see* Fabry-Pérot cavity
c axis of crystals, 267
C_{60} (fullerene) doped liquid crystal, 357
CCl_4 (carbon tetrachloride), 317
Cd (cadmium), 936
CDH (constricted double heterojunction) laser, 937
CDM (code division multiplexing), 1085, 1094
CdS (cadmium sulfide), 306, 800
CdSe (cadmium selenide), 800
CdTe (cadmium telluride), 306
CH_4 (gaseous methane), 512
C_6H_6 (benzene), 316, 317, 1017
$C_{10}H_{16}$ (turpentine), optically active, 412
$(CH_3)_2CO$ (acetone), 239
C_6H_4 $(CH_3)NO_2$ (nitrotoluene), 1017
$C_6H_5NO_2$ (nitrobenzene), 316, 317, 1017, 1035
$C_{12}H_{22}O_{11}$ (natural sugar), 414
C_2H_5OOH (alcohol), 512
Chain of optical amplifiers, 882

Channeled substrate planar (CSP) laser, 936, 943
Characteristic equations, 605
  for EH modes of optical fibers, 723
  for HE modes of optical fibers, 723
  for even TE modes, 622
  for even TM modes, 610, 628, 632, 634
  for odd TE modes, 622
  for odd TM modes, 610, 612, 628, 633, 636
  for TE modes of optical fibers, 719
  for TM modes of optical fibers, 719
  of coupled slab guides, 643
  of optical fibers, 717, 718
  of rectangular guides, 657
Characteristic temperatures
  of the external quantum efficiency, 951
  of the threshold current, 951
Characteristic wave impedance, 124
Chemical lasers, 893, 895
Chemical vapor deposition (CVD), 777
$\chi^{(1)}$, 1018
$\chi^{(2)}$, 523, 525, 1018
$\chi^{(3)}$, 523, 527, 1018
Chiral, 355
  smectic C liquid crystal, 354, 356
Chirp noise, 955
Chirped aperiodic Bragg grating, 748
  photoimprinted in fiber cores, 755
Cholesteric type liquid crystal, 341, 342
  as an optically active substance, 412
Chromium doped gallium arsenide (GaAs:Cr), photorefractive effect of, 512
Chromium oxide (CrO_3), 894
Cinnabar (HgS)
  as an optically active substance, 412
  electrooptic properties of, 305
Circle diagrams for polarization, 365
Circle function, 25
  Fourier transform of, 25
Circular apertures in random spacing, 37
Circular birefringence, 413
Circularly polarized wave, 364
Circularly polarizing sheets, 409
  for antiglare, 409
Cladding of fibers, 382, 694
Cladding layer of optical slab guides, 606
Clamped by the Boltzmann distribution, 855
Cleaved surfaces, 1013
CO_2 (carbon dioxide) laser, 512, 894
Code division multiplexing (CDM), 1085, 1094
Code format
  Manchester, 1097
  nonreturn to zero (NRZ), 774, 1096, 1140
  return to zero (RZ), 774, 1096, 1141
Coefficient matrix method, 606, 622, 623
Coherent communication systems, 807
  jitter in, 819
Coherent detection, 807, 812, 970, 1094
  immune to both polarization and phase jitter, 826
Collision of solitons, 1070
Color holograms, 87, 92
Comets' tails, 249
Comb function, 30
  Fourier transform of, 31
Commensurate, 528
Compensator, 380, 494
  Babinet compensator, 380
  Soleil compensator, 382
Complementary error function, 1126
  approximate expression of, 1127
Complementary mode patterns, 735
Complex fields, 452
Complex (phase) conjugate waves, 507
Complex refractive index of silver, 672
Component field ratio in the complex plane, 452
Component waves, 431, 615
Computer-controlled method of jitter control, 820
Computer generated holograms, 97, 98
Concave end mirrors of gas lasers, 893
Conduction band, 898, 988, 998, 1151
Confined light in the lateral (horizontal) direction, in laser diodes, 937
Confined light in the vertical (transverse) direction, 937
Confinement factor, 861, 866, 994
Confinement of injection current, 933
Conflection, 666
Conflection lens, 667
Confocal cavity, 236
Confocal reflector UV light cavity, 775, 783
Conjugate waves, 504, 507, 512, 521
Conical lens (axicon), 245, 246
Connectors
  fiber, 790
  loss, 790, 1130, 1131, 1138, 1144, 1146
Conservation of energy in elliptical polarization, 437
Constant azimuth θ and ellipticity ϵ lines of Argand diagrams, 455
Constricted double heterojunction (CDH) laser, 937
Contact layer in electrode lamination, 680
Contact potential of a junction, 932, 1155
Continuous wavelength tuning by combining I_p and I_b, 975
Conventions for expressing time dependence, 3
Conversion of joules into electron volts, 997
Conversion retarder of TM–TE converters, 499
Conversion of state of polarization
  from circular to linear, 484
  from elliptic to linear, 389
  from linear to circular, 386
  from unknown to linear, 389
Convolution, 14, 68, 69
Copper chloride (CuCl), electrooptic properties of, 306
Core-cladding interface of optical fibers, 694
Core glass of optical fibers, 694

Core layer of slab guides, 606
Correction factor of Gaussian beams, 210
Correlation operation
  autocorrelation, 67, 69
  countermeasures for differences in scale, 74
  cross-correlation, 68, 69
Corrugated reflectors, 957, 959, 964
Costas PLL (phase locked loop), 809
Cotton–Mouton effect, 327
Cotton type prism, 404
Coulomb field, 1152
Coulombic force, 341
Coupled guides, 643, 665
Coupled wave equations, 526
Coupler (tunable), 740
Coupling between rectangular guides, 664
Coupling circuits to preamplifiers, 1106, 1110
Coupling length, 664, 666, 740
Coupling loss, 1130, 1138, 1144, 1146
Coupling mechanism of slab guide couplers, 651
Critical angle
  anisotropic media, 285, 296, 575, 576
  isotropic media, 131, 142, 694
CrO_3 (chromium oxide), 894
Crooks, Sir William, 259
Cross-correlation, 68, 69, 72
Cross-sectional ellipse, 307, 313, 399, 580
Crosstalk between channels, 859
Cryptograph, 78
Crystal(s)
  with inversion symmetry, 523, 599
  without second order nonlinearity, 523, 525
  under various external fields, 302
Crystal axis
  c axis, 267
  crystallographic axis, 307
  z axis, 307
Crystal cut, 307
Crystallography, 13, 307
CS_2 (carbon disulphide), 239, 316, 1017
CSP (channeled substrate planer) laser diode, 934, 936, 943
CuCl (copper chloride), 306
Cutoff
  condition in fibers, 719
    EH modes, 723, 724
    HE modes, 723, 724
    TE mode, 719
    TM mode, 720
  condition in planar optical guides, 612
    of TM modes, 612, 613
    of buried optical guides, 661
CVD (chemical vapor deposition), 777
Cylindrical-coordinates Fourier transform, 16

D-cut, 757
D line of sodium spectrum, 414
D modes in nonlinear guides, 1048
d-rotary (right-handed) quartz, 412, 415
Dark current of photodiodes, 1103
Dark light soliton, 1077
DBR (distributed Bragg reflection) laser, 957, 963
DC-PBH (double-channel planar buried heterostructure) laser, 932, 934, 936, 942, 1134
De Broglie, Louis, 985
DEks plane, 278
Decomposition of elliptically polarized light, 375
Decrypting, 80, 108, 560
Degenerate modes, 768
Delay-and-detect circuit, 825, 826, 827
$\delta(x)$ (delta function), 28
  convolution of, 29
  Fourier transform of, 29
  train of, 30
Demountable fiber connectors, 786
Denisyuk hologram, 92. *See also* Lippman hologram
Density $m(\nu)$ of modes per unit frequency per unit volume of blackbody radiators, 843, 847, 1160
Density of states, 898, 987, 988
Depletion region, 1155
Depolarized (unpolarized) wave, 364
Depth of focus of Gaussian beams, 213, 218
Designing fiber-optic communication systems, 1129
Detection of light by stimulated Brillouin scattering effects, 815, 830
Devices for integrated optics, 655
Dextrorotary (d-rotary) optical activity, 412, 415
D-fiber, 708, 744, 757, 866
DFB (distributed feedback) grating, nonlinear, 1019
DFB (distributed feedback) laser, 896, 958, 961, 963, 970, 977
Di Francia, Toralds, 666
Dichoric mirror, 167, 197, 865
Dichoric polarizer, 394
Dielectric constant, 112
  absolute, 112
  nonlinear, 1022
  relative, 112, 266
  relationship with refractive index, 112, 266
Differential phase shift keying (DPSK), 824, 825, 828
Diffraction-free beams, 237
Diffraction from array, 32
  of finite dimension, 37
  of infinite dimension, 32, 36
Diffraction limited, 45
Diffraction patterns, 9, 1162
  far field, 11
  near field, 13
Diffraction unlimited images, 150
Diffusion constant, 1158
Diffusion equation, 1157
Diffusion length
  of electrons, 932
  of holes, 932

Diffusion potential, 932
Digital modulation, 1083
Digital system 1 (DS-1), 1101
Digital system design, 1144
Digital video disk (DVD), 161
Diode laser, *see* Laser diodes
Diode pumped solid-state laser, 967
Dipole antenna, 102, 103, 556, 557
Dipole–dipole interations, 341
Dipole moment, 265
  permanent, 265, 266
Dirac delta function ($\delta(x)$), *see* $\delta(x)$ (delta function)
Direct detection, 796, 805, 806, 830, 1086
Direction of the wavenormal, 761, 762
Directional couplers, 605, 1018
Discotic liquid crystals, 341, 344
Dislocation spots caused by γ rays, 697
Dispersion, 699, 1086
  anomalous, 1056, 1066
  material, 699, 701, 752, 1086
  mode, 699, 696, 703, 769
  normal, 1056
  waveguide, 699, 701, 752, 1086
Dispersion compensator, 741, 755, 759
Dispersion countermeasures, 1051
Dispersion equation, 637, 638
Dispersion parameter, 749, 769, 1055, 1072
  definition of, 749
  of graded-index fibers, 768, 769
Dispersion-shifted fibers, 703, 705, 748, 749, 755, 769, 1051
Distortion of light pulses during transmission, 635, 709, 838
Distributed amplifier approach, 1111
Distributed Bragg reflection (DBR) laser, 957, 963
Distributed element preamplifier, 1106
Distributed feedback (DFB) grating, 1019
Distributed feedback (DFB) laser, 896, 958, 961, 963, 970, 977
Distributed FET amplifiers, 1111
Divergence of Gaussian beams, 212, 218
DMOAP (*N*,*N*-dimethyl-*N*-octadecyl-3-aminopropyltrimethoxysilyl chloride), 353
DNA, 77, 160, 253
Doppler cooling, 255
Doppler shift, 254
Dominant mode, 612, 724
Donor atoms, 1152
Dopant-diffused channel, 936
Double balanced mixer, 814
Double heterojunction, 930
Double sideband (DSB) modulation, 1086, 1087
Double sideband suppressed carrier (DSBC) modulation, 1087
Double-channel planar buried heterostructure (DC-PBH) laser, 934, 936, 942, 1134
Double-clad fiber, 866
Double-crucible method, 775
DPSK (differential phase shift keying), 824, 825, 828
Drawing into optical fibers, 782
Drift of the operating point, 1092
Drift force, 334
DS-1 (digital system 1), 1101
DS-4 (digital system 4) PCM, 1145
DSB (double sideband) modulation, 1086, 1087
DSBC (double sideband suppressed carrier) modulation, 1087
D-shaped inner cladding, 866
Dual-mode fiber, 739, 758
  dispersion compensator, 741, 757
  tunable coupler, 740
DVD (digital video disk), 161
Dye-doped liquid crystals, 357
Dye lasers, 893, 894
Dynodes, 796, 797

E^x_{pq}, 656, 657
E^y_{pq}, 656
EASLM (electrically addressed spatial light modulator), 350
ECB (electrically controlled birefringence), 345, 357
E mode
  in linear optical fibers, 722
  in nonlinear optical guides, 1043, 1048
EDFA, *see* Erbium-doped fiber amplitude
Edge-emitting LED, *see* LED
Edge enhancement of images, 63
Edible laser, 896
Effective core area, 1072
Effective densities of states, 926
Effective index method for rectangular optical guides, 661
Effective index of refraction *N*, 619, 659
  of e-waves, 279, 280, 282, 290, 294
Effective masses, 925, 926
  of electrons, 898
  of holes, 899
EH modes, 722, 731, 732
$EH_{\mu\nu}$, 732
Eigenvalue equation of *M*, 428, 433
  eigenvalue of, 429, 433
  eigenvector of, 429, 430
  orthogonal eigenvectors of, 429, 430
Eikonal, 760
Eikonal equation, 759, 760
Einstein's *A* coefficient, 842, 900
Einstein's *B* coefficient, 843, 899
Elastooptic effect, 304, 317, 324
  due to principal strain, 318
  due to shearing strain, 318
  table of elastooptic constants, 320
Electric flux density, 112
Electric noise power of EDFA, 868

Electrical S/N of output from PIN photodiodes, 876
Electrically addressed spatial light modulator (EASLM), 350
Electrically controlled birefringence (ECB) cell, 345, 357
Electrode configurations of integrated optics, 680, 681
Electron-beam lithography, 98
Electron density, 898
Electron mass, 924
Electron microscope, 161
Electrooptic effect, 304, 1017
  Kerr, 316
  material, 605
  table of, 305
Electrooptic material, 605
Electrooptic modulator, 1086
  of amplitude, 1086
  of phase, 1096
Electrooptic probe, 313
Element fibers of fiber cables, 784
Element patterns of diffraction, 40, 88
Eliminating wave front distortion by phase conjugate waves, 508
Ellipse or hyperbola, 434
Ellipsoid, 286
Elliptical core fiber, 707
Elliptically polarized wave, 364, 819
  azimuth of major axes, 431
  converted from linearly polarized wave, 367
  converted into linearly polarized wave, 389
Ellipticity, 431, 436
Embedded guide, 671
Emission
  spontaneous, 842
  stimulated, 842
Encryption scheme, 80, 108, 560
  encrypted card, 80, 108, 560
Energy gap, 804, 901, 925, 945, 996, 998
Energy states, 985
  in bulk semiconductors, 985
  in quantum wells, 988
Energy transfer in photorefractive crystals, 335
Enhanced photorefractive effect by external electric fields, 334
Envelope function of solitons, 1059
  dynamics of, 1064
Equivalent circuit of PIN photodiodes, 1102
Er^{+3} (erbium), 844, 845, 848, 853, 854, 858, 891
Erbium chloride ($ErCl_3$), 837
Erbium-doped fiber amplifier (EDFA)
  ASE (amplified spontaneous emission), noise of, 847
  basics of, 834
  chain of, 882
  compared to SLA, 838
  cross-talk of, 859
  electric noise power of, 868
  ESA (excited state absorption) of, 855
  gain of, 838, 852, 901
  noise figure of, 880
  pump light of, 853, 864
  rate equations of, 848
  saturation of, 856
  spectral line shape of, 196, 839
$ErCl_3$ (erbium chloride), 837
ESA (excited state absorption), 855, 889
Etalon, 167, 179
Evanescent field, 58, 134, 139, 142
  due to total internal reflection, 135
  effective depth of penetration, 135
  generating, 147
  graphical solution of, 145
  microscope by, 150
  probes to detect, 154
Evanescent wave, *see* Evanescent field
Even TM modes in planar optical guides, 608, 609, 610, 611
e-wave (extraordinary wave), 269, 272, 273, 365
Excess noise in an APD, 1117
Excess noise index, 1117
Excimer laser, 894, 893
Excited dimer, 894
Excited state absorption (ESA), 855, 889
Exposure meter, 800
External quantum efficiency, 951
External terminal potential of laser diodes, 928
Extinction ratio, 916
Extraordinary wave (e-wave), 269, 272, 273, 365
"eyes" of Panda fiber, 407

F (fluorine), 895
Fabrication of optical fibers, 775
  by one-stage process, 775
  by two-stage process, 777
Fabrication of preform rods, 775, 777
Fabry–Pérot cavity, 166, 568
  antiresonance of, 174
  arbitrary angle of incidence of light, 170
  determining the spacing of reflectors, 202, 203
  finesse of, 194, 195, 196, 197
  free spectral range, 171
  modes, 232, 233
  resolving power, 192
  resonance, 174
  stability diagram, 231
  superstructure grating, 978
  tunable fiber, 195
Fabry–Pérot dichroic filter, 197
Fabry–Pérot (FP) type laser, 904
Fabry–Pérot type fiber laser, 1005
Far field region, 11
  divergence of Gaussian beams, 212
  divergence of a Gaussian beam after a lens, 218
  Fraunhofer region, 11
Faraday effect, 326, 327, 329, 415

Faraday rotator mirror (FRM), 1005, 1007
Fast axis of anistropic media, 367
FDM (frequency division multiplexing), 1085, 1094, 1099
$FeCl_3$ (ferric chloride), 777
Fe:$LiNbO_3$ (Fe doped lithium niobate), 91
Fermat's principle, 666
Fermi–Dirac distribution function, 898, 899, 992, 1151, 1157
Fermi–Dirac energy level, 898, 1151, 1152
Fermi level gap, 925
Ferric chloride ($FeCl_3$), 777
Ferroelectric liquid crystals, 354
FETs (field effect transistors), 1111
FIB (focused ion beam), 158
Fibers. *See also* Optical fibers
  side pit, 707
  side tunnel, 707
  with internal Bragg gratings, 744, 755
Fiber amplifier, types of, 836
Fiber Bragg gratings, 741, 1010
Fiber connectors, 790
Fiber coupler, adaptive, 517
Fiber Fabry–Pérot filter, 195
Fiber lasers, erbium-doped, 1004
Fiber-loop polarization controller (FPC), 824
Fiber-loop retarders, 382
  design of, 383
  how to use, 385
  $\lambda/2$ and $\lambda/4$ of, 385, 386
Fiber optic communication, 1081–1150
  amplifiers for, 1085
  coherent detection system, 807
  design of, 1129
  homodyne detection, 1143
  modulated light for, 1082, 1085
  multiplexing, 1097
  overview of, 1082
  receivers, 1085
  soliton, 1073
  space optic, 105
Fiber-optic gyros, 1012
Fiber Raman lasers, 1009
Fiber sensing, 747
50-Ω circuit with mismatch in coupling to a photodetector, 1106
Field effect transistor (FET), 1111
Film layer as core layer, 606
Finesse, 194, 195, 197
  effective, 194
  flatness, 194
  reflection, 194
Fingerprint detection, 77, 103, 105, 521
First-order Bessel functions of the second kind, 714
FLAG (fiber-optic link around the globe), 693
Flame brushing method to enhance photosensitivity, 743
Flatness finesse, 194
Fleming, Sir John A., 396
Fluorescent glow, 834, 835
Fluorinated alkyl methacrylate copolymer for plastic fibers, 706
Fluorine (F), 895
Fluorozirconate glass (ZBLAN) fiber for upconversion fiber amplifiers, 889
FM noise, 955
FM (frequency modulation) signal, 1088, 1092, 1094
Focal beam parameter (Rayleigh range of a Gaussian beam), 209
Focal cavity, 235, 236
Focal length, 43
Focal plane, 47
Focused ion beam (FIB), 158
Fold-unfold-fold (FUF) law, 667
Forward four-wave mixing, 533, 534, 537
Forward pumping of EDFA, 864
Forward pumping versus backward pumping, 864
Fourier–Bessel transform, 18, 27
Fourier–Hankel transform, 18, 561
Fourier optics, 1
Fourier transformable property of convex lenses, 46
Fourier transform in cylindrical coordinates, 16, 18, 243
  Fourier–Bessel transform, 18, 27
  Fourier–Hankel transform, 18, 561
Fourier transform in rectangular coordinates, 16, 20
  of $\delta(x)$, 28
  of derivatives, 64
  of Fourier transforms, 69
  of logarithms, 76
  of the point spread function, 14
  of shah functions, 32
  of sinusoidal functions, 71, 103, 556
Fourier transform spectroscopy, 166
Four-level model of fiber amplifiers, 836
Fourth order nonlinearity, 541, 599
Four-wave mixing (FWM), 506, 1051
  forward, 533, 534, 537
  in a general direction, 526
FP (Fabry–Pérot) type laser, 904
FPC (fiber-loop polarization controller), 824
Fraunhofer diffraction pattern, 11
Fraunhofer region, 11
Free electrons, 1151
Free space optical communication link, 258, 1085
  immune to turbulence, 516
Free spectral range (FSR), 179, 182
  in terms of frequency, 181, 235
  in terms of wavelength, 179, 181
Frequency chirp, 254
Frequency discriminator, 1094
Frequency division multiplexing (FDM), 1085, 1094, 1099
Frequency modulation (FM), 1088, 1092, 1094
Frequency response of PIN diodes, 1104

Frequency shift keying (FSK), 1089, 1095
Fresnel field, 13
Fresnel–Kirchhoff diffraction formula, 11, 58, 243, 545
Fresnel–Kirchhoff integral, 56, 544
Fresnel reflection equation, 121
Fresnel region, 13
Fresnel rhomb, 137
Fresnel's explanation of optical activity, 413
Front focal plane, 47, 71
FRM (Faraday rotator mirror), 1005, 1007
FSK (frequency shift keying), 1089, 1095
FSR, *see* Free spectral range
FUF (fold-unfold-fold) law, 667
Full width at half maximum (FWHM), 197, 1071
Fullerene (C_{60}) doped liquid crystals, 357
Full waveplate, 385
Fundamental soliton, 1069, 1071
Fused silica (SiO_2)
  elastooptic properties of, 320
  Kerr effect of, 1017
Fusen, 485
FWHM (full width at half maximum), 197, 1071
FWM, *see* Four-wave mixing

GaAs, *see* Gallium arsenide
GaAs:Cr (chromium doped gallium arsenide), 512
GaAs MESFET, 1111
$Ga_xIn_{1-x}As_yP_{1-y}$ laser, 945
Gabor-type hologram, 81
Gain compression, 838
Gain constant
  with respect to distance, 910
  with respect to time, 910
Gain guiding, 937, 939, 942
Gain of optical amplifiers, 838, 850, 852, 895
Gain spectrum
  of laser diodes, 902, 904, 923, 928
  of MQW lasers, 992, 995, 1004
Gallium arsenide (GaAs)
  elastooptic properties of, 320
  electrooptic properties of, 306
  photorefractive properties of, 332
Galvanometric scanner, 101
Gamma-ray (γ-ray) irradiation to increase fiber loss, 697, 698
Gaseous methane CH_4, stimulated Brillouin scattering, 512
Gas lasers, 893
Gas pipes for keeping moisture off, 786
Gaussian beam, 205
  amplitude distribution, 211
  beam waist, 208, 216, 217
  cross-sectional area of, 209
  depth of focus, 213
  in spherical mirror cavities, 227
  intensity distribution, 211
  location of waist, 208, 217
  q parameter of, 207, 209, 213, 214, 215, 216
  Rayleigh range of, 209, 218
  transformation, 214, 259
  with higher order modes, 223, 233, 235
  with the fundamental mode, 206
Gaussian distribution, 1124
Gaussian lens formula, 43, 44, 219
General guides, 625
Generalized wave equation, 269
GeO_2 (germanium dioxide), 697, 705, 706,
GeO_2 (germanium dioxide)-doped core, 694, 697
Geological surveys, 97
Geometrical optics theory, 605, 759
Germanium dioxide (GeO_2), 694, 697, 705, 706
GH (guest-host) liquid crystal cell, 353
Glan–Air polarizing prism, 402, 404
Glan–Faucault prism, 402, 404
Glan–Thompsom polarizing prism, 404
Glass, photoreflective media, 512
Goos–Hänchen shift, 141, 142
Graded-index fiber, 705, 759, 762, 768, 769, 770
Graded index (GRIN) lens, 330
Gradient, 288
Graphical solution
  of evanescent waves, 145
  of polarization states, 365
  of propagation in an uniaxial crystals, 269, 270
Grating
  arrayed waveguide, 673, 676, 1099
  photoinduced Bragg, 741, 747, 755
  reflection from Bragg, 957
  scattering from Bragg, 963
  transcendental equations, 611
Grating spectroscopy, 166
Gravitational acceleration, 572
Gravitational wave, 257
GRIN (graded index) lenses, 330
Group delay in optical fibers, 699, 749, 759, 1055
Group index, 750, 768
Group velocity, 748, 1055
Group velocity dispersion (GVD), 1053, 1055, 1057, 1058, 1069
Guard ring, 803
Guest-host (GH) liquid crystal cell, 353
Guide, slab optical, *see* Optical slab guides
Guided nonlinear boundary wave, 1030
Guided waves, 681, 1031
GVD (group velocity dispersion), 1053, 1055, 1057, 1058, 1069

H mode, 722
Half-mirror (HM), 65, 505, 1005
Half-wave fiber loop, 383
Half-waveplate ($\lambda/2$ plate), 383, 385
  how to use, 385
  polyimide, 677
Handedness of circular polarization, 364, 374, 439

HBT (heterojunction bipolar transistor), 1111
HE modes, 722, 731, 735, 736
  cutoffs, 723
HE_{11} mode, 732
$HE_{\nu\mu}$, 731, 735, 736
Heavyside's step function, 25, 992
HeCd (helium-cadmium) laser for blue light, 101
Helium-neon (HeNe) laser for red light, 101, 893
HEMT (high electron mobility transistor), 1111
HeNe (helium-neon) laser for red light, 101, 893
Hermite Gaussian beam (higher order modes), 223, 226
Hermite polynomial of order n, 224
Heterodyne detection, 807, 830
Heterojunction, 930
Heterojunction biopolar transistor (HBT), 1111
HF (holey fiber), 706
HF (hydrofluoric acid), 158
HgS (cinnabar), optically active, 305, 412
High density data storage, 87, 161, 162, 163
High electron mobility transistor (HEMT), 1111
High impedance (HZ) with compensating circuit, 1107, 1110
High pass spatial filter, 63
High-powered fiber amplifier, 866
Higher order Gaussian beam, 226
Hilbert transform, 1090
H_2O (water), *see* Water
HM (half-mirror), 505, 1005
Holes in semiconductors, 1153
Hole burning, 939, 941
  spatial, 968, 969, 1004, 1007
  spectral, 945
Hole density, 898
Holey optical fiber (HF), 706
Hologram, 81
  Gabor-type, 81
  off-axis type, 85
Holographic writing of photoinduced Bragg grating, 743
Holography, 81
  acoustic, 95, 97
  applications of, 92
  color, 87, 92
  computer-generated, 97
  interference, 93
  microwave, 96
  nonoptical, 95
  to explain phase conjugate, 504, 530
  thin emulsion, 88
  synthetic aperture, 97
  vibration measurement, 45
  video display, 99
  volume, 87
  white light, 87
Homeotropic orientation, 344, 345
Homodyne detection, 339, 340, 809, 812, 817, 830, 1096, 1143
  by stimulated Brillouin scattering, 817
Homogeneous broadening, 841
Homogeneous orientation, 344, 345
Homojunction, 930
Horizontal and vertical external fields, 684
Horizontal external field
  in rib guides, 684
  in embedded guides, 683
  periodic, 684
Horseshoe crab's eyes, 447, 585
How to use waveplates, 385
  full-wave plate, 385
  half-wave plate, 385
  quarter-wave plate, 386
Huygens' principle, 15, 16
Huygens' wavelet, 299
Huygens' wavelet ellipsoid, 577
Hybrid modes, 709, 723
  designation of, 729, 730
Hydrogen to create transmission loss, 695, 697
Hydroxyl (OH) formation, 695, 697
Hyperbola or ellipse, 433
Hysteresis curve, 1021
HZ (high impedance) circuit, 1107, 1110

Iceland spar (calcite crystal) ($CaCO_3$), 277, 290, 402
IF (intermediate frequency) amplifier, 808, 1094
IM (intensity modulation), 313, 1086, 1087
Image intensifier, 352
Image processing, *see* Picture processing
Immersed guide, 672
Immobile charges, 332, 1154
Impedance
  characteristic wave, 124
  electric, 112
  intrinsic, 111, 112, 869
  of p-n junction, 1159
Impedance approach, referring to the normal direction, 124
Impulse response function of free space, 14
Impulse train function, 30
Incoherent–coherent converter, 352
Incremental impedance of a p-n junction, 1159
Index of refraction, 112, 266
  relationship to the dielectric constant, 112, 266
Index-matching fluids, 954
$In_{1-x}GaAs_yP_{1-y}$ (gallium arsenide doped with indium and phosphorus) laser, 895
Indicatrix, 264, 285, 292
  absorption indicatrix, 400
  distortion of, 302
Indium tin oxide (ITO), 188
Infrared absorption associated with the vibration of Si — O network, 708
Inhomogeneous broadening, 841

Injection current, density of, 910, 911
InP:Fe (iron-doped indium phosphide), 512
In-phase component, 810, 1090
InP:Ti (titanium-doped indium phosphide), 512
Insertion loss, 132
Integrated optics, 605
  arrayed-waveguide grating (AWG) for, 673
  couplers, 678
  electrodes, 681
  modulators, 312, 1086
  optical guides, 670
  optical interconnects, 982
  optical magic T, 678
Intensity modulation (IM), 313, 1086, 1087
Intensity pattern of diffraction, 22, 27
Interdigital electrode, 324
Interference filter, 167
Interference holography, 93
Interferometers, 167, 605
Intermediate frequency (IF), 1094
Internal writing of Bragg gratings, 743
Internal reflection, total, 130, 145, 147, 296
Interrogation of handwritten letters, 521
Intradyne detection, 812, 830
Intrinsic impedance, 111, 112, 869
Intrinsic impedance η_0 of a vacuum, 111
Intrinsic impedance η_1 of a medium, 111
Intrinsic type semiconductors, 1151
Invariance of the location of the input pattern, 50
Invariance of the state of polarization, 362
Inversion, population, 836, 847, 864
Inversion symmetry of crystals, 541, 599
Ionic dipole, 818
Iron doped indium phosphide (InP:Fe), photorefractive effect of, 512
Isopachic fringe pattern, 95
Isolator, optical, 328, 330, 838, 954, 1005
ITO (indium tin oxide) electrodes, transparent electrode of, 188
I-type layer, 803
I-V characteristics of p-n junctions, 1156

Jacobian elliptic function with modulus m, 1027
Jitter in coherent communication systems, 819
Joining fibers, 786
Joint transform correlator (JTC), 64, 70
Jones matrix of, 421, 444
  eigenvectors, 428
  half-wave plates, 424
  polarizers, 422
  quarter-wave plates, 424
  retarders, 424, 425
  rotators, 425
Jones vector, 421
JTC (joint transform correlator), 64, 70
  compared to VLC, 72

$\kappa - \beta$ diagram, 617, 618
$\kappa - \gamma$ diagram, 618
k coordinate expression of Snell's law, 145, 164
KBr (potassium bromide), 708
KCl (potassium chloride), 708
KDP (potassium dihydrogen phosphate), electrooptic properties of, 305, 316, 321, 1018
KD_2PO_4(K-KDP), 305
k-matching, 113
Kepler, Johannes, 131, 249
Kerr, John, 304, 316
  cells, 317
  coefficients, 316
  constants, 317
  electrooptic effect, 304, 316, 1019, 1051
  media, 512, 518, 527, 1017
Keying device, 1095
Kink in the emission characteristic curve of laser diodes, 941
Kirchhoff, Gustav Robert, 12, 13, 56, 58, 545
Korteweg–de Vries (KdV) differential equation, 1049
Kr_2 (krypton) for excimer lasers, 894

l-rotary (left-handed) quartz, 412, 415
λ/4 shift DFB lasers, 961
Laminated aluminium polyethylene (LAP), 697
Laminated electrodes, 680
Laplacian operator, 711
Laser cavity, 175
Laser(s)
  chemical, 895
  definition of, 833
  dye, 894
  edible, 895
  erbium-doped fiber, 1004
  fiber Raman, 1009
  gas, 893
  ring, 904
  semiconductor, 895
  solid state, 894, 967
Laser cooling of atoms, 254
  Doppler cooling, 254
  polarization gradient cooling, 254
Laser diode(s) (LD)
  amplitude modulation characteristics, 919
  arrays of, 980
  beam patterns, 946
  conditions of laser oscillation, 905
  confinement of carriers, 930
  confinement of injection current, 933
  confinement of light, 937
  distributed Bragg reflector (DBR), 957
  distributed feedback (DFB), 958
  gain spectrum, 896, 902
  laser noise, 952
  light transmitter, 1082
  modulation characteristics, 919

Laser diode(s) (LD) (*Continued*)
multi-quantum well (MQW), 984
noise, 952
output power, 913, 926
quantum dot, 1002
quantum well, 984
quantum wire, 999
rate equations, 909
relaxation oscillation, 916, 917
selection of light sources, 1011
single frequency, 956
temperature dependence, 951
threshold current density, 912
tunable, 189, 970, 977
turn on delay, 914, 926
unwanted FM, 944, 1086, 1095
wavelength tuning, 970
Laser frequency stabilizer, 176
Laser oscillation
condition for, 905
explanation of, 908
Latitude lines of β (or ϵ) of Poincaré sphere, 478
Lausent type saccharimeter, 417
Law of superposition, 548–550, 1070
LD, *see* Laser diode
Leaky wave, 150
LED, *see* Light-emitting diode
Leith–Upatnieks type hologram, 82
Lengths of major and minor axes of elliptical polarization, 316
Lens (convex lens), 40
collimating property, 42
diffraction limited, 45
finite aperture, 102, 103
focal length, 43
Fourier transformable property, 46
imaging property, 43
radiation pressure, 260
Levorotary (l-rotary) optical activity, 412
LiF (lithium fluoride), 708
Lifetime, 834
of electrons in the conduction band, 910
of photons inside a cavity, 910, 920
Light computer, 1021
Light confinement in laser diodes in the horizontal (lateral) direction, 937
Light confinement in laser diodes in the vertical (transverse) direction, 937
Light detectors used in optics laboratories, 796, 800
Light-emitting diode (LED), 952, 1007, 1011, 1013, 1082, 1092
characteristics, 1007
structure, 1008
Light path (optical path), expression of, 761, 762
Light to pump EDFA, 837
Light rays, direction of, 274, 275, 299
Light transmitters, 893, 1082
Light tweezers, 226, 249, 251, 253
$LiIO_3$ (lithium iodate), 306
$LiNbO_3$, *see* Lithium niobate
Linear core layer sandwiched by nonlinear cladding layers, 1037
Linear normalized thickness, 1047
Linear polarization (LP) mode, 709, 726, 727, 736
Linear polarizer, 365, 394, 587
Linearly birefringent, 413
Linearly polarized $LP_{m\mu}$ mode, 726. *See also* $LP_{m\mu}$ modes
Linearly polarized wave, 363
Lines of constant parameters on Argand diagram
azimuth θ, 455
ellipticity ϵ, 458
Li_2O (lithium oxide) to lower the melting point of glass, 706
Lippmann hologram, 92. *See also* Denisyuk hologram
Liquid carbon disulphide CS_2, stimulated Brillouin scattering (SBS) of, 512
Liquid crystal(s), 341
cholesteric, 342
discotic, 344
nematic, 343
smectic, 343
Liquid crystal devices, 346
Fabry–Pérot resonator, 346
fiber filter, 176, 188
microlens, 347
rotatable waveplate, 346
spatial light modulator (SLM), 349, 350, 351, 352
television, 350
Lithium iodate ($LiIO_3$), electrooptic properties of, 306
Lithium fluoride (LiF), 708
Lithium oxide (LiO_2), 706
Lithium niobate $LiNbO_3$
deposited on lithium tantalate, 296–299, 575
elastooptic effect of, 320
electrooptic effect of, 144, 305, 307, 312, 358, 579, 649, 821
photorefractive effect of, 332, 511
$LiTaO_3$, *see* Lithium tantalate
Lithium tantalate ($LiTaO_3$)
electrooptic properties of, 296, 298, 299, 305, 321
photorefractive effect of, 511
LO (local oscillator) light, 192, 807
Local oscillator (LO) light, 192, 807
Local oscillator intensity noise power (LOIN), 1122
Location of the Gaussian beam waist, 208, 217
Longitudinal external field to bulk waves, 681
Longitudinal lines of the Poincaré sphere, 479
Longitudinal mode number, 972
Longitudinal modes, 846, 909
Long-line effect, 954
Loose-tube covered fiber, 784

Lorentz lineshape function, 840
Loss, fiber
  bending, 407
  cavity, 173
  connector, 790, 1130, 1131, 1138, 1144, 1146
  coupling, 1130, 1138, 1144
  fiber transmission, 694, 1138, 1146
  increase due to hydrogen and gamma-ray irradiation, 695
  splicing, 790, 1130, 1131, 1138, 1146
Low pass spatial filter, 62
Low-work-function material, 796
LP (linear polarization) mode, 709, 727, 736
$LP_{m\mu}$ modes, 726, 728, 730, 735, 737
  designation of, 729
  field pattern of, 735
Lumped element approach, 1106
Lyot–Ohman filter, 363, 385

m^2 (beam propagation factor), 220
Mach–Zehnder fiber coupler, 745
Mach–Zehnder interferometer, 679, 683, 793, 1019
Mach–Zehnder light modulator, 685
Magnesium fluoride (MgF_2), 681, 708
Magnesium oxide (MgO) to lower melting point of glass, 706
Magnetic flux densities, 112
Magnetic flux leaking out of transformer, 145
Magnetic permeability, 111, 326
Magnetooptic effect, 326
Magnetooptic property of a Nd:YAG crystal, 968
Magnification of images, 219, 220
Major and minor axes of elliptically polarized waves, 372, 431, 433
Major principal transmittance k_1 of polarizer sheets, 396
MAN (metropolitan area network), 866
Manchester Code, 1097
Manipulation of micron-sized spheres with light beams, 249
Mass fusion splicing machine of fibers, 786, 787
Material dispersion of optical fiber, 699, 701, 752
Matrix
  coefficient, 623
  Jones, 421
  transmission, 625, 626, 630
Maximum available power from source, 1114
Maximum energy coupling in four-wave mixing, 528
Maximum modulation frequency, 920
Maxwell's equations, 268, 521, 607, 609, 610, 709, 715
  boundary conditions of continuity, 112, 121
MCVD (modified chemical vapor deposition) method, 778
Mellin transform, 76
Memory disks, high density, 161
Mercury flash lamp, 894
Meridional rays, 693, 718, 771
Mesa structure (steep hill structure) laser diodes, 995
MESFET (metal semiconductor field effect transistor), 1110
Metal field plate to prevent stray light into detector diodes, 803
Metal guide for integrated optics, 672
Metal semiconductor field effect transistor (MESFET), 1110
Metastable lifetime of Er^{3+}, 891
Methane (CH_4), 512
Method by trial solution, 1024
Method of separation of variables, 1022, 1160
Metropolitan area network (MAN), 866
MgF_2 (magnesium fluoride), 681, 708
MgO (magnesium oxide) to lower the melting point of glass, 706
Michelson interferometry, 166
Microbending loss of optical fibers, 783, 784
Microbes, 363
Microchannel plate to enhance electron density, 800
Microfiche recording, 93
Microscope, Schlieren, 63
Microwave hologram, 96
Minimum detectable power of receivers, 1113
Minor principal transmittance k_2 of polarizer sheets, 396
Mirrors
  dichroic, 167
  half, 65, 505
  phase conjugate, 504
MISER (monolithic isolated single-mode end-pumped ring) laser, 968, 1012
MMI (multimode interference) splitter, 674
MNA (2-methyl-4-nitroaniline) crystal for phase modulation, 182
Modal noise, laser diode, 955
Modal retarder of TM–TE converter, 499
Mode
  density, 843, 847, 1160
  dispersion of optical fiber, 612, 699, 703, 769
  hopping of laser diodes, 945
  in graded index fibers, 759, 767, 768, 772
  in nonlinear guides, 1043, 1048
  in step index fibers, 718
  inside cavities, 223, 226, 235
  of propagation, 605
  number with the periodic boundary condition, 1163
  number in slab guides, 612
  patterns in graded-index fibers, 770
  pattern in rectangular guides, 662
  patterns in slab guides, 615
Mode converter, 497, 502, 684, 685
Mode hopping, 945
Mode-index lens, 620

Mode locking, 945
Mode number, *see* Mode
Modified Bessel function, 714
  of the first kind, 714
  of the second kind, 714
Modified chemical vapor deposition (MCVD) method, 778
Modified ray model method, 606, 622, 636
Modulation format
  amplitude (AM), 1086, 1087, 1083
  amplitude shift keying (ASK), 812, 1089, 1095
  double sideband (DSB), 1086, 1087
  double sideband suppressed carrier (DSBC), 1087
  frequency (FM), 1088, 1092
  frequency shift keying (FSK), 1089, 1095
  intensity (IM), 313, 1086, 1087
  phase (PM), 1088, 1092, 1094
  phase shift keying (PSK), 1089, 1095
  pulse amplitude (PAM), 1088, 1094
  pulse code (PCM), 1084, 1089, 1094
  pulse duration modulation (PDM), 1088, 1094
  pulse position (PPM), 1088, 1094
  quadrature amplitude (QAM), 1087, 1090
  single sideband (SSB), 1087, 1090
  sinusoidal, 103
  vestigial sideband (VSB), 1087, 1091
Modulation index, 1086
Modulators, 312, 605, 1018
Mole fraction, 923
Molecular reorientation of nematic liquid crystals, 345
Momentum of photons, 250, 572
Momentum-matching, 113
Monochromator, 175
Monocoated fiber, 784
Monolithic isolated single-mode end-pumping ring (MISER) laser, 968, 1012
Monomode fiber (single mode fiber), 700, 705, 728
Monomode optical guide, 612
Morse code, 1095
Movement of a point on the Poincaré sphere, 494
  along constant latitude (β), 497
  along constant longitude (θ), 494
Movement parallax, 92
MQW laser, *see* Multi-quantum-well laser
MQW (multi-quantum-well) nonlinear layers, 1019
Multiexposed hologram, 519
Multi-ingredient fiber, 706
Multimode graded-index fiber, 705, 769
Multimode interference (MMI) splitter, 674
Multimode step-index fiber, 703
Multiple D fiber, 866, 867
Multiplexer, 677
Multiplexing (MX), 1085, 1097
  code-division (CDM), 1085, 1094
  frequency division (FDM), 1085, 1094, 1099
  time division (TDM), 1085, 1094, 1100
  wavelength division (WDM), 191, 674, 676, 1085, 1098
Multiplexing fiber sensor, 747
Multiplier layers of APD, 803
Multi-quantum-well (MQW) laser, 984, 992, 994–999
  energy states, 988
  gain curve, 992
  threshold current, 994
Multi-quantum-well (MQW) nonlinear layer, 1019
Multiunit fiber cable, 784
MX, *see* Multiplexing

$N \times N$ matrix connection by AWG, 676
NA (numerical aperture), 151, 693, 694, 701, 703, 706, 770
N_α (zero gain electron density of laser media), 902
NaCl (sodium chloride), 341, 708
$NaClO_3$ (sodium chlorate), optically active, 412
Na_2O (sodium oxide) to lower the melting point of glass, 706
Narrow stripe electrodes of laser diodes, 934
N,*N*-dimethyl-N-octadecyl-3-amino-propyltrimethoxysilyl chloride (DMOAP), 353
Natural sugar ($C_{12}H_{22}O_{11}$), 414
Nd:YAG (neodymium-doped yttrium-aluminum-garnet) laser
  rod, 890, 1178
  solid-state laser, 967, 1012
Near field (Fresnel field), 13
  imaging, 110
  of optical microscopes, 150
  optics, 110
Nebulae, Faraday effect of, 363
Negative birefringence, 267
Negative feedback circuit, 1109
Nematic liquid crystal, 341, 343, 347, 357
NEP (noise equivalent power), 1117
Neodymium (Nd)-doped fiber amplifier, 836, 837
Neodymium YAG laser, *see* Nd:YAG laser
Neodymium YAG rod, 889
NH_4F (ammonium fluoride), 158
Nicol prism, 402
Nitrobenzene ($C_6H_5NO_2$)
  Kerr constant of, 316, 317
  Kerr effect of, 1017, 1035
Nitrotoluene ($C_6H_4(CH_3)NO_2$), Kerr effect of, 1017
NLC (nematic liquid crystal), 341, 343, 347, 357
Nondeformed strain, 304
Noise
  associated with relaxation oscillation, 955
  due to external optical feedback, 953
  due to fluctuations in temperature and injection current, 955
  due to mode hopping, 955

due to spontaneous emission, 955
excess index, 1117
FM, 955
in an APD, 1117
in detector systems, 1113
partition, 955
quantum limited, 1116, 1131
shot, 1113
thermal, 1114
thermal noise limited, 1116, 1131
Noise equivalent power (NEP), 1117
Noise figure, 880, 883
Nonlinear
differential equations, 550, 1022, 1024
distributed feedback (DFB) gratings, 1019
Kerr media, 512, 518
normalized thicknesses, 1041, 1048
parameter, 1072, 1077
saturable absorbers, 945
Schrödinger equation, 1067
susceptibility of photorefractive crystals, 511
Nonlinear-index coefficient n_2, 1023, 1066
Nonoptical holographies, 95
Nonpolarizing beam splitter (NPBS), 411, 822, 830
Nonradiative transitions, 835
Nonreciprocal effects, 713
Nonreturn to zero (NRZ) code, 774, 1096, 1140
Normal dispersion region, 1056
Normal to ellipse, expression of, 432
Normalized guide index b, 659
Normalized propagation parameter, 728
Normalized pumping rate of fiber amplifiers, 850
Normalized steady-state population difference, 849
Normalized thickness of optical guides, 611, 1041, 1047, 1048
"nose" of Panda fibers, 407
NPBS (nonpolarizing beam splitter), 411, 822, 830
NRZ (nonreturn to zero) code, 774, 1096, 1140
n-type semiconductors, 1151, 1152
Numerical aperture (NA), 151, 693, 694, 701, 703, 706, 770
νth order Bessel function of the first kind, 713

O modes in nonlinear guides, 1043, 1048
OASLM (optically addressed spatial light modulator), 351
Odd TM modes of optical planar optical guides, 608, 609, 610
Off-axis type hologram, 85
Offset core fiber, 866
OH (hydroxyl) ion impurities, 695, 697
On-off modulation (ASK modulation), 812, 1095
One-stage processes of fabrication of optical fibers, 775
OPC (optical phase conjugate), 541
Operational amplifier, 1107
Optic axis of crystals, 267
Optical activity, 412, 415, 586
Optical amplifier
based on doped fiber, 833. *See also* Erbium doped fiber amplifier (EDFA)
based on photorefractive effect, 334, 338
Optical communication, *see* Fiber optic communication
Optical correlators, 64
Optical directional coupler, 988
Optical feedback, 903
Optical fiber(s)
cabling of, 783
characteristic equation of, 717, 718
cross sectional field distribution, 730
dispersion shifted, 703, 705, 749, 755, 1051
dominant mode of, 724
dual mode, 739, 757, 758
fabrication of, 775
gamma ray exposure, 697
graded index, 759, 762, 768, 769, 770
holey, 706
hydrogen exposure, 695
joining, 786
kinds of, 703
modes, 718, 726, 729, 736
multimode, 703
numerical aperture of, 693
other than silica based, 708
photoimprinted Bragg grating, 741, 1010
polarization preserving, 707
solution in the cladding region, 714
solution in the core region, 713
splicing, 786
transmission loss of, 694
Optical fiber communication, *see* Fiber optic communication
Optical fiber connector, 790, 1130
Optical fibers other than silica-based fibers, 708
Optical guide, *see* Optical slab guide
Optical guide coupler, 144
Optical interconnect, 99, 982
Optical isolator, 327, 1005
polarization dependent, 328
polarization independent, 330
Optical magic T, 678
Optical modulator, 312, 679, 1018
Optical phase conjugate (OPC), 541
Optical resonator, *see* Fabry–Pérot cavity
Optical signal multiplexing, 1085
Optical signal processing, *see* Spatial filters
Optical slab guides, 605, 606
asymmetric, 606, 638
characteristic equation, 610, 633, 634, 643
coefficient matrix approach, 622, 623
component waves, 615
coupled slab guide, 643, 651
dispersion equation, 612, 637, 638
effective index of refraction, 619
even modes, 608, 628

Optical slab guides (*Continued*)
field distribution, 615
mode cutoff, 612
mode number, 612
modified ray model, 636
normalized thickness, 611
odd modes, 608, 628
propagation constant, 612
symmetric, 606
transmission matrix method, 625, 630
TE modes, 607, 620
TM modes, 607
wave optics approach, 607
Optical tweezer, 226, 249, 251, 253
Optical waveguides
arrayed waveguide grating (AWG), 673
characteristic equation, 657
conflection, 666
conflection lens, 667
coupling between guides, 664
effective index method, 661
electrode configurations, 681
electrode structures, 680
magic T, optical, 678
mode converter, 685
mode patterns, 662
polyimide half-waveplate, 677
power divider, 673
practical example of designing, 659
rectangular, 655
types of waveguides for integrated optics, 670
Y junction, 673
Optically addressed spatial light modulator (OASLM), 351
Optically tunable optical filter, 341
Optimum length of fiber amplifiers, 867
Optimum number of EDFA repeater stations, 882
Optimum value of the multiplication factor of an APD, 1148
Optoelastic, *see* Elastooptic effects
Optoelectronic, *see* Electrooptic effects
Orbital electrons, 1151
Order of mode (mode number), *see* Mode
Ordinary wave (o-wave), 269, 272, 273, 365
Organohydrogen polysiloxane monomer as hardener, 697
Orientation polarization, 265
Orthogonal eigenvectors, 429
Orthogonality between constant θ and ϵ lines of the Poincaré sphere, 465
Orthoscopic image, 85
OVD (outside vapor deposition) method, 778, 780
Outlining input image, 64
Output power from laser diodes, 913, 926
Outside vapor deposition (OVD) method, 778, 780
Overdrive parameter of the injection current, 952
Overview of fiber-optic communication systems, 1082
o-wave (ordinary wave), 269, 272, 273, 365
Oxazine for dye lasers, 894

p wave (parallel wave), 118, 157
Pair production, 1155, 1159
PAM (pulse amplitude modulation), 1088, 1094
Panda fiber, 406, 707
"eyes" of, 407
"nose" of, 407
polarizer, 407
Parallax
accommodation, 93
binocular, 92
movement, 92
Parallel alignment of the reflectors of a Fabry–Pérot interferometer, 199
Parallel switching, 1092
Parallel wave (*p* wave), 118, 157
Paraxial, 11
Partition noise of laser diodes, 955
Parts per billion (ppb) (10^{-9}), 695
Pattern recognition, *see* Optical correlator; Phase conjugate optics
PBS (polarizing beam splitter), 412, 822, 830
PCS (plastic-clad silica) fiber, 775
PCM (pulse code modulation), 1084, 1089, 1094
PCVD (plasma chemical vapor deposition) method, 778, 779
PDFA (praseodymium-doped fiber amplifier), 836
PDLC (polymer-dispersed liquid crystal) type spatial light modulator (SLM), 352
PDM (pulse duration modulation), 1088, 1094
Period of relaxation oscillation, 930
Periodic boundary condition, 1161, 1163
Periodic horizontal external field, 684
Periodic vertical external field, 684
Permanent dipole moment, 265, 266
Permanent joint, 786
Permeability, magnetic, 111, 326
Perpendicular wave (s wave), 118, 157
Perturbation theory, 1164
Phase conjugate optics, 504
expressions of, 507
for distortion free amplification, 513
for eliminating wave front distortion, 508
for picture processing, 512, 519
for self-tracking, 514, 517
in real time, 511
mirror, 504, 508
Phase conjugate wave, 504, 521, 537, 539
explained by holography, 504
generation of, 506

Phase conjugation method used in dispersion compensators, 755
Phase correction factor of Fabry–Pérot resonators, 232
Phase discriminator, 1094
Phase diversity technique, 812
Phase grating, 334
Phase jitter, 819, 823
Phase-lock loop (PLL), 809
Phase matching, 113, 284
Phase modulation (PM), 1088, 1090, 1092, 1094
Phase modulators, 1018
Phase shifter, 310
Phase shift keying (PSK), 1089, 1095
Phase velocity, 3, 748
Phonons, 817, 835
Phosphorus oxychloride ($POCl_3$), 777
Phosphorus pentaoxide (P_2O_5), 697
Photocell, 800
Photochromic flexible guide, 672
Photoconductive effect, 800
Photoconductor cell, 800
Photocurrent, 1159
Photodetector, 1159
Photodiode, 800
Photoelastic effect, *see* Elastooptic effect
Photoelastic fringes, 95
Photoelastic sheet, 95
Photoelasticity, 94
Photoelectric effect, *see* Electrooptic effect
Photoinduced Bragg grating, 741, 747, 755
  applications of, 744
  methods of writing, 742
  photograph of, 745
Photolithography machine, 513
Photomultiplier tube (PMT), 796, 798
Photon
  flux density, 843
  lifetime, 920
  momentum, 250, 572
  punch press, 903
  radiation pressure, 249
Photon–electron converter, of a cathode, 796
Photon-induced pair production, 1103, 1159
Photon tunneling microscope, 152, 161
Photorefractive beam combiner for coherent homodyne detection, 339
Photorefractive effect, 331
  adaptive fiber couplers, 517
  cockroach theory, 333
  crystals, 77, 511, 517
  energy transfer, 335
  enhanced by external electric fields, 334
  fly-swatter theory, 332
  joint transform correlator, 77
  optical amplifiers, 334
  optically tunable filters, 341
  real time phase conjugate mirrors, 511
Photosensitivity, 741
Phototransistor, 800
Photovoltaic effect, 800
P–I curve (light power vs. injection current) of laser diodes, 927
Picture processing, 512, 519
  by phase conjugate mirrors, 519
  by spatial filters, 61
Piezoelectric effect, 100
Piezo transducer (PZT), 176
Pigtailed laser diode, 1130
PIN and PN photodetectors, 796, 801, 1085, 1104, 1151
  equivalent circuit of, 1102
Pinhole camera, 103, 558
Planar optical guides, *see* Optical slab guide
Planar-type W guide, 633
Planck's constant, 250, 797
Planck's factor of the radiation law, 1114
Plane mirror cavity, 234
Plane waves, 1, 268
Plano-convex lens, 41
Plasma chemical vapor deposition (PCVD) method, 778, 779
Plasma effect, 939, 944, 963
Plasma-enhanced modified chemical vapor deposition (PMCVD) method, 779
Plastic-clad silica (PCS) fiber, 775
Plastic fiber, 706
PLL (phase locked loop), 809
PM (phase modulation), 1088, 1092, 1094
PM (pulse modulation), 1094
PMCVD (plasma-enhanced modified chemical vapor deposition) method, 779
PMMA (polymethyl methacrylate), 706
PMT (photomultiplier tube), 796, 798
PN diode, 801
p-n junction, 1154
P_2O_5 (phosphorus pentaoxide), 697
Pockels effect, 1018
Pockels electrooptic effect, 304
  electrooptic properties, 305, 307
  for amplitude modulators, 312
  for phase shifters, 310
  for retarders, 311
$POCl_3$ (phosphorus oxychloride), 777
Poincaré, Henrie, 451
Poincaré sphere, 451
  constant θ and ε lines, 465
  converted from Argand diagram, 469
  fabrication of a Poincaré sphere, 483
  for solving polarizer problems, 485
  for solving retarder problems, 479
  how to construct, 451
  how to use, 451
  lines of constant ε, 458, 478
  lines of constant θ, 455, 479
  traces of, 490

Point-by-point writing of photoinduced Bragg grating, 744
Point spread function, 14
  in cylindrical coordinates, 243
  in rectangular coordinates, 14
Polariscope, 95, 363, 415
Polarizers, 362
  based on bending optical fibers, 404
  based on scattering, 408
  dichroic, 394
  pile-of-plates, 407
  polarizing beam splitter (PBS), 412
  polaroid sheets, 399
Polarization
  azimuth of, 366
  circular, 364
  elliptical, 364
  linear, 363
Polarization dependent optical isolator, 328
Polarization, electrical, 264, 266
Polarization diversity method of detection, 822
Polarization gradient laser cooling, 255
Polarization independent optical isolator, 330
Polarization jitter, 363, 819
Polarization jitter controls, 819
Polarization-maintaining fibers (or preserving), 404, 707, 819
Polarization microscopes, 363
Polarizing beamsplitter (PBS), 412, 822, 830
Polaroid polarizer, 399
Polygonal mirror, 101
Polyimide half-wave-plate ($\lambda/2$ plate), 677
Polymer-dispersed liquid crystal (PDLC) type spatial light modulator (SLM), 352
Polymethyl methacrylate (PMMA) for plastic fibers, 706
Population difference of carriers, 860
Population inversion of carriers, 836
Population inversion factor, 847, 864
Position vector, 6
Positive birefringence, 267
Positive carrier, 1153
Potassium bromide (KBr), 708
Potassium chloride (KCl), 708
Potassium dihydrogen phosphate (KDP)
  elastooptic properties of, 321
  electrooptic properties of, 305, 316, 1018
Power budget (requirement), 1130, 1144, 1148
Power density handling capability of fiber amplifiers, 866
Power density-interaction length for nonlinear effects, 816
Power dividers, 673
Power intensity, 114
Power level parameter for nonlinearity, 1041
Power required to establish a fundamental soliton, 1071
Power saturation of EDFA, 838
Power scrambler for optical networks, 677
Power transmittance k of a polarizer, 485
Poynting vector, 114, 117, 118, 124, 125, 275, 283
ppb (parts per billion) (10^{-9}), 695
PPM (pulse position modulation), 1088, 1094
PR, *see* Photorefractive effect
Practical aspects of Fabry–Pérot interferometers, 234
Practical aspects of optical fibers, 693
Praseodymium-doped fiber amplifier (PDFA), 836
Preform rod, optical fiber, 775, 777, 837
Preheating ends of fiber before splicing, 790
Preventing glare, 412, 419
Primitive (basic) communication system, 1101
Principal axes of an ellipse, 303, 431, 433, 434
Principal mode number of graded-index fibers, 768, 771
Principal section, 277, 278
Principal strains, 318
Principle of conflection, 667
Prisms, polarizing, 402
Probability of spontaneous emission, 836
Probes to detect evanescent fields, 154
Projection, back, 469
Propagation
  constant, 4, 607, 612, 728, 729, 759, 768, 1055
  direction of, 1, 2, 5, 7, 9
  vector, 4, 5
Propagation in anisotropic media, 263
  in uniaxial crystals, 270, 272
Pros and cons of 1.48-μm and 0.98-μm pump light, 853
Proton-bombarded region, 935
Pseudoscopic image, 85, 87, 89
PSK (phase shift keying), 1089, 1095
p-type semiconductor, 1151, 1153
Pulfich's refractometer, 164
Pulse amplitude modulation (PAM), 1088, 1094
Pulse broadening compensation in optical fibers by four-wave mixing, 537
Pulse code modulation (PCM), 1084, 1089, 1094
Pulse duration modulation (PDM), 1088, 1094
Pulse modulation, 1094
Pulse position modulation (PPM), 1088, 1094
Pump light of optical amplifiers, 834, 835, 837, 844
Pyrometer, 800
PZT (piezo transducer), 176, 207, 213, 215

q parameter, 209, 214, 216, 221, 222
Q of the cavity, 194
QAM (quadrature amplitude modulation), 1087, 1090
Quadratic electrooptic coefficients, 316
Quadratic phase factor, 48, 49, 559
Quadrature amplitude modulation (QAM), 1087, 1090
Quadrature component, 810, 1090
Qualitative explanation of laser oscillation, 908

Quantization of the propagation constant in graded index fibers, 766
Quantum dot laser, 1002
Quantum efficiency, 797, 803, 804, 831, 951
Quantum energy, 249
Quantum limited, 1116
Quantum size effect, 999
Quantum-well laser, multi-, 933, 945, 984
Quantum wire laser, 999
Quarter-wave fiber loop, 382, 383
Quarter-waveplate
  function of, 371, 377, 386, 394, 480, 822
  generate circular polarization by, 386
  how to use, 386, 389, 392, 393
  Jones matrix of, 424
  optical fiber, 382
  problems associated with, 394, 480, 483, 493
Quartz crystal (SiO_2), 412, 414, 447
  electrooptic properties, 305
  rotary power, 412, 414, 448
Quasi-Fermi levels, 898, 994
Quasi-Gaussian beam, 221
Quasi single longitudinal mode laser, 909

Radiation patterns
  antenna, 13, 95
  from a half wave dipole, 103
Radiation pressure, 166
  of laser light, 249
  of photons, 254
Radiation sensitive optical fiber, 698
Radiation therapy, 698
Radius of curvature of the wavefront of Gaussian beams, 208
Raised resistivity by proton bombardment, 934
Raman, 818
  amplifier, 818
  fiber laser, 1009
  oscillator, 818
  scattering, 816, 817
  spectrum, 818
Rate equations for the three-level model of Er^{3+}, 848
Rate equations of semiconductor lasers, 909
Rate of transitions of the carriers, 841
Ray, 275
  direction of, 275, 299
  path, 277
  velocity diagram, 579
Ray theory, 759
Rayleigh range of a Gaussian beam, 209
Rayleigh resolution criteria, 151
Rayleigh scattering, 694, 408, 448
Rayleigh–Sommerfeld diffraction formula, 54, 58, 59, 545
Real image, 85, 87
Real time correlation, 77
  of joint transform correlator (JTC), 77
  of Vander Lugt correlator (VLC), 77
Reciprocity theorem, 713
Recombination of electrons and holes, 898
Rectangle function, 20
  Fourier transform of, 20
Rectangular optical waveguide, 605, 655, 670
Reflectance
  at arbitrary incident angle, 124
  at normal incidence, 117
Reflected waves, 113
Reflecting telescope, Cassegrain, 107, 108
Reflection coefficients, 113, 114
  at arbitrary incident angle, 118
  electrical (at normal incidence), 113, 114
  magnetic (at normal incidence), 113, 114
Reflection coefficient for total internal reflection, 135
Reflection finesse, 194
Reflection of e-waves
  from anisotropic boundaries, 294
  in the case of total internal reflection, 295, 296
Refractive index, *see* Index of refraction
Refractive index gradient constant of Selfoc fibers, 763, 771
Relative core index step Δ, 758
Relative magnetic permeability, 111
Relaxation oscillation of laser diodes, 916, 917, 930
Relaxation time of laser diodes, 1013
Repeater stations, 833, 1081
Required frequency bandwidth for amplifying digital signals, 1139
Resolution
  of lens-type microscopes, 150
  of scanning near field optical microscope (SNOM), 154
Resolving power of Fabry–Pérot resonators, 192, 194
Resonance of Fabry–Pérot resonators, 167
  condition, 174
  frequency of pth longitudinal mode, 232, 233
  wavelength, 178
Responsivity R
  of photomultiplier tubes, 798
  of PIN and APD photodiodes, 803, 805
Retardance, 310, 365, 367, 431, 437, 579, 587
  measurement of, 392, 393
  microbes' pattern of, 363
Retarded time frame, 1064
Retarder, 310, 365, 378, 587
Revere, Paul, 1085
Return to zero (RZ) code, 774, 1096, 1141
Rhodamine 6G dye, 357, 894
Rib guide, 670
Ribbon fiber cable, 785, 786, 790
Ridge guide, 670

Ring-type fiber laser, 1004
Ring type solid state laser, 904, 967
Rise time of laser diode turn-on, 930
Rise-time requirement, 1132, 1145
Rochon prism, 404
Rotary power, 412, 414
  dextrorotary (right-handed) or *d*-rotary, 412
  leverotary (left-handed) or *l*-rotary, 412
Rotation and scaling of the input image in signal processing, 73
Rotation
  of coordinates, 309, 314
  of linearly polarized waves by the Faraday effect, 327
  of major axes of an ellipse, 303
  of polarization, 327, 412
Rotators, 365, 412
  Faraday effect, 326
  Fresnel's explanation of, 413
  optical activity, 412
  saccharimeters, 417
Rubber eraser, 303, 323
Ruby laser, 894
Rutile (TiO_2), elastooptic properties of, 321
RZ (return to zero) code, 774, 1096, 1141

S (sulfur), 936
S/N (signal to noise ratio), 1115, 1139
s wave (perpendicular, senkrecht in German) wave, 118, 157
Saccharimeter, 414, 417
SAM (stress-applying member), 707
Sampled function, 31
Saturated back-biased current, 1102, 1158
Saturation signal power intensity, 853, 856, 857
SBS (stimulated Brillouin scattering), 512, 602, 815, 817
Scalar wave approach, 53
Scanning electron microscope (SEM), 98, 161
Scanning of Fabry–Pérot cavities, 176
  by angle of incidence, 184
  by frequency of incident light, 190
  by index of refraction, 187
  by reflector spacing, 177
Scanning Fabry–Pérot spectrometer, 176
Scanning near-field optical microscopes (SNOM), 154
Scattering cross section, 254
Schlieren camera, 63, 103
Schrödinger equation, 985, 1067
Scott-Russell, John, 1049
Second harmonic generation (SHG), 299, 579, 1018
Second order nonlinear susceptibility, 523, 525, 1018
SE (surface emitting) laser, 956
Selection of light sources, 1011
Selfoc fiber, 763, 766, 770, 772
Selfoc lens, 330
Self-focusing in gain guiding laser diodes, 939, 941
Self-induced transparency due to the soliton effect, 1076
Self-oscillation of optical amplifiers, 838
Self-phase modulation (SPM), 1051, 1052, 1053, 1069
Self-pumped phase conjugation (SPPC), 518
Self-tracking capability of phase conjugate waves, 514, 517
Self-tracking of laser beams, 514
Sellmeier formula, 702
SEM (scanning electron microscope), 98, 161
Semiconductors
  acceptor atoms, 1153
  bandgap energy, 804, 901, 925, 945, 996, 998
  carrier concentration, 930
  carrier confinement, 930
  conduction band, 1151
  density of states, 898, 987
  donor atoms, 1152
  effective mass, 924
  Fermi energy level, 898, 1151, 1152
  Fermi occupancy probability function, 1151
  heterojunction, 930
  intrinsic layer, 1151
  n-type semiconductor, 1151, 1152
  p-type semiconductor, 1151, 1153
  p-n junction, 1154
  quantum wells, 933, 984, 999, 1002
  spontaneous emission, 254, 835, 836, 841, 842, 1007
  stimulated emission, 835, 836
  valence band, 1151
Semiconductor laser amplifier (SLA), 837, 838, 847, 895, 1085
  compared to EDFA, 838
Semidegenerate four-wave mixing, 533
Senarmont method for measuring retardance, 392, 491, 587
Senarmont prism, 404
Senkrecht wave (perpendicular, *s* wave), 118
Sensing strain, vibration, and temperature by optical fibers, 747
Sensitivity of photomultipliers, 797
Sensors, 747, 1012
Separation of variables, 1022, 1160
SF_6 (sulfur hexafluoride) for chemical lasers, 895
Shadowgrams, 97
Shah function, 30
  Fourier transform of, 31
Shearing strain, 318
SHG (second harmonic generation), 299, 579, 1018
Shift keying, 1095
Shifting theorem of the Fourier transform, 36
Short-wavelength-loss edge (SLE), 697
Shot noise, 1113
Shot noise limited, 1116
Shot noise power, 876

$SiCl_4$ (silicon tetrachloride), 777
Side pit fiber, 707
Side scan sonar, 96
Side tunnel fiber, 707
Sifting property of the δ-function, 29
Sigma laser, 1004, 1005
Sign conventions, 3, 368
Sign functions, 25
  Fourier transform of, 25
Signal saturation intensity of EDFA, 853, 856
Signal-spontaneous beat noise, 871
Signal to noise ratio (S/N), 1115, 1139
  for ASK modulation, 1121
  for homodyne detection, 1122
  for thermal noise limited case, 1131
  of output from PIN photodiodes, 876
Silica core flourine-added cladding fiber, 705
Silicon dioxide (SiO_2)-based fiber, 708
Silicon tetrachloride ($SiCl_4$), 777
SiO_2 (quartz), 305, 412, 681
Sinc function, 21
Sine-Gordon equation, 1076
Single crystal mixer, 814
Single heterojunction, 930
Single longitudinal mode (SLM) laser, 907, 956
Single-mode fiber (monomode), 700, 705, 728
Single-mode laser, 909
Single-mode optical guide, 612
Single sideband (SSB) modulation, 1087, 1090
Sketching hybrid mode patterns in optical fibers, 732
Sketching linearly polarized mode patterns in optical fibers, 735
Skew modes in optical fibers, 721
Skew rate, 1125
Skew ray in optical fibers, 693
SLA, *see* Semiconductor laser amplifier
Slab optical guide, *see* Optical slab guides
SLE (short-wavelength-loss edge), 697
Slipping-buffer layer in optical fiber cables, 783
Slit, apodized, 22
SLM (spatial light modulator), 349, 350, 351, 352
SLM (single longitudinal mode) laser, 907, 956
Slow axis of retarder, 367
Slowly varying envelope approximation (Bohr approximation), 529
Small signal amplitude modulation, 916
Smectic liquid crystal, 341
  "A" type, 343
  "C" type, 343, 354, 356
S/N (signal to noise ratio), *see* Signal to noise ratio,
Snell's law, 110, 112, 113, 121, 135, 292, 565, 566
  in k-coordinates, 145, 164
Snitzer proposal, 723
SnO_2 (tin dioxide) transparent electrode, 311
Sodium chlorate ($NaClO_3$), optically active, 412
Sodium chloride (NaCl), 341, 708
Sodium oxide (Na_2O) to lower the melting point of glass, 706
SNOM (scanning near field optical microscope), 154
Šolc filters, 385
Soleil compensator, 382
Solid-state laser, 893, 894, 895, 967
Soliton, 1049
  collisions of, 1070
  envelope function, 1059, 1064
  fundamental mode, 1069, 1071
  generation of, 1050, 1052
  history of, 1049
  optical communication system, 1073, 1077
  period, 1070, 1077
  (pulse width) $\times$ (amplitude) product, 1073, 1077
  waves, 1017, 1052, 1056, 1081
Sommerfeld, *see* Rayleigh–Sommerfeld diffraction formula
Soot, 779
Space charge polarization, 265
Spatial derivative operation in image processing, 64
Spatial filters, 61
  derivative operation, 64
  high pass, 63
  low pass, 62
  phase contrast, 63
Spatial frequency, 1, 4, 8
  approaches in Fourier optics, 52
  components, 60
Spatial hole burning, 968, 969, 1004, 1007
Spatial light modulator (SLM), 349, 350, 351, 352
Spatial mapping of microwave radiation, 343
Spectral hole burning, 945
Spectral lineshape, 839
Spectral measurements by Fabry–Pérot resonator without reflector spacing, 203
Spectroscopy, 4
Spectrum of laser diode emission, 920
Splicing fibers, 786, 866
Splicing losses, 790, 1130, 1138, 1146
Split field polarizer as an analyzer, 418
SPM (self-phase modulation), 1051, 1052, 1053, 1058, 1069
Spontaneous emission, 254, 835, 836, 841, 842, 1007
Spontaneous emission lifetime, 254, 842, 926
Spontaneous–spontaneous beat noise, 871, 872, 875
SPPC (self-pumped phase conjugation), 518
SRS (stimulated Raman scattering), 512, 818, 1009
SSB (single sideband) modulation, 1087, 1090
SSG (superstructure grating) laser diode, 977, 979
Stability condition of cavity resonators, 231, 234
Stable operation of optical tweezers, 253
Standardized Japanese system for telephone bit rate, 1102

Standardized U.S. system for telephone bit rates, 1102
Standing-wave boundary condition, 1161, 1163
Standing wave patterns, 89, 135, 235, 615, 636, 637, 969
Star coupler, 678
State of polarization, 371
  graphical solution of, 365
Static energy of induced polarization, 345
Steady-state solutions of rate equations, 911
Step and repeat function, 31, 64
Step function, 25
  Fourier transform of, 25
Step index fiber, theory of, 709
Step index multimode fiber, 703, 709, 1139
Stereoscopic motion picture, 362
Stimulated Brillouin scattering (SBS), 512, 602, 815, 817
Stimulated effects, 816
Stimulated emission, 835, 836, 842
Stimulated emission cross section, 843, 859
Stimulated Raman scattering (SRS), 512, 818, 1009
Stokes radiation, 816, 1009
Stopper layer of electrode structure, 681
Strain, 318
  fiber sensor, 747
  free, 304
  principal, 318
  shearing, 318
Stranded cable, 784, 786
Stranding pitch, 784
Streak camera, 798
Streak pattern due to edge diffraction, 25
Stress-applying member (SAM), 707
Stripe substrate laser, 936
Strip-loaded guide, 671
Strongly coupled quantum wells, 988
Subcarrier frequency, 1099
Substrate, 606
Sucrose ($C_{12}H_{22}O_{11}$), 412
Sulfur (S), 939
Sulfur hexafluoride (SF_6) for chemical lasers, 895
Summary of essential formulas for calculating states of polarization, 439
Summary of transformations of Gaussian beams by a lens, 219
Sunglasses, 362
Superheterodyne radio, 807
Superposition, law of, 549, 1070
Superstructure grating, 978
Superstructure grating (SSG) laser diode, 977, 979
Surface acoustic wave (SAW) light deflector, 324
Surface acoustooptic modulator, 99
Surface-emitting diode LED, 1008
Surface-emitting (SE) laser, 956
Surface wave, 135
Susceptibility
  electric, 266
  nonlinear, 511, 523, 1018
  tensor, 266
Switching electrodes, 685
Symmetric guide, 606
Symmetric matrix, 430
Synchronous homodyne receiver, 818
Synthetic aperture holography, 97
System loss, 1130

3Rs of repeaters, 833
TAT 1 (Trans Atlantic Transmission), 693
Taylor series expansion, 538, 1063
TDM (time division multiplexing), 1085, 1094, 1100
TE (transverse electric) wave, 607, 620, 719, 1030
$TE_{0\mu}$ mode, 720, 721, 732
Teleaxicon, 245
Telephone bit rates, standardized U.S., 1102
TE-like modes, 656
Tellurium dioxide (TeO_2), elastooptic properties of, 100, 322
TEM_{00} Gaussian mode, 226, 253
TEM_{11} Gaussian mode beam, 226, 253
Temperature dependence of L-I curves of laser diodes, 951
Temperature effect on the wavelength of laser diodes, 945
Temperature fluctuations measured by optical fibers, 747
Temporal frequency f, 4
Tensor susceptibility, 266
TeO_2 (tellurium dioxide), 100, 322
Terminal contact resistance, 920
Terminal potential, 928
Terraced substrate (TS) laser diode, 936
TE–TM converter, 502, 594, 685
Tetrahedral substrate, 666
TFT (thin film transistor), 350
Thallium bromide (TlBr), 708
Thallium monoxide (Tl_2O) to lower the melting point of glass, 706
Thermal noise, 876, 1114
Thermal noise limited, 1116
Thermister, 800
Thin emulsion holograms, 88, 108
Thin film transistor (TFT), 350
Third-order dispersion effect, 755
Third-order nonlinear electric susceptibility, 523, 527, 1018
Three-dimensional displays, 92
Three-layer optical guide with a linear core and nonlinear identical claddings, 1038
Three-level material, 836
Threshold carrier density, 927
Threshold current, 909, 928, 930
Threshold electron density, 928
Threshold gain, 928

Threshold light intensity of pump light, 860
Thulium (Tm^{3+}) for upconversion fiber amplifiers, 889
Time constant of photomultipliers, 798
Time constant of relaxation oscillation, 917
Time degradation of LD output power, 1130
Time-dependent solutions of rate equations, 859, 914
Time division multiplexing (TDM), 1085, 1094, 1100
Time reversed videotape, 504
Time reversed wave, 507
Tin dioxide (SnO_2) transparent electrode, 311
TiO_2 (rutile), 321
Titanium doped indium phosphide (InP:Ti), as a Kerr medium, 512
TJS (transverse junction stripe) laser, 936
Tl_2O (thallium monoxide) to lower the melting point of glass, 706
TlBr (thallium bromide), 708
Tm^{3+} (thulium) for upconversion fiber amplifiers, 889
TM (transverse magnetic) mode, 607, 719, 722
TM-like modes, 656, 657, 659
$TM_{0\mu}$ mode, 732
TM_0 mode of optical guides (dominant mode), 612
TM–TE converter, 497, 498
TN (twisted nematic) liquid crystal, 344, 350, 351, 359
TNSLM (twisted nematic liquid crystal spatial light modulator), 349
Top layer of electrode structures, 680
Topographic images obtained by AFM, 160
Total internal reflection
  of e-waves, 285, 296, 298
  of o-waves, 130, 145, 147, 605
  reflection coefficient for, 135
Transcendental equations, 611, 639
Transfer length of optical couplers, 651
Transformation of Gaussian beams by lenses, 214
Transformation of q parameters by lenses, 215
Transimpedance circuit (TZ), 1109, 1110, 1136
Transistor–transistor logic (TTL), 1124
Transition energy, 254
Transmission coefficients
  at arbitrary incident angle, 118
  at normal incidence, 114
  electrical, 113, 114
  magnetic, 113, 114
Transmission loss of fibers, 694
Transmission matrix method, 606, 622, 625, 626, 630
Transmittance
  at arbitrary incident angle, 124
  at normal incidence, 117
Transmitted waves, 113
Transmitters, 893
Transmitting satellite, 259
Transoceanic fiberoptic submarine communication, 1017
Transparent electrodes, 188, 311
Transverse electric (TE) wave, 607, 620, 719, 1030
Transverse external electric field to bulk waves, 681
Transverse junction stripe (TJS) laser, 936
Transverse magnetic (TM) wave, 607, 719
Trapping micron-sized dielectric spheres, 249
Triangle function, 21
  Fourier transform of, 22
Trimming electrodes, 684
Triple-photon excitation, 889
Trunk lines, 786
TS (terraced substrate) laser diode, 936
TTL (transistor–transistor logic), 1124
Tunable optical coupler, 740
Tunable optical filter
  electrically, 188, 189
  optically, 341
Tuning of the wavelength of a laser diode
  by Bragg reflector tuning current I_b alone, 973
  by combining I_p and I_b, 975
  by phase controller tuning current I_p alone, 973
Turn-on characteristics of laser diodes, 920
Turn-on delay time of laser diodes, 909, 914, 926, 930
Turpentine ($C_{10}H_{16}$), optical activity of, 412
Twisted nematic (TN) liquid crystal, 344, 350, 351, 359
Twisted nematic spatial light modulator (TNSLM) liquid crystal, 349
Two-frequency method to remove phase jitter, 823
Two-stage processes to fabricate optical fibers, 775, 837
Two stage process to fabricate erbium doped fibers, 837
Two-wave mixing gain of photorefractive materials, 518
Tyndall, John, 692
Types of optical fibers, 703
TZ (transimpedance) circuit, 1109, 1110, 1136

U-groove cable, 786
Unbounded field, 627
Unclamped condition, 304
Undersea transmission cable, 833
Unfolded, manipulation of conflection, 666, 667
Uniaxial crystals, 267
  condition of propagation in, 272
  graphical solution of propagation in, 270
  propagation inside, 270
Unit fiber cable, 784
Unit vector for the direction of propagation, 7
Unpolarized (depolarized) wave, 364
Upatnieks hologram, *see* Leith–Upatnieks type hologram

Upconversion fiber amplifier, 889
Upper limit on the modulation frequency of a laser diode, 909, 920, 930
U.S. standard rating DS-4 PCM, 1145
UV laser beam, 744
UV-light-cured resin, 697

V shape fiber loss curve, 694
VAD (vapor-phase axial deposition) method, 778, 780, 781
Valence band, 899, 998, 1151
Values of sin Δ to determine the handedness of elliptical polarization, 375
Vanadium tetrachloride (VCl_4), 777
Van der Waals forces, 341
Vander Lught correlator (VLC), 64, 66
Vapor-phase axial deposition (VAD) method, 778, 780, 781
VCl_4 (vanadium tetrachloride), 777
VCSEL (vertical cavity surface-emitting laser), 957, 981
Vector propagation constant, 4
Velocity matched Mach–Zehnder interferometer, 683
Velocity of the envelope, 748
Velocity surface of Huygens' wavelet ellipsoid, 578
Verdet constant, 327
Vernier effect, 977
Vertical cavity surface-emitting laser (VCSEL), 957, 981
Vertical external field in adjacent embedded guides, 683
Vertical external field
  in embedded guides, 683
  periodic, 684
Very near field, 150
Vestigial sideband (VSB) modulation, 1087, 1091
V-groove fiber, 786
V-grooved substrate buried heterostructure (VSB) laser, 936
Vibration fiber sensor, 747
Vibration measurement by holography, 95
Virtual image, 84
VLC (Vander Lugt correlator), 64
  compared to JTC, 72
Voice channels, number of, 1102
Voice recognition, 77
Volume hologram, 87
VSB (vestigial sideband), 1087, 1091
VSB (V-grooved substrate buried heterostructure) laser, 936

W guide, 635
Waist location of Gaussian beams, 208
  emergent from lenses, 217
Waist size of Gaussian beams, 208
  emergent from lenses, 216
Water (H_2O)
  as a nonlinear medium, 1017
  as a polarized medium, 266
  elastooptic properties of, 320
  Kerr constant of, 316, 317
Wave equations in cylindrical coordinates, 709, 710
Wave expressions of light, 132
Wavefront, 276
Waveguide, *see* Optical waveguides
Waveguide dispersion of optical fibers, 699, 701, 752
Waveguide index, optical fiber, 750
Wavelength, 4, 8, 110, 112
Wavelength converter, 352
Wavelength division multiplexing (WDM), 191, 674, 676, 1085, 1098, 1175
Wavelength filter, 684
Wavelength matching, 113
Wavelength-selective insertion into a WDM, 747, 1175
Wavelength-selective tapping in a WDM, 747
Wavelength shift of radiation from laser diodes, 943
Wavenormal, 275, 295, 759, 761, 762
  and Snell's law, 292
Wavenumber, 177, 890
Wave optics approach, propagation in planar optical guides, 606, 607
Waveplate, 379
  fiber type, 382, 383, 385
  full, 385
  half, 385
  how to use, 385, 386, 389, 392, 393
  quarter wave, 371, 377, 382, 386, 389, 393, 394, 480, 483, 493, 822
Wavevector method to analyze waves in anisotropic media, 264, 282
  diagram of, 283, 284
WDM (wavelength division multiplexing) system, 191, 674, 676, 1085, 1098, 1175
Weakly coupled quantum wells, 988
Weakly guiding approximation, 722, 726
Weyl expansion theorem, 545
White light hologram, 87
Wollaston polarizing prism, 404
Work function, 796
Writing photoinduced Bragg gratings in optical fibers, 742, 744

Xe_2 (Xenon) as medium for excimer lasers, 894
X-ray analysis, 89, 97
X-ray crystallography, 13

Y-cut crystal, 307, 312
Y junction, 673
YAG (yittrium aluminum garnet) laser, 101, 894, 967
Y-branch laser diode array, 983

$Y_3Al_5O_{12}$ (YAG) yittrium aluminum garnet, 101, 894, 967
$Y_3Fe_5O_{12}$ (YIG) yittrium iron garnet, 329, 330
YIG (yittrium iron garnet), 329, 330
Yittrium aluminum garnet (YAG), 101, 894, 967
Yittrium iron garnet (YIG), 329, 330

ZBLAN fiber, 889
Z-cut crystal, 307
Zero gain electron density N_α of laser media, 902
Zeroth-order Bessel functions of the second kind, 714
Zinc (Zn), 936
Zinc chloride ($ZnCl_2$), 708
Zinc oxide (ZnO), electrooptic properties of, 306
Zinc sulfide (ZnS), elastooptic properties of, 320
Zinc telluride (ZnTe), electrooptic properties of, 306
Zn (zinc), 936
$ZnCl_2$ (zinc chloride), 708
ZnO, *see* Zinc oxide
ZnS, *see* Zinc sulfide
ZnTe, *see* Zinc telluride