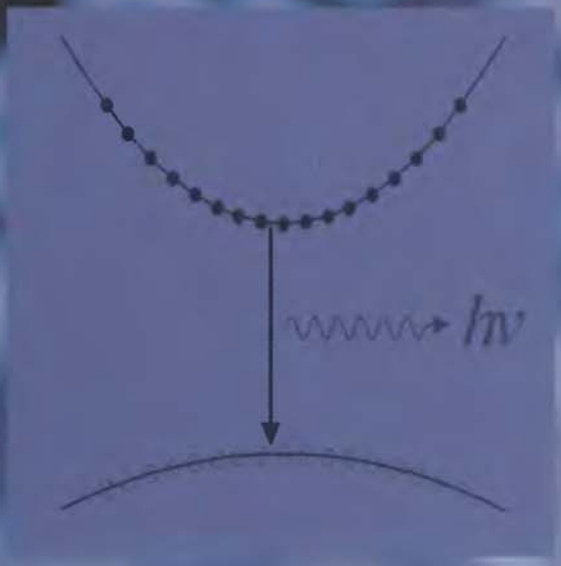
AMNON YARIV
POCHI YEH

PHOTONICS

OPTICAL ELECTRONICS IN MODERN COMMUNICATIONS



SIXTH EDITION

THE OXFORD SERIES IN ELECTRICAL AND COMPUTER ENGINEERING

Adel S. Sedra, Series Editor

Allen and Holberg, *CMOS Analog Circuit Design, 2nd edition*
Bobrow, *Elementary Linear Circuit Analysis, 2nd edition*
Bobrow, *Fundamentals of Electrical Engineering, 2nd edition*
Burns and Roberts, *An Introduction to Mixed-Signal IC Test and Measurement*
Campbell, *The Science and Engineering of Microelectronic Fabrication, 2nd edition*
Chen, *Digital Signal Processing*
Chen, *Linear System Theory and Design, 3rd edition*
Chen, *Signals and Systems, 3rd edition*
Comer, *Digital Logic and State Machine Design, 3rd edition*
Comer, *Microprocessor-Based System Design*
Cooper and McGillem, *Probabilistic Methods of Signal and System Analysis, 3rd edition*
DeCarlo and Lin, *Linear Circuit Analysis, 2nd edition*
Dimitrijev, *Principles of Semiconductor Devices*
Dimitrijev, *Understanding Semiconductor Devices*
Fortney, *Principles of Electronics: Analog & Digital*
Franco, *Electric Circuits Fundamentals*
Ghausi, *Electronic Devices and Circuits: Discrete and Integrated*
Guru and Hiziroğlu, *Electric Machinery and Transformers, 3rd edition*
Guru and Warrier, *Electrical Circuits: Analysis and Design*
Houts, *Signal Analysis in Linear Systems*
Jones, *Introduction to Optical Fiber Communication Systems*
Krein, *Elements of Power Electronics*
Kuo, *Digital Control Systems, 3rd edition*
Lathi, *Linear Systems and Signals, 2nd edition*
Lathi, *Modern Digital and Analog Communication Systems, 3rd edition*
Lathi, *Signal Processing and Linear Systems*
Martin, *Digital Integrated Circuit Design*
Miner, *Lines and Electromagnetic Fields for Engineers*
Parhami, *Computer Architecture*
Parhami, *Computer Arithmetic*
Roberts and Sedra, *SPICE, 2nd edition*
Roulston, *An Introduction to the Physics of Semiconductor Devices*
Sadiku, *Elements of Electromagnetics, 3rd edition*
Santina, Stubberud, and Hostetter, *Digital Control System Design, 2nd edition*
Sarma, *Introduction to Electrical Engineering*
Schaumann and Van Valkenburg, *Design of Analog Filters*
Schwarz and Oldham, *Electrical Engineering: An Introduction, 2nd edition*
Sedra and Smith, *Microelectronic Circuits, 5th edition*
Stefani, Savant, Shahian, and Hostetter, *Design of Feedback Control Systems, 4th edition*
Tsividis, *Operation and Modeling of the MOS Transistor, 2nd edition*
Van Valkenburg, *Analog Filter Design*
Warner and Grung, *Semiconductor Device Electronics*
Wolovich, *Automatic Control Systems*
Yariv and Yeh, *Photonics: Optical Electronics in Modern Communications, 6th edition*
Żak, *Systems and Control*

Photonics

Optical Electronics in Modern Communications

SIXTH EDITION

Amnon Yariv
California Institute of Technology

Pochi Yeh
University of California, Santa Barbara

New York Oxford
OXFORD UNIVERSITY PRESS
2007

Oxford University Press, Inc., publishes works that further Oxford University's objective of excellence in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2007 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016
http://www.oup.com

Oxford is a registered trademark of Oxford University Press

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, mechanical, photocopying, recording, or otherwise, without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data

Yariv, Amnon.
Photonics : optical electronics in modern communications / Amnon Yariv, Pochi Yeh.—6th ed.
p. cm.—(The Oxford series in electrical and computer engineering)
Rev. ed of: Optical electronics in modern communications. 5th ed., c1997.
Includes bibliographical references and index.
ISBN-13: 978-0-19-517946-0
ISBN 0-19-517946-3
1. Photonics. I. Yeh, Pochi. II. Yariv, Amnon. Optical electronics in modern communications. III. Title. IV. Series.

TA1520.Y37 2006
621.382'7—dc22

2005047270

Printing number: 9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

Contents

Preface xi

Chapter 1 Electromagnetic Fields and Waves 1

1.0 Introduction 1
1.1 Maxwell's Equations and Boundary Conditions 1
1.2 Energy Density and Poynting Vector 4
1.3 Monochromatic Fields and Complex-Function Formalism 6
1.4 Wave Equations and Monochromatic Plane Waves 8
1.5 Chromatic Dispersion and Group Velocity 13
1.6 Polarization States and Representations (Stokes Parameters and Poincaré Sphere) 19
1.7 Electromagnetic Propagation in Anisotropic Media (Crystals) 30
1.8 Plane Waves in Uniaxially Anisotropic Media—Phase Retardation 36
1.9 Jones Matrix Method 41
1.10 Elementary Theory of Coherence 56
Problems 59
References 65

Chapter 2 Rays and Optical Beams 66

2.0 Introduction 66
2.1 Ray Matrices 66
2.2 Skew Rays and Reentrant Rays 72
2.3 Rays in Lenslike Media 73
2.4 Wave Equation in Quadratic Index Media and Beams 77
2.5 Gaussian Beams in Homogeneous Media 79
2.6 Fundamental Gaussian Beam in a Lenslike Medium—The *ABCD* Law 83
2.7 Gaussian Beams in Lens Waveguide 87
2.8 High-Order Gaussian Beam Modes in a Homogeneous Medium 88
2.9 Gaussian Beam Modes in Quadratic Index Media 91
2.10 Propagation in Media with a Quadratic Gain Profile 95
2.11 Elliptic Gaussian Beams 97
2.12 Beam Propagation and Diffraction Integral 99
Problems 106
References 109

Chapter 3 Guided Waves in Dielectric Slabs and Fibers 110

3.0 Introduction 110
3.1 TE and TM Confined Modes in Symmetric Slab Waveguides 110

3.2 TE and TM Confined Modes in Asymmetric Slab Waveguides 118
3.3 Step-Index Circular Dielectric Waveguides (Linearly Polarized Modes in Optical Fibers) 126
3.4 Effective Index Theory 137
3.5 Waveguide Dispersion in Optical Fibers 140
3.6 Attenuation in Silica Fibers 145
Problems 149
References 153
Additional Reading 155

Chapter 4 Optical Resonators 156
4.0 Introduction 156
4.1 Fabry–Perot Etalon 160
4.2 Fabry–Perot Etalons as Optical Spectrum Analyzers 170
4.3 Optical Resonators with Spherical Mirrors 172
4.4 Mode Stability Criteria 176
4.5 Modes in a Generalized Resonator—Self-Consistent Method 178
4.6 Resonance Frequencies of Optical Resonators 180
4.7 Losses in Optical Resonators 183
4.8 Ring Resonators 184
4.9 Multicavity Etalons 194
4.10 Mode Matching and Coupling Loss 204
Problems 206
References 209
Additional Reading 210

Chapter 5 Interaction of Radiation and Atomic Systems 211
5.0 Introduction 211
5.1 Atomic Transitions and Electromagnetic Waves 211
5.2 Atomic Polarizability and Dielectric Constant 213
5.3 Classical Electron Model 214
5.4 Dispersion and Complex Refractive Index 216
5.5 Lineshape Function—Homogeneous and Inhomogeneous Broadening 221
5.6 Induced Transitions—Absorption and Amplification 225
5.7 Gain Saturation in Homogeneous Laser Media 230
5.8 Gain Saturation in Inhomogeneous Laser Media 232
Problems 235
References 236

Chapter 6 Theory of Laser Oscillation and Some Specific Laser Systems 237
6.0 Introduction 237
6.1 Fabry–Perot Laser 237
6.2 Oscillation Frequency 242

6.3 Three- and Four-Level Lasers 244
6.4 Power in Laser Oscillators 246
6.5 Optimum Output Coupling in Laser Oscillators 248
6.6 Multimode Laser Oscillation and Mode Locking 251
6.7 Mode Locking in Homogeneously Broadened Laser Systems 265
6.8 Pulse Length Measurement and Narrowing of Chirped Pulses 273
6.9 Giant Pulse (Q-Switched) Lasers 281
6.10 Hole Burning and the Lamb Dip in Doppler-Broadened Gas Lasers 287
6.11 Some Specific Laser Systems 290
6.12 Frequency Comb and Optical Frequency Metrology 303
Problems 308
References 309
Additional Reading 312

Chapter 7 Chromatic Dispersion and Polarization Mode Dispersion in Fibers 313
7.0 Introduction 313
7.1 Chromatic Dispersion in Optical Transmission Systems 313
7.2 Optical Pulse Spreading in Dispersive Media 317
7.3 Polarization Effects in Optical Fibers 322
7.4 Principal States of Polarization 325
7.5 Vector Analysis of Polarization Mode Dispersion 329
7.6 High-Order PMD and Compensators 346
Problems 350
References 353

Chapter 8 Nonlinear Optics 354
8.0 Introduction 354
8.1 On the Physical Origin of Nonlinear Polarization 354
8.2 Second-Order Nonlinear Phenomena—General Methodology 355
8.3 Electromagnetic Formulation and Optical Second-Harmonic Generation 358
8.4 Other Second-Order Nonlinear Processes 369
8.5 Quasi Phase Matching 377
8.6 Third-Order Nonlinear Optical Processes 380
8.7 Stimulated Brillouin Scattering 387
8.8 Four-Wave Mixing and Phase Conjugation 392
8.9 Frequency Tuning in Parametric Oscillation 399
Problems 402
References 404

Chapter 9 Electro-optic Modulation of Laser Beams 406
9.0 Introduction 406
9.1 Linear Electro-optic Effect 406
9.2 Electro-optic Modulation—Phase, Amplitude 418

9.3 High-Frequency Modulation Considerations 427
9.4 Electroabsorption and Electroabsorption Modulators 431
9.5 Electro-optical Effect in Liquid Crystals 434
9.6 Acousto-optic Effect (Photoelastic Effect) 440
9.7 Scattering of Light by Sound 446
9.8 Bragg Diffraction—Coupled-Wave Analysis 450
9.9 Bragg Cells and Beam Deflectors 458
Problems 461
References 463

Chapter 10 Noise in Optical Detection and Generation 465
10.0 Introduction 465
10.1 Limitations Due to Noise Power 466
10.2 Noise—Basic Definitions and Theorems 469
10.3 Spectral Density Function of a Train of Randomly Occurring Events 471
10.4 Shot Noise 473
10.5 Johnson Noise 475
10.6 Spontaneous Emission Noise in Laser Oscillators 479
10.7 Phasor Derivation of Laser Linewidth 484
10.8 Coherence and Interference 491
10.9 Error Probability in a Binary Pulse Code Modulation System 496
Problems 499
References 500

Chapter 11 Detection of Optical Radiation 501
11.0 Introduction 501
11.1 Optically Induced Transition Rates 501
11.2 Photomultiplier 503
11.3 Noise Mechanisms in Photomultipliers 505
11.4 Heterodyne Detection with Photomultipliers 507
11.5 Photoconductive Detectors 511
11.6 The *p-n* Junction 517
11.7 Semiconductor Photodiodes 521
11.8 Avalanche Photodiode 529
11.9 Power Fluctuation Noise in Lasers 532
Problems 536
References 537
Additional Reading 538

Chapter 12 Wave Propagation in Periodic Media 539
12.0 Introduction 539
12.1 Periodic Media 539

12.2 Periodic Layered Media—Bloch Waves 545
12.3 Bragg Reflectors 555
12.4 Coupled-Wave Analysis 560
12.5 Periodic Waveguides 573
12.6 Spectral Filters and Fiber Bragg Gratings 582
12.7 Chirped and Tapered Index Gratings 587
12.8 2-D and 3-D Periodic Media (Photonic Crystals) 594
Problems 600
References 601

Chapter 13 Waveguide Coupling 602
13.0 Introduction 602
13.1 General Properties of Modes 602
13.2 Dielectric Perturbation Theory and Mode Coupling 607
13.3 Coupling of Two Parallel Waveguides—Directional Coupler 611
13.4 Coupling of N Parallel Identical Waveguides—Supermodes 618
13.5 Phase Matching and Frequency Selective Coupling—Multiplexing 622
13.6 Mode Converters 626
Problems 630
References 632

Chapter 14 Nonlinear Optical Effects in Fibers 633
14.0 Introduction 633
14.1 Kerr Effect and Self-Phase Modulation 633
14.2 Cross-Phase Modulation—Polarization 637
14.3 Nondegenerate Four-Wave Mixing 641
14.4 Partially Degenerate Four-Wave Mixing 653
14.5 Optical Solitons 663
Problems 670
References 671

Chapter 15 Semiconductor Lasers—Theory and Applications 673
15.0 Introduction 673
15.1 Some Semiconductor Physics Background 674
15.2 Gain and Absorption in Semiconductor (Laser) Media 680
15.3 GaAs/$Ga_{1-x}Al_xAs$ Lasers 686
15.4 Some Real Laser Structures 691
15.5 Direct-Current Modulation of Semiconductor Lasers 696
15.6 Gain Suppression and Frequency Chirp in Current-Modulated Semiconductor Lasers 700
15.7 Integrated Optoelectronics 709
Problems 711
References 712

Chapter 16 Advanced Semiconductor Lasers 714
16.0 Introduction 714
16.1 Carriers in Quantum Wells (Advanced Topic) 715
16.2 Gain in Quantum Well Lasers 720
16.3 Distributed Feedback Lasers 724
16.4 Vertical Cavity Surface Emitting Semiconductor Lasers 738
Problems 746
References 746

Chapter 17 Optical Amplifiers 748
17.0 Introduction 748
17.1 Semiconductor Optical Amplifiers 749
17.2 Erbium-Doped Fiber Amplifiers 752
17.3 Amplified Spontaneous Emission 755
17.4 Optical Amplification in Fiber Links 761
17.5 Raman Optical Amplifiers 767
Problems 774
References 776

Chapter 18 Classical Treatment of Quantum Noise and Squeezed States 778
18.0 Introduction 778
18.1 The Uncertainty Principle and Quantum Noise 778
18.2 Squeezing of Optical Fields 787
Problems 795
References 796

Appendix A Wave Equation in Cylindrical Coordinates and Bessel Functions 797
Appendix B Exact Solutions of the Step-Index Circular Waveguide 802
Appendix C Kramers–Kronig Relations 812
Appendix D Transformation of a Coherent Electromagnetic Field by a Thin Lens 817
Appendix E Fermi Level and Its Temperature Dependence 820
Appendix F Electro-optic Effect in Cubic $\bar{4}3m$ Crystals 823
Appendix G Conversion for Power Units and Attenuation Units 827

Author Index 828
Subject Index 830

Preface

The field of photonics, sometimes referred to as optical electronics, has continued to evolve vigorously during the last decade, thus justifying a major updating of the last (fifth) edition. The present edition has a broader theoretical underpinning and includes new and important subjects.

The book continues the tradition of introducing basic principles in a systematic self-contained treatment with minimal reliance on outside sources. It describes the physics and methodology of operation of the basic optoelectronic components of importance to optical communications and optical electronics. The book is intended to serve both as a text for students in electrical engineering and applied physics as well as a reference book for engineers and scientists working in those fields.

The present edition reflects two major efforts on our part: (1) the addition of new topics related to technology development in optical electronics and communications (and the omission of some less important topics) and (2) the refinement and improvement of materials already in the fifth edition. In the revision process, we decided to tailor the new edition for students, researchers, and engineers in the area of optical communications who are interested in learning how to generate and manipulate optical radiation and how to put this knowledge to work in analyzing and designing photonic components for the transmission of information. The presentation and inclusion of topics also reflect comments and suggestions from many anonymous reviewers and instructors.

Specifically, the main new features of this edition are:

1. The introduction of Stokes parameters and the Poincaré sphere for the representation of polarization states in birefringent optical networks.
2. The use of Fermat's principle for the derivation of rays, beam propagation, and the Fresnel diffraction integral.
3. The use of matrix methods for treating wave propagation in coupled resonator optical waveguides (CROWs).
4. Matrix treatment of multicavity etalons and multilayer structures.
5. Matrix treatment of mode coupling and supermodes in mode-locked lasers.
6. Chromatic dispersion, polarization mode dispersion (PMD) in fibers, and their compensation.
7. Nonlinear optical effects in fibers: self-phase modulation, cross-phase modulation, stimulated Brillouin scattering (SBS) and stimulated Raman scattering (SRS) in fibers, optical four-wave mixing, and spectral reversal (phase conjugation) in fibers.
8. Electroabsorption and waveguide electro-optic Mach–Zehnder modulators.
9. Periodic layered media, fiber Bragg gratings and photonic crystals, and Bragg reflection waveguides.
10. Optical amplifiers: semiconductor optical amplifiers, erbium-doped fiber amplifiers, and Raman amplifiers.

As in the earlier editions, we assume a basic background in electromagnetic theory and familiarity with Maxwell's equations and electromagnetic wave propagation in the bulk and in waveguides. An elementary acquaintance with quantum mechanics is recommended but may be acquired en route.

A generous use of numerical examples is intended to help bridge the gap between theory and applications.

The authors thank their students and colleagues as well as the many reviewers and lecturers whose comments constituted an important factor in the revision.

Pasadena, California Amnon Yariv
Santa Barbara, California Pochi Yeh

CHAPTER 1

ELECTROMAGNETIC FIELDS AND WAVES

1.0 INTRODUCTION

This book deals with many important subjects in optical electronics and their applications in modern optical communications, where optical waves are employed as carriers of information for local and long distance transmission. In this chapter, we review some of the most important and basic properties of electromagnetic radiation. These background materials are included for completeness and as a ready source of reference.

We begin by reviewing Maxwell's equations, which govern the propagation of optical waves in various media, including free space, optical crystals, periodic media, optical fibers, and waveguides. We then describe the boundary conditions on the electric and magnetic field vectors. One of the most important contributions of Maxwell's equations is the prediction of the existence of electromagnetic waves in free space. We discuss the storage as well as the transport of energy that accompany the propagation of electromagnetic waves. These are followed by a derivation of the wave equation and an analysis of the monochromatic plane waves and some of their important properties. We also discuss the polarization states of optical waves and their representation in terms of Jones vectors, Stokes parameters, and the Poincaré sphere. Comprehensive coverage of the propagation of optical waves in anisotropic media is included. This is followed by a detailed discussion of the Jones calculus, which is a powerful technique for analyzing the propagation of optical waves in birefringent systems. In the end, we present a brief discussion of an elementary theory of coherence.

1.1 MAXWELL'S EQUATIONS AND BOUNDARY CONDITIONS

In the great work *Treatise of Electricity and Magnetism*, the Scottish physicist James Clerk Maxwell published in 1873 his original finding of the electromagnetic theory of light. His theory has led to many important discoveries, including the existence of electromagnetic waves. Based on his theory, all electric, magnetic, electromagnetic, and optical phenomena are governed by the same fundamental laws of electromagnetism. These laws are written mathematically in terms of the Maxwell equations (in MKS units):

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \tag{1.1-1}$$

$$\nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J} \tag{1.1-2}$$

$$\nabla \cdot \mathbf{D} = \rho \tag{1.1-3}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1.1-4}$$

In these equations, **E** and **H** are the electric field vector (in volts per meter) and magnetic field vector (in amperes per meter), respectively. These two field vectors are often employed to describe an electromagnetic field or an optical wave. The quantities **D** and **B** are called the electric displacement vector (in coulombs per square meter) and magnetic induction vector (in webers per square meter), respectively. These two vectors are introduced to include the effect of the electromagnetic field on matter. The quantities ρ and **J** are the electric charge density (in coulombs per cubic meter) and electric current density vector (in amperes per square meter), respectively. The electric charge and current may be considered the source of the electromagnetic radiation, represented by the **E** and **H** vector fields. These four equations completely determine the electromagnetic field and are the fundamental equations of the theory of such fields, that is, of electrodynamics.

In the field of optical electronics and optical communications, one often deals with the transmission and propagation of electromagnetic radiation in regions of space where both charge density and current density are zero. In fact, if we set $\rho = 0$ and $\mathbf{J} = 0$ in Maxwell's equations, we find that nonzero solutions exist. This means that an electromagnetic field can exist even in the absence of any charges or currents. Electromagnetic fields occurring in media in the absence of charges are called electromagnetic waves. Maxwell's equations (Equations (1.1-1) to (1.1-4)) consist of 8 scalar equations that relate a total of 12 variables, 3 for each of the 4 field vectors **E**, **H**, **D**, and **B**. They cannot be solved uniquely unless the relationships between **B** and **H** as well as that between **E** and **D** are known. To obtain a unique determination of the field vectors, Maxwell's equations must be supplemented by the so-called constitutive equations (or material equations),

$$\mathbf{D} = \varepsilon \mathbf{E} = \varepsilon_0 \mathbf{E} + \mathbf{P} \tag{1.1-5}$$

$$\mathbf{B} = \mu \mathbf{H} = \mu_0 \mathbf{H} + \mu_0 \mathbf{M} \tag{1.1-6}$$

where the constitutive parameters ε and μ are tensors of rank 2 and are known as the dielectric tensor (or permittivity tensor) and the permeability tensor, respectively; **P** and **M** are electric and magnetic polarizations, respectively. In vector notation, the dielectric tensor and the permeability tensor are often written in terms of 3×3 matrices. When an electromagnetic field is present in matter, the electric field can perturb the motion of electrons and produce a distribution of charge separation. This leads to a dipole polarization per unit volume. Analogously, the magnetic field can also produce a magnetization in materials having a permeability that is different from μ_0. The constant ε_0 is called the permittivity of a vacuum and has a value of 8.854×10^{-12} F/m. The constant μ_0 is known as the permeability of a vacuum. It has, by definition, the exact value of $4\pi \times 10^{-7}$ H/m.

If the material medium is isotropic, both ε and μ reduce to scalars. For most applications in optical electronics and optical communications, the quantities ε and μ can be assumed to be independent of the field strengths. However, if the fields are sufficiently strong, such as obtained, for example, by focusing a laser beam or by applying a strong dc electric field to an

electro-optic crystal, the dependence of these quantities on **E** and **H** must be considered. These nonlinear optical effects will be discussed later in this book.

Boundary Conditions

Maxwell's equations can be solved in regions where both ε and μ are continuous. In optical electronics and optical communications, one often deals with situations in which the physical properties (characterized by ε and μ) change abruptly across one or more surfaces. This occurs, for example, in dielectric waveguides, which consist of transparent layers of different dielectric constants (or refractive indices). Although the physical properties may change abruptly across the dielectric interfaces, there exist continuity relationships of some of the components of the field vectors at the dielectric boundary. These continuity conditions can be derived directly from Maxwell's equations.

We consider a boundary surface separating two media with different dielectric permittivity and permeability (medium 1 and medium 2). To obtain the boundary conditions for **B** and **D**, we construct a thin cylinder over a unit area of the surface, as shown in Figure 1.1a. The end faces of the cylinder are parallel to the surface. We now apply the Gauss divergence theorem

$$\iiint \nabla \cdot \mathbf{F}\, dV = \iint \mathbf{F} \cdot d\mathbf{S} \tag{1.1-7}$$

to both sides of Equations (1.1-3) and (1.1-4). The surface integral reduces, in the limit as the height of the cylinder approaches zero, to an integral over the end faces only. This leads to

$$\begin{aligned} \mathbf{n} \cdot (\mathbf{B}_2 - \mathbf{B}_1) &= 0 \\ \mathbf{n} \cdot (\mathbf{D}_2 - \mathbf{D}_1) &= \sigma \end{aligned} \tag{1.1-8}$$

where **n** is the unit vector normal to the surface directed from medium 1 into medium 2, $\mathbf{B}_1$ and $\mathbf{D}_1$ are the field vectors in medium 1 in the immediate vicinity of the boundary, $\mathbf{B}_2$ and $\mathbf{D}_2$ are the field vectors in medium 2 in the immediate vicinity of the boundary, and σ is the surface charge density (in coulombs per square meter). These boundary conditions are often written

$$\begin{aligned} B_{2n} &= B_{1n} \\ D_{2n} - D_{1n} &= \sigma \end{aligned} \tag{1.1-9}$$

where $B_{1n} = \mathbf{B}_1 \cdot \mathbf{n}$, $B_{2n} = \mathbf{B}_2 \cdot \mathbf{n}$, $D_{1n} = \mathbf{D}_1 \cdot \mathbf{n}$, and $D_{2n} = \mathbf{D}_2 \cdot \mathbf{n}$. In other words, the normal component of the magnetic induction is always continuous, and the difference between the normal components of the electric displacement vector **D** is equal in magnitude to the surface charge density σ.

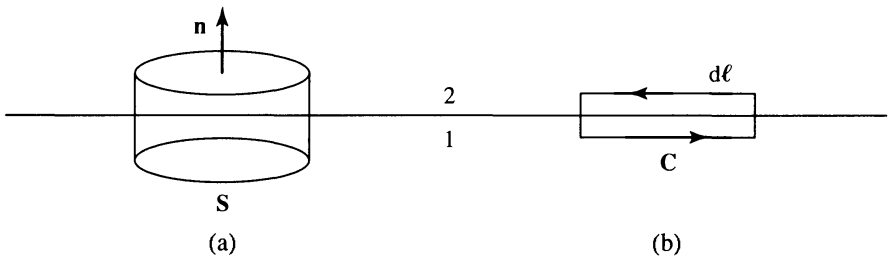


Figure 1.1 Boundary between medium 1 and medium 2. **n** is the unit vector normal to the surface directed from medium 1 into medium 2. $d\ell$ is a differential element along the rectangular contour C.

For the field vectors **E** and **H**, we draw a rectangular contour with two long sides parallel to the surface of discontinuity, as shown in Figure 1.1b. We now apply the Stokes theorem

$$\iint \nabla \times \mathbf{F} \cdot d\mathbf{S} = \int \mathbf{F} \cdot d\ell \tag{1.1-10}$$

to both sides of Equations (1.1-1) and (1.1-2). The contour integral reduces, in the limit as the width of the rectangle approaches zero, to an integral over the two sides only. This leads to

$$\begin{aligned} \mathbf{n} \times (\mathbf{E}_2 - \mathbf{E}_1) &= 0 \\ \mathbf{n} \times (\mathbf{H}_2 - \mathbf{H}_1) &= \mathbf{K} \end{aligned} \tag{1.1-11}$$

where **K** is the surface current density (in amperes per meter), and $\mathbf{E}_1$, $\mathbf{H}_1$, $\mathbf{E}_2$, and $\mathbf{H}_2$ are the field vectors in the immediate vicinity of the boundary in media 1 and 2, respectively. Again, the boundary conditions for the electric field and magnetic field vectors (1.1-11) are often written as

$$\begin{aligned} \mathbf{E}_{2t} &= \mathbf{E}_{1t} \\ \mathbf{H}_{2t} - \mathbf{H}_{1t} &= \mathbf{K} \end{aligned} \tag{1.1-12}$$

where the subscript t means the tangential component of the field vector. (Note: The tangential components of these field vectors to the boundary surface are still vectors in the tangential plane of the surface.) In other words, the tangential component of the electric field vector **E** is always continuous at the boundary surface, and the difference between the tangential components of the magnetic field vector **H** is equal to the surface current density **K**.

In many areas of optical electronics and optical communications, one often deals with situations in which the surface charge density σ and the surface current density **K** both vanish. This occurs, for example, at the boundary between core and cladding of dielectric waveguides. It follows that, in such cases, the tangential components of **E** and **H**, and the normal components of **D** and **B** are continuous across the interface separating media 1 and 2. These boundary conditions are important in solving many wave propagation problems in optical electronics and optical communications, including guided waves in optical fibers and dielectric slab waveguides.

1.2 ENERGY DENSITY AND POYNTING VECTOR

It has been known for some time that light carries energy with it and is a form of electromagnetic radiation. The sun delivers an enormous amount of energy (with an intensity of about 1 kW/m^2 at sea level) to the planet Earth everyday. As we indicated, the first and most conspicuous success of Maxwell's theory was the prediction of the existence of electromagnetic waves and the transport of energy. We now consider two of the most important aspects of electrodynamics: the energy stored with an electromagnetic wave and the power flow associated with an electromagnetic wave. To derive the energy density and the power flow, we consider the conservation of energy in a small volume element in space. The work done per unit volume by an electromagnetic field is $\mathbf{J} \cdot \mathbf{E}$, which may also be considered as the energy dissipation per unit volume. This energy dissipation must be connected with the net decrease of energy density and the power flow out of the volume. According to Equation (1.1-2), the work done by the electromagnetic wave can be written

$$\mathbf{J} \cdot \mathbf{E} = \mathbf{E} \cdot (\nabla \times \mathbf{H}) - \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} \tag{1.2-1}$$

If we now employ the vector identity

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{H} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \mathbf{H}) \tag{1.2-2}$$

and use Equation (1.1-1), the right side of Equation (1.2-1) becomes

$$\mathbf{J} \cdot \mathbf{E} = -\nabla \cdot (\mathbf{E} \times \mathbf{H}) - \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} - \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} \tag{1.2-3}$$

If we now further assume that the material medium involved is linear in its electromagnetic properties (i.e., ε and μ are independent of the field strength), Equation (1.2-3) can be written

$$\frac{\partial U}{\partial t} + \nabla \cdot \mathbf{S} = -\mathbf{J} \cdot \mathbf{E} \tag{1.2-4}$$

where U and $\mathbf{S}$ are defined as

$$U = \tfrac{1}{2}(\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) \tag{1.2-5}$$

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \tag{1.2-6}$$

The scalar U represents the energy density of the electromagnetic fields and has the dimension of joules per cubic meter. The vector $\mathbf{S}$, representing the energy flux, is called the Poynting vector and has the dimensions of joules per square meter per second. It is consistent to view $|\mathbf{S}|$ as the power per unit area (watts per square meter) carried by the field in the direction of $\mathbf{S}$. The quantity $\nabla \cdot \mathbf{S}$ thus represents the net electromagnetic power outflow per unit volume. Equation (1.2-4) is known as the continuity equation or Poynting's theorem. It represents the conservation of energy for the electromagnetic field. The conservation laws for the linear momentum of electromagnetic fields can be obtained in a similar way (see Problem 1.5).

In a region where there's no electric current ($\mathbf{J} = 0$), the continuity equation becomes

$$\frac{\partial U}{\partial t} + \nabla \cdot \mathbf{S} = 0 \tag{1.2-7}$$

Thus, in a current-free region, the decrease of electromagnetic energy density in a volume is a result of the outflow of electromagnetic energy through the surface of the volume.

Dipolar Dissipation

Using the material equations (1.1-5) and (1.1-6), the work done per unit volume by an electromagnetic field, $\mathbf{J} \cdot \mathbf{E}$ (or Equation (1.2-3)), can be written

$$\mathbf{J} \cdot \mathbf{E} = -\nabla \cdot (\mathbf{E} \times \mathbf{H}) - \mu_0 \mathbf{H} \cdot \frac{\partial \mathbf{H}}{\partial t} - \varepsilon_0 \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t} - \mu_0 \mathbf{H} \cdot \frac{\partial \mathbf{M}}{\partial t} - \mathbf{E} \cdot \frac{\partial \mathbf{P}}{\partial t} \tag{1.2-8}$$

Of particular interest in this book is the last term,

$$\mathbf{E} \cdot \frac{\partial \mathbf{P}}{\partial t} \tag{1.2-9}$$

which represents the power per unit volume expended by the field on the electric dipoles. This power goes into an increase in the electrostatic energy stored by the dipoles as well as into supplying the dissipation that may accompany the change in **P**.

1.3 MONOCHROMATIC FIELDS AND COMPLEX-FUNCTION FORMALISM

In optical electronics and optical communications, one often deals with monochromatic light, which serves as a carrier of information. It is known that monochromatic light has a unique angular frequency of oscillation. The field vectors of monochromatic light are sinusoidal functions of time. For the purpose of simplifying the algebraic manipulations, we introduce the complex-function (analytic function) formalism. To illustrate this, we consider, as an example, the function

$$a(t) = |A| \cos(\omega t + \alpha) \tag{1.3-1}$$

where ω is the angular frequency (in units of radian per second), $|A|$ is the amplitude, and α is the phase. If we define the complex amplitude of $a(t)$ by

$$A = |A| e^{i\alpha} \tag{1.3-2}$$

then we can rewrite Equation (1.3-1) as

$$a(t) = \mathrm{Re}[Ae^{i\omega t}] \tag{1.3-3}$$

In complex-function formalism, we will often represent $a(t)$ as

$$a(t) = Ae^{i\omega t} \tag{1.3-4}$$

instead of by Equation (1.3-1) or (1.3-3). This, of course, is not strictly correct so that when this happens it is always understood that what is meant by Equation (1.3-4) is the real part of $A \exp(i\omega t)$. It is important to note that the angular frequency ω is related to the real frequency ν (in units of hertz) by the following equation:

$$\omega = 2\pi\nu \tag{1.3-5}$$

In most situations, when linear operations such as differentiation and integration are involved, the replacement of Equation (1.3-3) by the complex form (1.3-4) poses no problems. The exceptions are cases that involve the product (or powers) of monochromatic field vectors such as energy density and Poynting's vector. In these cases, we must use the real form of the field vectors.

To illustrate the case where the distinction between the real and complex forms is not necessary, consider the problem of taking the derivative of $a(t)$. Using the real form (1.3-1), we obtain

$$\frac{d}{dt} a(t) = \frac{d}{dt} |A| \cos(\omega t + \alpha) = -\omega |A| \sin(\omega t + \alpha) \tag{1.3-6}$$

If we use instead the complex form (1.3-4), we get

$$\frac{d}{dt} a(t) = \frac{d}{dt} Ae^{i\omega t} = i\omega Ae^{i\omega t} \tag{1.3-7}$$

Taking, as agreed, the real part of the last expression and using Equation (1.3-2), we obtain Equation (1.3-6). In fact, the complex form is most convenient when linear operations are involved.

As an example of a case in which we have to use the real form of the function, consider the product of two sinusoidal functions $a(t)$ and $b(t)$, where

$$a(t) = |A| \cos(\omega t + \alpha) = \mathrm{Re}[Ae^{i\omega t}] \tag{1.3-8}$$

$$b(t) = |B| \cos(\omega t + \beta) = \mathrm{Re}[Be^{i\omega t}] \tag{1.3-9}$$

with $A = |A|e^{i\alpha}$ and $B = |B|e^{i\beta}$. Using the real functions, we get

$$a(t)b(t) = \frac{|A||B|}{2}[\cos(2\omega t + \alpha + \beta) + \cos(\alpha - \beta)] \tag{1.3-10}$$

If we were to evaluate the product $a(t)b(t)$ using the complex form of the functions, we would get

$$a(t)b(t) = ABe^{i2\omega t} = |A||B|e^{i(2\omega t+\alpha+\beta)} \tag{1.3-11}$$

Comparing with the last result, Equation (1.3-10) using the real form, we notice that the time-independent term $\cos(\alpha - \beta)$ in parentheses is missing. In addition, the real part of Equation (1.3-11) is a factor of 2 larger than the corresponding term in Equation (1.3-10). Thus, the use of the complex form led to an erroneous result. Generally, the product of the real parts of two complex numbers may not be equal to the real part of the product of these two complex numbers. In other words, if x and y are two arbitrary complex numbers, the following is generally true:

$$\mathrm{Re}[x] \cdot \mathrm{Re}[y] \neq \mathrm{Re}[xy] \tag{1.3-12}$$

Time Averaging of Sinusoidal Products

In optical electronics, one often deals with optical waves with field vectors that are fast varying functions of time. For example, the period of a monochromatic light with a wavelength of $\lambda = 1.5\ \mu\mathrm{m}$ is $T = \lambda/c = 0.5 \times 10^{-14}$ s. Most optical detectors are unable to respond at such a short period of time. One often considers the time-averaged values rather than the instantaneous values of many physical quantities such as Poynting's vector and energy density. It is frequently necessary to find the time average of the product of two sinusoidal functions of the same frequency:

$$\langle a(t)b(t)\rangle = \frac{1}{T}\int_0^T |A| \cos(\omega t + \alpha)\, |B| \cos(\omega t + \beta)\, dt \tag{1.3-13}$$

where $a(t)$ and $b(t)$ are given by Equations (1.3-8) and (1.3-9), and the angular brackets denote time averaging; $T = 2\pi/\omega$ is the period of oscillation. Since the integrand in the above equation is periodic with a period T, the averaging can be performed over any time period that is an integral number of T. By using Equation (1.3-10), we obtain directly

$$\langle a(t)b(t)\rangle = \tfrac{1}{2}|A||B| \cos(\alpha - \beta) \tag{1.3-14}$$

This last result can be written in terms of the complex amplitudes A and B, defined immediately following Equation (1.3-9), as

$$\langle a(t)b(t)\rangle = \tfrac{1}{2}\mathrm{Re}[A^*B] \tag{1.3-15}$$

or in terms of the analytic form of $a(t)$ and $b(t)$ directly as

$$\langle \mathrm{Re}[a(t)]\mathrm{Re}[b(t)]\rangle = \tfrac{1}{2}\mathrm{Re}[a(t)^*b(t)] \tag{1.3-16}$$

where the asterisk superscript indicates the complex conjugate. The time dependence on the right side of Equation (1.3-16) disappears because both $a(t)$ and $b(t)$ have the same sinusoidal time dependence $\exp(i\omega t)$. These two results, Equations (1.3-15) and (1.3-16), are important and will find frequent use throughout the book. Equation (1.3-16) is also useful when the frequencies of the two functions are slightly different.

By using the complex formalism (or analytic representation) for the field vectors **E**, **H**, **D**, and **B**, the time-averaged Poynting vector (1.2-6) and the energy density (1.2-5) for sinusoidally varying fields are given by

$$\mathbf{S} = \tfrac{1}{2}\mathrm{Re}[\mathbf{E} \times \mathbf{H}^*] \tag{1.3-17}$$

and

$$U = \tfrac{1}{4}\mathrm{Re}[\mathbf{E} \cdot \mathbf{D}^* + \mathbf{H} \cdot \mathbf{B}^*] \tag{1.3-18}$$

respectively. The dipolar dissipation (1.2-9) for sinusoidally varying fields can be written

$$P_D = \tfrac{1}{2}\mathrm{Re}\left(\mathbf{E}^* \cdot \frac{\partial \mathbf{P}}{\partial t}\right) \tag{1.3-19}$$

If an electromagnetic field is written

$$\begin{aligned} \mathbf{E} &= \mathbf{E}_0 e^{i\omega t} \\ \mathbf{H} &= \mathbf{H}_0 e^{i\omega t} \end{aligned} \tag{1.3-20}$$

and the polarization **P** is written

$$\mathbf{P} = \varepsilon_0 \chi \mathbf{E}_0 e^{i\omega t} \tag{1.3-21}$$

where χ is the electric susceptibility coefficient, then the dipolar dissipation can be written

$$P_D = \tfrac{1}{2}\mathrm{Re}(i\omega\varepsilon_0\chi\mathbf{E}_0 \cdot \mathbf{E}_0^*) \tag{1.3-22}$$

In general, χ is a complex number that leads to a phase delay between the polarization **P** and the electric field **E**. The phase delay depends on the imaginary part of χ. If the material response to the electric field is instantaneous (i.e., no phase delay), then χ is a real number. This medium is lossless according to Equation (1.3-22). If the complex susceptibility is written as

$$\chi = \chi' - i\chi'' \tag{1.3-23}$$

where χ' is the real part and χ'' is the imaginary part, then the dipolar dissipation can be written

$$P_D = \tfrac{1}{2}\omega\varepsilon_0\mathbf{E}_0 \cdot \mathbf{E}_0^*\chi'' \tag{1.3-24}$$

We note that the imaginary part of the electric susceptibility is responsible for the dipolar dissipation.

1.4 WAVE EQUATIONS AND MONOCHROMATIC PLANE WAVES

Two of the most important results of Maxwell's equations are the wave equations and the existence of electromagnetic waves that are solutions to them. Here we derive the wave equations in isotropic material media and then consider the propagation of electromagnetic

plane waves in homogeneous and isotropic media, where both ε and μ are scalar constants. Vacuum is, of course, the best example of such a "medium." Glasses with uniform compositions are material media that can be treated as homogeneous and isotropic.

To derive the wave equations, we start from Maxwell's equations (1.1-1) to (1.1-4). We limit our attention to regions where both charge density ρ and current density $\mathbf{J}$ vanish. Using the constitutive equations $\mathbf{D} = \varepsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$, the first two of Maxwell's equations can be written

$$\nabla \times \mathbf{E} + \mu \frac{\partial \mathbf{H}}{\partial t} = 0 \tag{1.4-1}$$

$$\nabla \times \mathbf{H} - \varepsilon \frac{\partial \mathbf{E}}{\partial t} = 0 \tag{1.4-2}$$

We now take a curl operation of Equation (1.4-1) and eliminate $\nabla \times \mathbf{H}$ by using Equation (1.4-2). This leads to

$$\nabla \times (\nabla \times \mathbf{E}) + \mu\varepsilon \frac{\partial^2}{\partial t^2} \mathbf{E} = 0 \tag{1.4-3}$$

The first term in the above equation can be expanded as

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} \tag{1.4-4}$$

Using Equation (1.1-3) and $\mathbf{D} = \varepsilon\mathbf{E}$, we find that for homogeneous and isotropic media, the first term on the right side of Equation (1.4-4) is zero. Thus, Equation (1.4-3) becomes

$$\nabla^2 \mathbf{E} - \mu\varepsilon \frac{\partial^2}{\partial t^2} \mathbf{E} = 0 \tag{1.4-5}$$

This is the wave equation for the electric field vector $\mathbf{E}$ in homogeneous and isotropic media. A similar equation for the magnetic field vector $\mathbf{H}$ can be written

$$\nabla^2 \mathbf{H} - \mu\varepsilon \frac{\partial^2}{\partial t^2} \mathbf{H} = 0 \tag{1.4-6}$$

These are the standard electromagnetic wave equations. They are satisfied by the well-known monochromatic plane wave solution

$$\psi = Ae^{i(\omega t - \mathbf{k}\cdot\mathbf{r})} \tag{1.4-7}$$

where A is a constant and is often called the amplitude. In Equation (1.4-7), the angular frequency ω is related to the magnitude of the wavevector $\mathbf{k}$ by

$$|\mathbf{k}| = \omega\sqrt{\mu\varepsilon} \tag{1.4-8}$$

and ψ can be any Cartesian component of $\mathbf{E}$ and $\mathbf{H}$.

We now examine the meaning of this solution. The expression (1.4-7) represents the field component as a function of time and space. At any point in space, the field is a sinusoidal function of time. In addition, at each given moment, the field is a sinusoidal function of space. It is clear that the field has the same value for coordinates $\mathbf{r}$ and times t, which satisfy

$$\omega t - \mathbf{k} \cdot \mathbf{r} = \text{constant} \tag{1.4-9}$$

where the constant is arbitrary and determines the field value. The above equation determines a plane normal to the wavevector $\mathbf{k}$ at any instant of time. This plane is called a surface of

constant phase. The surfaces of constant phases are often referred to as wavefronts. The electromagnetic wave represented by Equation (1.4-7) is called a plane wave because all the wavefronts are planar. It is easily seen (from Equation (1.4-9)) that the surfaces of constant phase travel in the direction of **k** with a velocity whose magnitude is

$$v = \frac{\omega}{|\mathbf{k}|} \tag{1.4-10}$$

This is the phase velocity of the wave. If we set $t = 0$ in Equation (1.4-7) and examine the spatial variation of the wave amplitude, the separation between two neighboring field peaks (i.e., the wavelength) is

$$\lambda' = \frac{2\pi}{k} = 2\pi\frac{v}{\omega} \tag{1.4-11}$$

where $k = |\mathbf{k}|$ and the prime indicates the wavelength of light inside the medium. In optical electronics and optical communications, λ is reserved for the wavelength of light in a vacuum.

The value of the phase velocity is a property of the medium and can be expressed in terms of the dielectric constant ε and the magnetic permeability μ. From Equations (1.4-8) and (1.4-10), we obtain

$$v = \frac{1}{\sqrt{\mu\varepsilon}} \tag{1.4-12}$$

The phase velocity of the electromagnetic radiation in a vacuum is

$$c = \frac{1}{\sqrt{\mu_0\varepsilon_0}} = 299{,}792{,}458 \text{ m/s} \tag{1.4-13}$$

whereas in material media it has the value

$$v = \frac{c}{n} \tag{1.4-14}$$

where

$$n = \sqrt{\frac{\mu\varepsilon}{\mu_0\varepsilon_0}} \tag{1.4-15}$$

Most transparent media in optical electronics are nonmagnetic and have a magnetic permeability μ_0. In that case, $n = \sqrt{\varepsilon/\varepsilon_0}$ and is called the index of refraction of the medium. Table 1.1 lists the index of refraction of some common optical electronic materials. We must keep in mind, however, that ε, and therefore n, of a nonmagnetic material ($\mu = \mu_0$) are functions of the frequency ω. The variation of n with frequency ω gives rise to the well-known phenomenon of chromatic dispersion in optics. In a dispersive medium, the phase velocity of a light wave depends on the frequency ω. For example, red light travels faster than blue light in most optical glasses. The physical origin of the refractive index and its dispersion is discussed in some detail in Chapter 5.

We now turn our attention to the vector nature of the electromagnetic field and the requirements of satisfying Maxwell's equations. Using the complex formalism, we write the electromagnetic fields of the monochromatic plane wave in the forms

TABLE 1.1 Index of Refraction of Selected Materials

	Wavelength (μm)									
Material	0.488	0.5	0.5145	0.6328	1.0	1.3	1.55	3.0	5.0	10.6
As-S glass	2.786	2.77	2.75	2.606	2.478	2.449	2.437	2.416	2.407	2.378
BaF_2	1.478	1.4779	1.477	1.473	1.4686	1.467	1.466	1.4611	1.451	1.3928
$Bi_{12}GeO_{20}$			2.55	2.54						
$Bi_{12}SiO_{20}$				2.54						
$Bi_{12}TiO_{20}$				2.55						
CaF_2	1.437	1.4366	1.4362	1.433	1.429	1.427	1.426	1.418	1.399	1.2803
CdS, n_o			2.747	2.46	2.334			2.279	2.266	2.226
CdSe, n_o					2.55	2.50	2.48	2.454	2.446	2.43
CdTe					2.84			2.81	2.77	2.69
CsBr	1.712	1.710	1.707	1.694	1.679	1.675	1.673	1.670	1.668	1.663
CsI	1.810	1.806	1.802	1.781	1.757	1.751	1.749	1.744	1.742	1.739
CuBr	2.201	2.184	2.167	2.102						
CuCl	2.019	2.01	2.002	1.9613	1.924	1.92	1.91	1.903	1.901	1.90
CuI	2.437	2.422	2.405	2.321						
GaAs					3.5	3.41	3.38	3.35	3.29	3.135
GaP			3.66	3.38	3.17	3.07	3.05	2.97	2.94	2.90
GaSb								3.898	3.824	3.843
Ge								4.045	4.0163	4.0029
InAs									3.46	3.42
InP					3.327	3.21	3.17	3.11	3.08	3.05
InSb										3.95
Intran 1 (MgF_2)					1.378	1.377	1.376	1.364	1.337	
Intran 2 (ZnS)					2.291			2.256	2.245	2.1902
Intran 3 (CdF_2)					1.429			1.418	1.399	1.2817
Intran 4 (ZnSe)					2.485			2.44	2.432	2.4034
Intran 5 (MgO)					1.723			1.692	1.637	
KBr	1.572	1.57	1.568	1.558	1.5443	1.541	1.540	1.536	1.534	1.525
KCl	1.498	1.497	1.496	1.488	1.4799	1.478	1.477	1.474	1.471	1.454
KI	1.686	1.684	1.68	1.661	1.64	1.636	1.634	1.6284	1.626	1.6191
LiF	1.395	1.394	1.394	1.392	1.387	1.385	1.383	1.367	1.327	1.05
MgF_2		1.3703			1.363					
MgF_2 crystal, n_o	1.380	1.3798	1.379	1.377	1.374	1.372	1.371			
MgF_2 crystal, n_e	1.392	1.3917	1.391	1.389	1.385	1.384	1.382			
MgO		1.76			1.7237			1.6922	1.6262	
PbF_2	1.786	1.782	1.779	1.761	1.742			1.724	1.708	1.625
Quartz, crystal, n_o	1.55	1.549	1.548	1.543	1.535	1.531	1.528	1.50	1.417	
Ruby, Al_2O_3, n_o	1.776	1.775	1.774	1.766						
Sapphire, Al_2O_3, n_o	1.776	1.775	1.774	1.766	1.756	1.745	1.736	1.7115	1.6239	
Si						3.502	3.476	3.432	3.422	3.418
SiC, crystal	2.698	2.691	2.682	2.64	2.583					
SiO_2, fused silica		1.462	1.462	1.457	1.45	1.447	1.444	1.4185		
$SrTiO_3$, crystal	2.489	2.477	2.461	2.389	2.316			2.231	2.122	
Ta_2O_5	2.27	2.26	2.25	2.16	2.03	1.99				
TiO_2, rutile, n_o	2.731	2.712	2.694	2.59	2.484					
ZnO, crystal, n_o	2.064	2.051	2.044	1.99	1.944	1.932	1.928	1.9072		
ZnS		2.42	2.409	2.351	2.32	2.28	2.27	2.26	2.25	
ZnS, sphalerite	2.433	2.421	2.406	2.354	2.293	2.28	2.27			
ZnS wurtzite, n_o	2.43	2.421	2.41	2.352	2.301					
ZnSe		2.7	2.7	2.59	2.48	2.47	2.46	2.44	2.43	2.392
ZnTe				2.984	2.79	2.75	2.73	2.71	2.70	2.70

$$\mathbf{E} = \mathbf{u}_1 E_0 e^{i(\omega t - \mathbf{k}\cdot\mathbf{r})} \tag{1.4-16}$$

$$\mathbf{H} = \mathbf{u}_2 H_0 e^{i(\omega t - \mathbf{k}\cdot\mathbf{r})} \tag{1.4-17}$$

where $\mathbf{u}_1$ and $\mathbf{u}_2$ are two constant unit vectors, and E_0 and H_0 are complex amplitudes that are constant in space and time. In a homogeneous charge-free medium, the divergence Maxwell equations are $\nabla \cdot \mathbf{E} = 0$ and $\nabla \cdot \mathbf{H} = 0$, which, when applied to Equations (1.4-16) and (1.4-17), result in

$$\mathbf{u}_1 \cdot \mathbf{k} = \mathbf{u}_2 \cdot \mathbf{k} = 0 \tag{1.4-18}$$

This means that $\mathbf{E}$ and $\mathbf{H}$ are both perpendicular to the direction of propagation. For this reason, electromagnetic waves are said to be transverse waves. The transverse condition (1.4-18) holds for all four field vectors of the plane wave in homogeneous and isotropic media. In a general anisotropic charge-free medium, only the field vectors $\mathbf{D}$ and $\mathbf{B}$ of a plane wave are perpendicular to the direction of propagation.

The curl Maxwell equations provide further restrictions on the field vectors. These are obtained by substituting Equations (1.4-16) and (1.4-17) into Equation (1.1-1) and are given by

$$\mathbf{u}_2 = \frac{\mathbf{k} \times \mathbf{u}_1}{|\mathbf{k}|} \tag{1.4-19}$$

and

$$H_0 = \frac{E_0}{\eta}, \quad \eta = \sqrt{\frac{\mu}{\varepsilon}} \tag{1.4-20}$$

This shows that the triad $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{k})$ forms a set of mutually orthogonal vectors and that the field vectors $\mathbf{E}$ and $\mathbf{H}$ are in phase and in constant ratio provided ε and μ are real. The parameter η in Equation (1.4-20) has the dimension of resistance and is called the impedance of space. In a vacuum,

$$\eta_0 = \sqrt{\mu_0/\varepsilon_0} = 377\ \Omega$$

The plane wave we have just derived is a transverse wave propagating in the direction $\mathbf{k}$. It represents a time-averaged energy flux given, according to Poynting's theorem (1.3-17), by

$$\mathbf{S} = \frac{1}{2\eta}|E_0|^2 \mathbf{u}_3 = \frac{\mathbf{k}}{2\omega\mu}|E_0|^2 = \frac{\mathbf{k}}{2\omega\mu}|\mathbf{E}|^2 \tag{1.4-21}$$

where $\mathbf{u}_3 = \mathbf{u}_1 \times \mathbf{u}_2$ is a unit vector in the direction of $\mathbf{k}$. The time-averaged energy density is

$$U = \tfrac{1}{2}\varepsilon|E_0|^2 = \tfrac{1}{2}\varepsilon|\mathbf{E}|^2 \tag{1.4-22}$$

We note from Equations (1.4-21) and (1.4-22) that the energy flux is directed along the direction of propagation. We also note from Equations (1.4-13) and (1.4-15) that we have the relation

$$S = vU \tag{1.4-23}$$

This equation simply states that the energy is flowing at a speed $v = c/n$ along the direction of propagation. It is important to note that this is valid only in media without dispersion. When chromatic dispersion is present, the energy is flowing at a speed known as the group velocity.

1.5 CHROMATIC DISPERSION AND GROUP VELOCITY

In the previous section we discussed the plane wave solutions to Maxwell's equations and studied some of the basic properties. Only monochromatic waves with a definite frequency and wavenumber were treated. In the field of optical communications, short pulses of laser radiation are employed to transmit information. The finite duration of an optical pulse results in a finite spread of frequencies or, equivalently, wavelengths. The propagation of an optical pulse in a linear medium can be described in terms of the propagation of an appropriate superposition of plane waves with different frequencies, since Maxwell's equations are linear. In a linear medium, the induced polarization is proportional to the electric field. If the medium is dispersive, the phase velocity depends on the frequency. This is known as the chromatic dispersion. As a result of the chromatic dispersion, different frequency components of the pulse propagate with different speeds. This can lead to a change of the shape or even a spreading of the pulse. In addition, the velocity of energy flow of an optical pulse in a dispersive medium may be different from the phase velocity. This is a subtle subject and requires careful investigation.

These effects on the propagation of an optical pulse due to the chromatic dispersion can be described by representing the pulse as a sum of many monochromatic plane wave components. Each of these plane wave components is a solution of Maxwell's equations. Since the equations are linear, the sum is also a solution of Maxwell's equations. In general, we must replace the summation by an integration. For the sake of clarity in introducing the basic concepts, we will only consider the case of a beam of polarized light which is uniform in the transverse directions. If $A(k)$ denotes the amplitude of the monochromatic plane wave component with wavenumber k, then the electric field of the pulse can be written

$$E(z, t) = \int_{-\infty}^{\infty} A(k) e^{i[\omega(k)t - kz]}\, dk \tag{1.5-1}$$

where the angular frequency is written $\omega(k)$ to show its dependence on k. This integral satisfies Maxwell's equations, since the integrand is a basic monochromatic plane wave solution to the same equations. Note that if we view $E(z, t)$ at some instant of time, say, $t = 0$, then $A(k)$ is formally the Fourier transform of $E(z, 0)$. In other words,

$$E(z, 0) = \int_{-\infty}^{\infty} A(k) e^{-ikz}\, dk \tag{1.5-2}$$

We will then refer to $|A(k)|^2$ as the Fourier spectrum (or power spectrum) of $E(z, t)$.

The relation between ω and k (known as the dispersion relation) is given by Equation (1.4-8) for the electromagnetic field. In the following discussion, we assume that both k and $\omega(k)$ are real. A typical optical pulse and its Fourier spectrum are shown in Figure 1.2.

A beam of polarized optical pulse is usually characterized by its center frequency ω_0 (or its corresponding wavenumber k_0), and the frequency spread $\Delta\omega$ around ω_0 (or the corresponding spread in wavenumber Δk). Typically, $A(k)$ is sharply peaked around k_0 (i.e., $\Delta k \ll k_0$). To study the propagation of such an optical pulse in space and time, we expand $\omega(k)$ around the value k_0 in terms of a Taylor series,

$$\omega(k) = \omega_0 + \left(\frac{d\omega}{dk}\right)_{\omega_0} (k - k_0) + \cdots \tag{1.5-3}$$

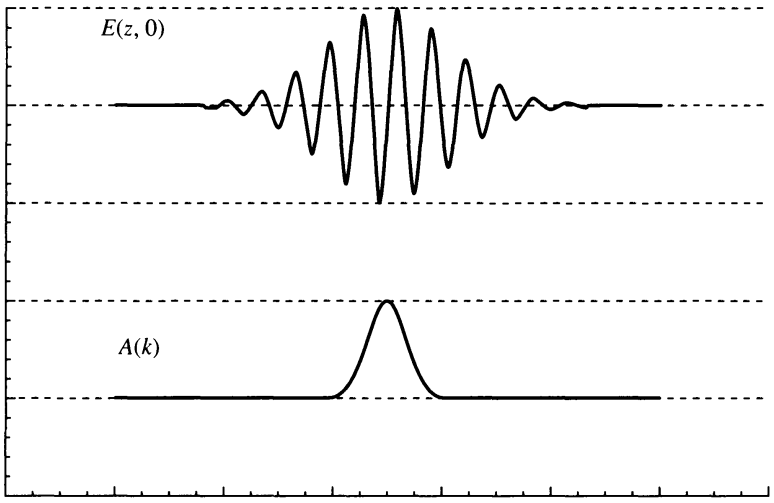


Figure 1.2 An optical pulse of finite extent and its Fourier spectrum in wavenumber (k) space.

and then substitute for ω from Equation (1.5-3) into Equation (1.5-1). Thus Equation (1.5-2) can be written

$$E(z, t) = e^{i(\omega_0 t - k_0 z)} \int_{-\infty}^{\infty} A(k) \exp\left\{ i \left[\left(\frac{d\omega}{dk} \right)_{\omega_0} t - z \right] (k - k_0) \right\} dk \tag{1.5-4}$$

where we neglect the higher-order terms in $(k - k_0)$. The integral in Equation (1.5-4) is a function of the composite variable

$$\xi = \left[z - \left(\frac{d\omega}{dk} \right)_{\omega_0} t \right] \tag{1.5-5}$$

only and may be called the envelope function $V(\xi)$. Thus the amplitude of the optical pulse can be written

$$E(z, t) = e^{i(\omega_0 t - k_0 z)} V \left[z - \left(\frac{d\omega}{dk} \right)_{\omega_0} t \right] \tag{1.5-6}$$

This shows that, apart from an overall phase factor, the optical pulse travels along undistorted in shape, with a velocity

$$v_g = \left(\frac{d\omega}{dk} \right)_{\omega_0} \tag{1.5-7}$$

which is called the group velocity of the pulse. Figure 1.3 shows the propagation of an optical pulse in a dispersive medium in this approximation. The approximation is legitimate only when the distribution is sharply peaked at k_0 and the frequency ω is a smoothly varying function of k around k_0. If the electromagnetic energy density of an optical pulse is associated with the absolute square of the amplitude, it is obvious that in this approximation the group velocity represents the speed of transport of energy. We note that, in the case of a pulse, the phase velocity and group velocity are, in general, different. The phase velocity, which is usually greater than the group velocity, has the same significance as in the case of a plane wave. It is the velocity needed to stay on top of a given wave crest or valley.

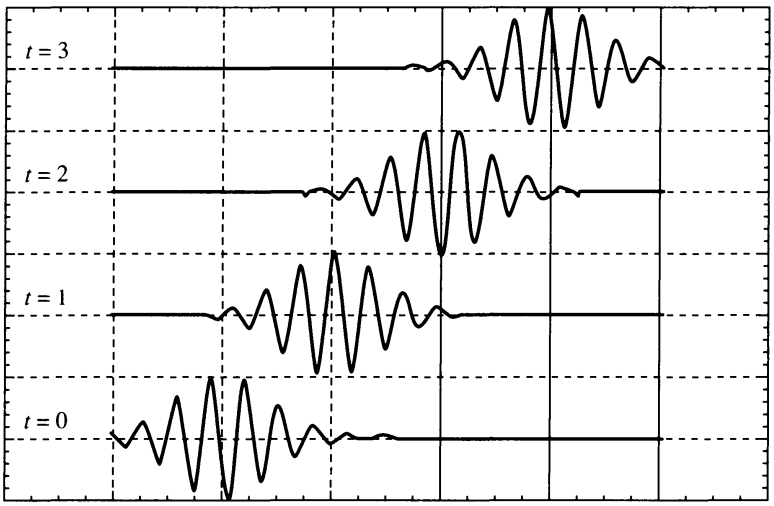


Figure 1.3 Propagation of an optical pulse.

In optical communications, optical pulses are often propagating in silica fibers, which exhibit chromatic dispersion. The dispersion is usually described by the index of refraction, $n(\omega)$, as a function of frequency (or wavelength), and the relation between ω and k is given by

$$k = n(\omega)\frac{\omega}{c} \tag{1.5-8}$$

where c is the speed of light in vacuum. The phase velocity is

$$v_p = \frac{c}{n(\omega)} \tag{1.5-9}$$

and is greater than or smaller than c depending on whether $n(\omega)$ is smaller or larger than unity. From Equations (1.5-7) and (1.5-8), the group velocity is given by

$$v_g = \frac{c}{n + \omega(dn/d\omega)} \tag{1.5-10}$$

For normal dispersion, $dn/d\omega > 0$, the group velocity is less than the phase velocity. In regions of anomalous dispersion, however, $dn/d\omega$ can become large and negative. Then the group velocity differs greatly from the phase velocity and can be greater than c. The latter case occurs only when $dn/d\omega$ is very negative, which is equivalent to a rapid variation of ω as a function of k, thus making the approximation illegitimate. In most transparent media, the materials exhibit normal dispersion. Anomalous dispersion occurs usually in the spectral regime within the absorption bands.

Equation (1.5-10) is often written, equivalently, in terms of wavelength as

$$v_g = \frac{c}{n - \lambda(dn/d\lambda)} \tag{1.5-11}$$

Group Velocity Dispersion and Pulse Spreading

We have seen in an earlier discussion that an optical pulse propagating in a dispersive medium can remain undistorted in shape provided the approximation (1.5-3) is valid. In high-speed optical communications, the pulse width can be on the order of picoseconds. In this case, the amplitude $A(k)$ is no longer sharply peaked around the central wavenumber k_0. Consequently, the next higher-order term

$$\frac{1}{2}\left(\frac{d^2\omega}{dk^2}\right)_{\omega_0}(k-k_0)^2 \tag{1.5-12}$$

cannot be neglected. The pulse will no longer remain unchanged and will, in general, spread as the pulse travels. The spreading of the pulses in optical communication networks can lead to a degradation or loss of the signal transmission.

The spreading of the optical pulse can be accounted for by noting that the group velocity v_g may not be the same for each spectral component of the pulse. This is known as the group velocity dispersion (GVD). If Δk is the width of the spread in wavenumbers, then the group velocity spread is on the order of

$$\Delta v_g = \left(\frac{d^2\omega}{dk^2}\right)_{\omega_0}\Delta k \tag{1.5-13}$$

As the optical pulse propagates, one expects a spread in position on the order of $\Delta v_g\, T$, where T is the time of flight of a reference frequency component. The group velocity spread is often written

$$\Delta v_g = \left(\frac{dv_g}{d\omega}\right)_{\omega_0}\Delta\omega \tag{1.5-14}$$

where $\Delta\omega$ is the width of spread of the angular frequency, or equivalently,

$$\Delta v_g = \left(\frac{dv_g}{d\lambda}\right)_{\omega_0}\Delta\lambda \tag{1.5-15}$$

where $\Delta\lambda$ is the width of spread of the wavelength.

In optical communications, signals are represented in terms of a sequence of pulses. Each of the pulses has a well-defined temporal duration. For example, in a 40 Gb/s transmission system, the pulse duration is around 25 ps. The pulse spreading in terms of picosecond is an important measurement of the signal integrity. To minimize the bit errors in transmission, the pulse spreading is often required to be less than a fraction of the pulse duration. Thus, in fiber-optic communications, the group velocity dispersion is often characterized by a parameter

$$D = \frac{1}{L}\frac{dT}{d\lambda} \tag{1.5-16}$$

where T is the pulse transmission time through a length L of the fiber. Physically, D is a measure of the pulse spreading (in units of seconds) per unit bandwidth per unit length of the medium of transmission. Using $T = L/v_g$, it can be shown that

$$T = L\left(\frac{n}{c} + \frac{\omega}{c}\frac{dn}{d\omega}\right) = L\left(\frac{n}{c} - \frac{\lambda}{c}\frac{dn}{d\lambda}\right) \tag{1.5-17}$$

and

$$D = -\frac{1}{c\lambda}\left(\lambda^2\frac{d^2n}{d\lambda^2}\right) \tag{1.5-18}$$

or equivalently,

$$D = -\frac{2\pi c}{\lambda^2}\frac{d^2k}{d\omega^2} \tag{1.5-19}$$

We note that the group velocity dispersion is related to the second-order derivative of the index of refraction with respect to frequency (or wavelength).

EXAMPLE: DISPERSION IN SILICA (SiO_2)

Most optical fibers for communications are made of silica glass. Pure silica glass exhibits extremely low absorption loss. The material, however, exhibits a chromatic dispersion. An accurate empirical expression of the refractive index of fused silica can be written

$$n(\lambda) = c_0 + c_1\lambda^2 + c_2\lambda^4 + \frac{c_3}{(\lambda^2 - a)} + \frac{c_4}{(\lambda^2 - a)^2} + \frac{c_5}{(\lambda^2 - a)^3} \tag{1.5-20}$$

where

$$\begin{aligned} c_0 &= 1.4508554 \\ c_1 &= -0.0031268 \\ c_2 &= -0.0000381 \\ c_3 &= 0.0030270 \\ c_4 &= -0.0000779 \\ c_5 &= 0.0000018 \\ a &= 0.035 \end{aligned}$$

and λ is in units of micrometers.

Figure 1.4 shows the index of refraction of silica glass as a function of wavelength. We note that the index of refraction is a decreasing function of wavelength. This is typical in the normal dispersion regime.

Using Equation (1.5-11), we obtain the group velocity of optical propagation in silica glass. Figure 1.5 shows v_g/c as a function of wavelength. It is clear that the group velocity depends

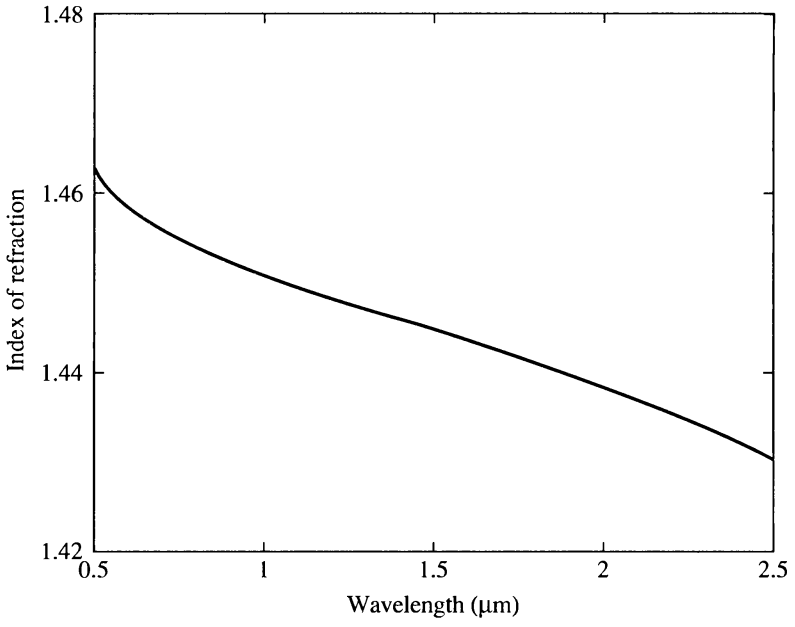


Figure 1.4 The index of refraction of pure silica glass as a function of wavelength.

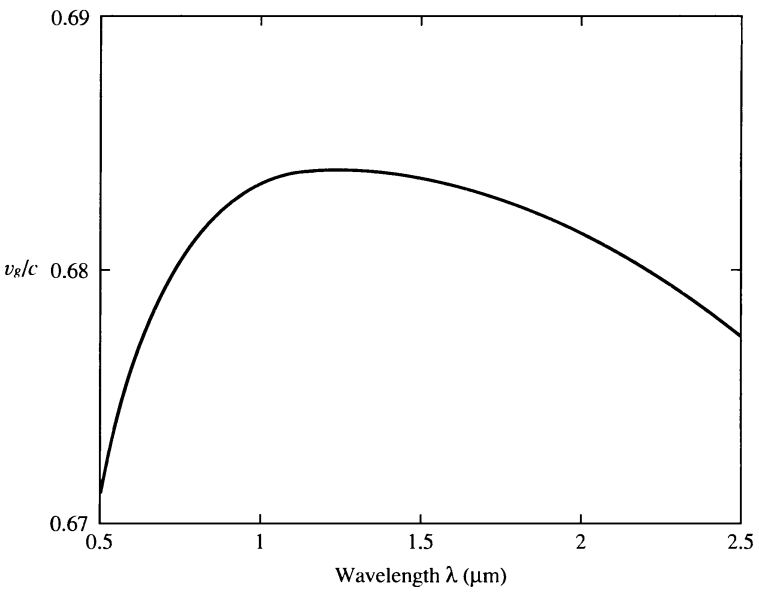


Figure 1.5 Group velocity in units of c in silica glass as a function of wavelength.

on the wavelength. The group velocity of light in pure silica exhibits a maximum at a wavelength near $\lambda = 1.275$ μm. For optical transmission of a pulse in the spectral regime around $\lambda = 1.55$ μm, the short-wavelength components travel faster than the long-wavelength components of the optical pulse. The disparity in the group velocity leads to pulse broadening and a descending frequency chirp in the pulse.

Using Equation (1.5-18), we obtain the group velocity dispersion in silica glass. Figure 1.6 shows the group velocity dispersion D as a function of wavelength. We note that D becomes zero at a wavelength near $\lambda = 1.275$ μm. The group velocity dispersion due to bulk silica glass is around 20 ps/nm-km in the spectral regime around $\lambda = 1500$ nm.

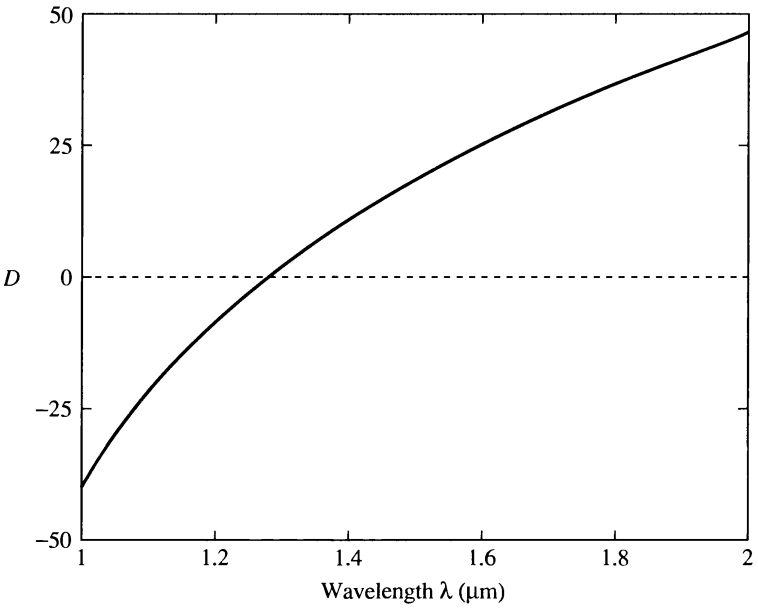


Figure 1.6 Group velocity dispersion of silica glass in units of ps/nm-km as a function of wavelength.

1.6 POLARIZATION STATES AND REPRESENTATIONS (STOKES PARAMETERS AND POINCARÉ SPHERE)

In the electromagnetic theory of light, beams of light are represented by electromagnetic waves propagating in space. A beam of light is often represented by its electric field vector, which oscillates in time and space as the beam propagates. In an isotropic medium, the direction of oscillation (or vibration) is always perpendicular to the direction of propagation. Being a transverse wave, there are two independent directions of oscillation. In an isotropic medium (e.g., glass, vacuum), the two mutually independent directions of oscillation can be chosen arbitrarily. If these two components of oscillation are totally uncorrelated, the resultant direction of oscillation is random and the beam of light is said to be unpolarized. For all thermal sources in nature, the direction of vibration is random. A beam of light is said to be linearly polarized if its electric field vector vibrates in one particular direction. In this section, we describe the polarization states of light and various ways of representing the states of polarization.

The main objective of this book is to provide comprehensive coverage of optical electronics and its application in optical communications. Many of the components in optical communications involve the use of polarized light and anisotropic optical media. Here we introduce the concept of polarization by considering a beam of monochromatic plane waves propagating in an isotropic and homogeneous medium. The beam can be represented by its electric field $\mathbf{E}(\mathbf{r}, t)$, which can be written

$$\mathbf{E} = \mathbf{A}\cos(\omega t - \mathbf{k}\cdot\mathbf{r}) \tag{1.6-1}$$

where ω is the angular frequency, $\mathbf{k}$ is the wavevector, and $\mathbf{A}$ is a constant vector representing the amplitude. The magnitude of the wavevector k is related to the frequency by the following equation:

$$k = n\frac{\omega}{c} = n\frac{2\pi}{\lambda} \tag{1.6-2}$$

where n is the index of refraction of the medium, c is the speed of light in vacuum, and λ is the wavelength of light in vacuum. For materials with some absorption, the index of refraction is a complex number. Reflecting the transverse nature, the electric field vector is always perpendicular to the direction of propagation,

$$\mathbf{k}\cdot\mathbf{E} = 0 \tag{1.6-3}$$

For mathematical simplicity, the monochromatic plane wave in Equation (1.6-1) is often written

$$\mathbf{E} = \mathbf{A}\exp[i(\omega t - \mathbf{k}\cdot\mathbf{r})] \tag{1.6-4}$$

with the understanding that only the real part of the right side represents the actual electric field. This is the analytic representation described in Section 1.3. The polarization state of a beam of monochromatic light is specified by its electric field vector $\mathbf{E}(\mathbf{r}, t)$. The time evolution of the electric field vector is exactly sinusoidal; that is, the electric field must oscillate at a definite frequency. For the purpose of describing various representations of the polarization states, we consider propagation along the z axis. Being a transverse wave, the electric field vector must lie in the xy plane. The two mutually independent components of the electric field vector can be written

$$E_x = A_x \cos(\omega t - kz + \delta_x)$$
$$E_y = A_y \cos(\omega t - kz + \delta_y) \tag{1.6-5}$$

where we have used two independent and positive amplitudes A_x and A_y and have added two independent phases δ_x and δ_y to reflect the mutual independence of the two components. With the amplitudes being positive, the phase angles are defined in the range $-\pi < \delta_{x,y} \leq \pi$. Since the x component and the y component of the electric field vector can oscillate independently at a definite frequency, one must consider the effect produced by the vector addition of these two oscillating orthogonal components. The problem of superposing two independent oscillations at right angles to each other and with the same frequency is well known and is completely analogous to the classical motion of a two-dimensional harmonic oscillator. The orbit of a general motion is an ellipse, which corresponds to oscillations in which the x and y components are not in phase. For optical waves, this corresponds to elliptic polarization states. There are, of course, many special cases that are important in optics. These include linear polarization states and circular polarization states.

We start with the discussion of two special cases of interest. Without loss of generality, we consider the time evolution of the electric field vector at the origin $z = 0$. According to Equation (1.6-5), the electric field components are written

$$E_x = A_x \cos(\omega t + \delta_x)$$
$$E_y = A_y \cos(\omega t + \delta_y) \tag{1.6-6}$$

We define a relative phase as

$$\delta = \delta_y - \delta_x \tag{1.6-7}$$

where, again, δ is limited to the region $-\pi < \delta \leq \pi$.

Linear Polarization States

A beam of light is said to be linearly polarized if the electric field vector vibrates in a constant direction (in the xy plane). This occurs when the two components of oscillation are in phase ($\delta = \delta_y - \delta_x = 0$) or π out of phase ($\delta = \delta_y - \delta_x = \pi$); that is,

$$\delta = \delta_y - \delta_x = 0 \quad \text{or} \quad \pi \tag{1.6-8}$$

In this case, the electric field vector vibrates sinusoidally along a constant direction in the xy plane defined by the ratio of the two components,

$$\frac{E_y}{E_x} = \frac{A_y}{A_x} \quad \text{or} \quad -\frac{A_y}{A_x} \tag{1.6-9}$$

Since the amplitudes A_x and A_y are independent, the electric field vector of linearly polarized light can vibrate along any direction in the xy plane. Linearly polarized light is often called plane polarized light. If we examine the space evolution of the electric field vector at a fixed point in time (say, $t = 0$), the components of the electric field vector can be written

$$E_x = A_x \cos(-kz + \delta_x)$$
$$E_y = A_y \cos(-kz + \delta_y) \tag{1.6-10}$$

with $\delta = \delta_y - \delta_x = 0$ or π. We note that the sinusoidal curve traced by the components in space is confined in a plane defined by Equation (1.6-9). The vibration of the electric

field vector is confined in this plane. Thus the beam of light is said to be plane polarized. The terms plane polarized light and linearly polarized light are interchangeable. Linear polarization states are most widely used in optics because of their simplicity and ease of preparation.

Circular Polarization States

The other special case of importance is that of the circular polarization state. A beam of light is said to be circularly polarized if the electric field vector undergoes uniform rotation in the xy plane. This occurs when $A_x = A_y$ and

$$\delta = \delta_y - \delta_x = \pm\pi/2 \tag{1.6-11}$$

According to our convention, the beam of light is right-handed circularly polarized when $\delta = -\pi/2$, which corresponds to a counterclockwise rotation of the electric field vector in the xy plane; and left-handed circularly polarized when $\delta = \pi/2$, which corresponds to a clockwise rotation of the electric field vector in the xy plane. Our convention for labeling right-handed and left-handed polarization is consistent with the terminology of modern physics in which a photon with a right-handed circular polarization has a positive angular momentum along the direction of propagation (see Problem 1.36). However, some optics books adopt the opposite convention. The opposite convention arises from the description of the evolution of the electric field vector in space (see Problem 1.16).

It is interesting to note that the conditions of equal amplitude and $\pm\pi/2$ phase shift for circular polarization states are valid in any set of perpendicular coordinates in the xy plane. In other words, when the electric field vector of a circularly polarized light is decomposed into any two mutually perpendicular components, the amplitudes are always equal and the phase shift is always $\pm\pi/2$.

Elliptic Polarization States

A beam of light is said to be elliptically polarized if the curve traced by the end point of the electric field vector is an ellipse (in the xy plane). This is the most general case of a polarized light. Both linear polarization states and circular polarization states are special cases of elliptic polarization states. At a given point in space (say, $z = 0$), Equation (1.6-5) is a parametric representation of an ellipse traced by the end point of the electric field vector. The equation of the ellipse can be obtained by eliminating ωt in Equation (1.6-6). After several steps of elementary algebra, we obtain

$$\left(\frac{E_x}{A_x}\right)^2 + \left(\frac{E_y}{A_y}\right)^2 - 2\frac{\cos\delta}{A_x A_y}E_x E_y = \sin^2\delta \tag{1.6-12}$$

Equation (1.6-12) is an equation of a conic. From Equation (1.6-6) it is obvious that this conic section is confined in a rectangular region with sides parallel to the coordinate axes and whose lengths are $2A_x$ and $2A_y$. Therefore, the curve must be an ellipse. Thus, we find that the polarization states of light are, in general, elliptical. A complete description of an elliptical polarization state includes the orientation of the ellipse relative to the coordinate axes and the shape and sense of revolution of the electric field vector. In general, the principal axes of the ellipse are not in the x and y directions. By using a transformation (rotation) of the coordinate

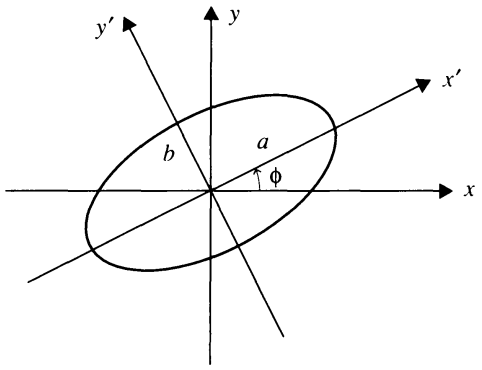


Figure 1.7 Polarization ellipse.

system, we are able to diagonalize Equation (1.6-12). Let x' and y' be the new set of axes along the principal axes of the ellipse. Then the equation of the ellipse in this new coordinate system becomes

$$\left(\frac{E_{x'}}{a}\right)^2 + \left(\frac{E_{y'}}{b}\right)^2 = 1 \tag{1.6-13}$$

where a and b are the lengths of the principal semiaxes of the ellipse, and $E_{x'}$ and $E_{y'}$ are the components of the electric field vector in this principal coordinate system.

Let ϕ be the angle between the x' axis and the x axis (see Figure 1.7). Then the lengths of the principal axes are given by

$$\begin{aligned} a^2 &= A_x^2 \cos^2\phi + A_y^2 \sin^2\phi + 2A_xA_y \cos\delta \cos\phi \sin\phi \\ b^2 &= A_x^2 \sin^2\phi + A_y^2 \cos^2\phi - 2A_xA_y \cos\delta \cos\phi \sin\phi \end{aligned} \tag{1.6-14}$$

The angle ϕ can be expressed in terms of A_x, A_y, and $\cos\delta$ as

$$\tan 2\phi = \frac{2A_xA_y}{A_x^2 - A_y^2}\cos\delta \tag{1.6-15}$$

It is important to note that $\phi + \pi/2$ is also a solution, if ϕ is a solution of the equation. The sense of revolution of an elliptical polarization is determined by the sign of $\sin\delta$. The end point of the electric vector will revolve in a clockwise direction if $\sin\delta > 0$ and in a counterclockwise direction if $\sin\delta < 0$. Figure 1.8 illustrates how the polarization ellipse changes with varying phase difference δ.

The ellipticity of a polarization ellipse is defined as

$$e = \pm\frac{b}{a} \tag{1.6-16}$$

where a and b are the lengths of the principal semiaxes. The ellipticity is taken as positive when the rotation of the electric field vector is right-handed and negative otherwise. With this definition, $e = \pm 1$ for circularly polarized light.

As discussed earlier, light is linearly polarized when the tip of the electric field vector **E** moves along a straight line. When it describes an ellipse, the light is elliptically polarized. When it describes a circle, the light is circularly polarized. If the end point of the electric field vector is seen to move in a counterclockwise direction by an observer facing the approaching wave, the field is said to possess right-handed polarization. Figure 1.8 also illustrates the sense of revolution of the ellipse.

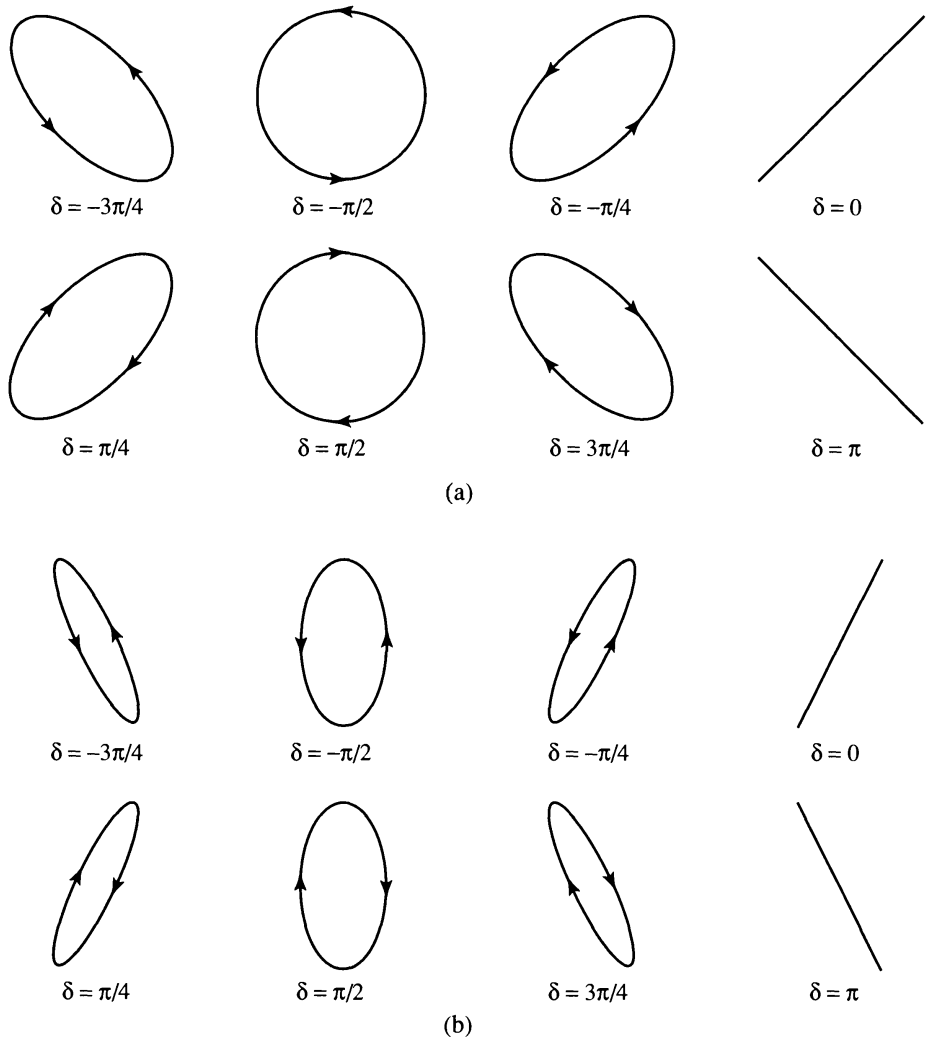


Figure 1.8 Polarization ellipses at various phase angles δ: (a) $E_x = \cos(\omega t - kz)$, $E_y = \cos(\omega t - kz + \delta)$; (b) $E_x = \frac{1}{2}\cos(\omega t - kz)$, $E_y = \cos(\omega t - kz + \delta)$.

An elliptic polarization state can always be decomposed into two mutually orthogonal components. The relative phase shift between these two components can be anywhere between $-\pi$ and π. However, in the principal coordinate system, the relative phase shift between the two orthogonal components is always $-\pi/2$ or $\pi/2$, depending on the sense of revolution.

Complex-Number Representation

From the previous discussion we found how the polarization state of a beam of light can be described in terms of the amplitudes and the phase angles of the x and y components of the electric field vector. In fact, all the information about the polarization state of a wave is contained in the complex amplitude **A** of the plane wave (Equation (1.6-4)). Therefore, a complex number χ defined as

$$\chi = e^{i\delta} \tan \psi = \frac{A_y}{A_x} e^{i(\delta_y - \delta_x)} \tag{1.6-17}$$

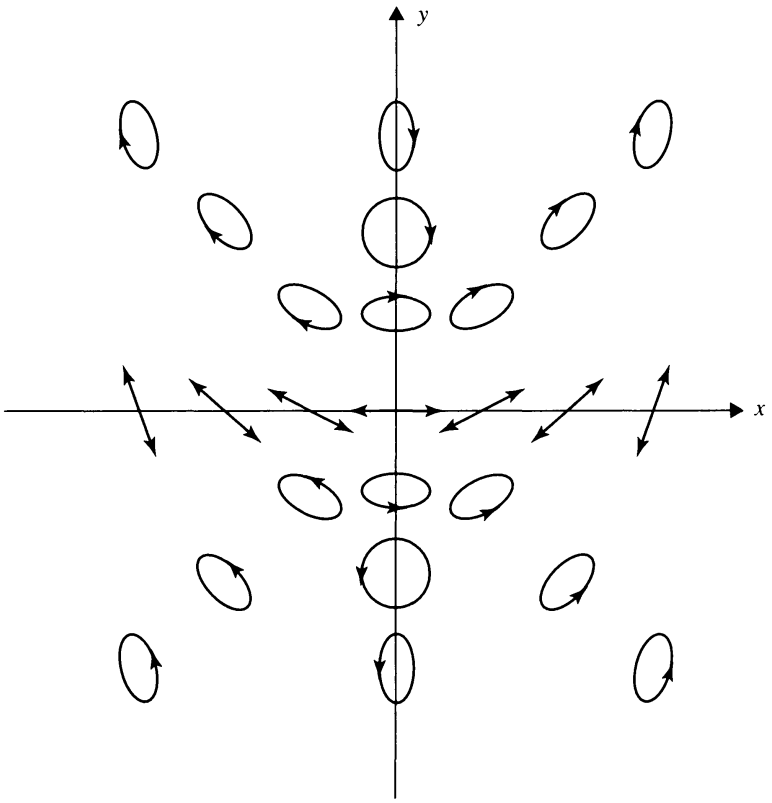


Figure 1.9 Complex-number representation of polarization states. Each point on the complex plane represents a unique state of polarization. The origin corresponds to a linearly polarized state along the x direction. Point (1, 0) represents a linearly polarized state with an azimuth angle of 45°, (0, 1) represents a left-handed circularly (LHC) polarized state, and (0, −1) represents a right-handed circularly (RHC) polarized state.

is sufficient to describe the polarization states. The angle ψ is defined to be between 0 and $\pi/2$. A complete description of the ellipse of polarization, which includes the orientation, sense of revolution, and ellipticity (see Equation (1.6-16)), can be expressed in terms of δ and ψ. Figure 1.9 illustrates various different polarization states in the complex plane. It can be seen from the figure that all the right-handed elliptical polarization states are in the lower half of the plane, whereas the left-handed elliptical polarization states are in the upper half of the plane. The origin corresponds to a linear polarization state with direction of oscillation parallel to the x axis. Thus each point on the complex plane represents a unique polarization state. Each point on the x axis represents a linearly polarized state with different azimuth angles of oscillation. Only two points $(0, \pm 1)$ correspond to circular polarization. Each point of the rest of the complex plane corresponds to a unique elliptical polarization state.

The inclination angle ϕ and the ellipticity angle θ ($\theta \equiv \tan^{-1} e$) of the polarization ellipse corresponding to a given complex number χ are given by

$$\tan 2\phi = \frac{2\text{Re}[\chi]}{1 - |\chi|^2} = \tan 2\psi \cos \delta \tag{1.6-18}$$

and

$$\sin 2\theta = -\frac{2\text{Im}[\chi]}{1 + |\chi|^2} = -\sin 2\psi \sin \delta \tag{1.6-19}$$

Jones Vector Representation

The Jones vector, introduced in 1941 by R. C. Jones [1], is very convenient for description of the polarization state of a plane wave. In this representation, the plane wave (1.6-4) is expressed in terms of its complex amplitudes as a column vector

$$\mathbf{J} = \begin{bmatrix} A_x e^{i\delta_x} \\ A_y e^{i\delta_y} \end{bmatrix} \tag{1.6-20}$$

Note that the Jones vector is a complex vector; that is, its elements are complex numbers. **J** is not a vector in the real physical space; rather, it is a vector in an abstract mathematical space. To obtain, as an example, the real x component of the electric field, we must perform the operation $E_x(t) = \text{Re}[J_x e^{i\omega t}] = \text{Re}[A_x e^{i(\omega t+\delta_x)}]$.

The Jones vector contains complete information about the amplitudes and phases of the electric-field-vector components. It thus specifies the polarization state of the wave uniquely. If we are only interested in the polarization state of the wave, it is convenient to use the normalized Jones vector, which satisfies the condition

$$\mathbf{J}^* \cdot \mathbf{J} = 1 \tag{1.6-21}$$

where the asterisk ($*$) denotes complex conjugation. Thus a beam of linearly polarized light with the electric field vector oscillating along a given direction can be represented by the Jones vector

$$\begin{bmatrix} \cos\psi \\ \sin\psi \end{bmatrix} \tag{1.6-22}$$

where ψ is the azimuth angle of the oscillation direction with respect to the x axis. The state of polarization that is orthogonal to the state represented by Equation (1.6-22) can be obtained by substituting $\psi + \pi/2$ for ψ, leading to a Jones vector

$$\begin{bmatrix} -\sin\psi \\ \cos\psi \end{bmatrix} \tag{1.6-23}$$

The special case, when $\psi = 0$, represents linearly polarized waves whose electric field vector oscillates along the coordinate axes. These Jones vectors are given by

$$\hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{y}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{1.6-24}$$

Jones vectors for the right- and left-handed circularly polarized light waves are given by

$$\mathbf{R} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix} \tag{1.6-25}$$

$$\mathbf{L} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix} \tag{1.6-26}$$

These two states of circular polarizations are mutually orthogonal in the sense that

$$\mathbf{R}^* \cdot \mathbf{L} = 0 \tag{1.6-27}$$

Since the Jones vector is a column matrix of rank 2, any pair of orthogonal Jones vectors can be used as a basis of the mathematical space spanned by all the Jones vectors. Any polarization

state can be represented as a superposition of two mutually orthogonal polarization states $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, or **R** and **L**. In particular, we can resolve the basic linear polarization states $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ into two circular polarization states **R** and **L** and vice versa. These relations are given by

$$\mathbf{R} = \frac{1}{\sqrt{2}}(\hat{\mathbf{x}} - i\hat{\mathbf{y}}) \tag{1.6-28}$$

$$\mathbf{L} = \frac{1}{\sqrt{2}}(\hat{\mathbf{x}} + i\hat{\mathbf{y}}) \tag{1.6-29}$$

$$\hat{\mathbf{x}} = \frac{1}{\sqrt{2}}(\mathbf{R} + \mathbf{L}) \tag{1.6-30}$$

$$\hat{\mathbf{y}} = \frac{i}{\sqrt{2}}(\mathbf{R} - \mathbf{L}) \tag{1.6-31}$$

Circular polarization states are seen to consist of linear oscillations along the x and y directions with equal amplitude $1/\sqrt{2}$, but with a phase difference of $\pi/2$. Similarly, a linear polarization state can be viewed as a superposition of two oppositely sensed circular polarization states.

We have so far discussed only the Jones vectors of some simple special cases of polarization. It is easy to show that a general elliptic polarization state can be represented by the following Jones vector:

$$\mathbf{J}(\psi, \delta) = \begin{bmatrix} \cos\psi \\ e^{i\delta}\sin\psi \end{bmatrix} \tag{1.6-32}$$

This Jones vector represents the same polarization state as the one represented by the complex number $\chi = e^{i\delta}\tan\psi$. Table 1.2 shows the Jones vectors of some typical polarization states.

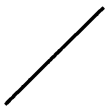
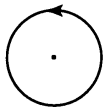
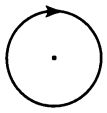
The most important application of Jones vectors is in conjunction with the Jones calculus. This is a powerful technique used for studying the propagation of plane waves with arbitrary states of polarization through an arbitrary sequence of birefringent elements and polarizers. This topic will be considered in some detail later in this chapter.

Stokes Parameters and Partially Polarized Light

By virtue of its nature, a monochromatic plane wave must be polarized; that is, the end point of its electric field vector at each point in space must trace out periodically an ellipse or one of its special forms, such as a circle or a straight line. However, if the light is not absolutely monochromatic, the amplitudes and relative phase δ between the x and y components can both vary with time, and the electric field vector will first vibrate in one ellipse and then in another. As a result, the polarization state of a polychromatic plane wave may be constantly changing. If the polarization state changes more rapidly than the speed of observation, we say the light is partially polarized or unpolarized depending on the time-averaged behavior of the polarization state. In optical electronics, one often deals with light having oscillation frequencies of about $10^{14}\ \text{s}^{-1}$, whereas the polarization state may change in a time period of 10^{-8} s depending on the nature of the light source.

We will limit ourselves to the case of quasi-monochromatic waves, whose frequency spectrum is confined to a narrow bandwidth $\Delta\omega$ (i.e., $\Delta\omega \ll \omega$). Such a wave can still be described by Equation (1.6-4), provided we relax the constancy condition of the amplitude **A**. Now ω denotes the center frequency, and the complex amplitude **A** is a function of time. Because the

TABLE 1.2 Various Representations of Polarization States

Polarization Ellipse	Jones Vector	(δ, ψ)	(ϕ, θ)	Stokes Vector
	$\begin{bmatrix}1\\0\end{bmatrix}$	$(0, 0)$	$(0, 0)$	$\begin{bmatrix}1\\1\\0\\0\end{bmatrix}$
	$\begin{bmatrix}0\\1\end{bmatrix}$	$(0, \pi/2)$	$(\pi/2, 0)$	$\begin{bmatrix}1\\-1\\0\\0\end{bmatrix}$
	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$	$(0, \pi/4)$	$(\pi/4, 0)$	$\begin{bmatrix}1\\0\\1\\0\end{bmatrix}$
	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$	$(\pi, \pi/4)$	$(-\pi/4, 0)$	$\begin{bmatrix}1\\0\\-1\\0\end{bmatrix}$
	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-i\end{bmatrix}$	$(-\pi/2, \pi/4)$	$(0, \pi/4)$	$\begin{bmatrix}1\\0\\0\\-1\end{bmatrix}$
	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\i\end{bmatrix}$	$(\pi/2, \pi/4)$	$(0, -\pi/4)$	$\begin{bmatrix}1\\0\\0\\1\end{bmatrix}$

bandwidth is narrow, $\mathbf{A}(t)$ may change only by a relatively small amount in a time interval $1/\Delta\omega$, and in this sense it is a slowly varying function of time. However, if the time constant of the detector, τ_D, is greater than $1/\Delta\omega$, $\mathbf{A}(t)$ may change significantly in a time interval τ_D. Although the amplitudes and phases are irregularly varying functions of time, certain correlations may exist among them.

To describe the polarization state of this type of radiation, we introduce the following time-averaged quantities:

$$\begin{aligned} S_0 &= \langle\langle A_x^2 + A_y^2 \rangle\rangle \\ S_1 &= \langle\langle A_x^2 - A_y^2 \rangle\rangle \\ S_2 &= 2\langle\langle A_x A_y \cos\delta \rangle\rangle \\ S_3 &= 2\langle\langle A_x A_y \sin\delta \rangle\rangle \end{aligned} \tag{1.6-33}$$

where the amplitudes A_x and A_y and the relative phase δ are assumed to be time dependent, and the double brackets denote averages performed over a time interval τ_D that is the characteristic time constant of the detection process. These four quantities are known as the *Stokes parameters* of a quasi-monochromatic plane wave. Note that all four quantities have the same dimension of intensity. It can be shown that the Stokes parameters satisfy the relation

$$S_1^2 + S_2^2 + S_3^2 \le S_0^2 \tag{1.6-34}$$

where the equality sign holds only for polarized waves.

It is a simple exercise to compute, from the definitions, various Stokes parameters of principal interest. Consider, for example, unpolarized light. There is no preference between A_x and A_y; consequently, $\langle\langle A_x^2 + A_y^2 \rangle\rangle$ reduces to $2\langle\langle A_x^2 \rangle\rangle$, and $\langle\langle A_x^2 - A_y^2 \rangle\rangle$ reduces to zero. The other quantities also reduce to zero because δ is a random function of time. If the field is normalized such that $S_0 = 1$, the Stokes vector representation of an unpolarized light wave is (1, 0, 0, 0). Similar reasoning shows that a horizontally polarized beam can be represented by the Stokes vector (1, 1, 0, 0), and a vertically polarized beam can be represented by (1, −1, 0, 0). Right-handed circularly polarized light ($\delta = -\pi/2$) is represented by (1, 0, 0, −1), and the left-handed circularly polarized light ($\delta = \pi/2$) is represented by (1, 0, 0, 1). From the definition, none of the parameters can be greater than the first S_0, which is normalized to 1. Therefore, each of the others lies in the range from −1 to 1. If the beam is entirely unpolarized, $S_1 = S_2 = S_3 = 0$. If it is completely polarized, $S_1^2 + S_2^2 + S_3^2 = 1$. The degree of polarization is therefore defined as

$$\gamma = \frac{(S_1^2 + S_2^2 + S_3^2)^{1/2}}{S_0} \tag{1.6-35}$$

According to Equation (1.6-35), this parameter γ is a real number between 0 and 1. It is thus very useful in describing the partially polarized light. The polarization preference of a partially polarized light can be seen directly from the sign of the parameters S_1, S_2, and S_3.

The parameter S_1 describes the linear polarization along the x or y axis; the probability that the light is linearly polarized along the x axis is $\frac{1}{2}(1 + S_1)$ and along the y axis, $\frac{1}{2}(1 - S_1)$. Thus the values $S_1 = 1, -1$ correspond to complete polarization in these directions. The parameter S_2 describes the linear polarization along directions at angles $\phi = \pm 45°$ to the x axis. The probability that the light is linearly polarized along these directions is, respectively, $\frac{1}{2}(1 + S_2)$ and $\frac{1}{2}(1 - S_2)$. Thus the values $S_2 = 1, -1$ correspond to complete polarization in these directions. Finally, the parameter S_3 represents the degree of circular polarization; the probability that the light wave has right-handed circular polarization is $\frac{1}{2}(1 - S_3)$, and left-handed circular polarization, $\frac{1}{2}(1 + S_3)$.

The Stokes parameters for a polarized light with a complex representation $\chi = e^{i\delta} \tan \psi$ are given by (according to Equation (1.6-17))

$$\begin{aligned} S_0 &= 1 \\ S_1 &= \cos 2\psi \\ S_2 &= \sin 2\psi \cos \delta \\ S_3 &= \sin 2\psi \sin \delta \end{aligned} \tag{1.6-36}$$

According to our convention, a positive S_3 corresponds to left-hand elliptical polarization ($\sin \delta > 0$, clockwise revolution).

Poincaré Sphere

Although the Stokes parameters are introduced for describing partially polarized light, they are also convenient parameters to describe the polarization states of polarized light. For polarized light, the Stokes parameters S_1, S_2, and S_3 can also be employed to represent the polarization states. Since $S_0 = 1$, all points with coordinate (S_1, S_2, S_3) are confined on the surface of a unit sphere in three-dimensional (3-D) space. This sphere is known as the Poincaré

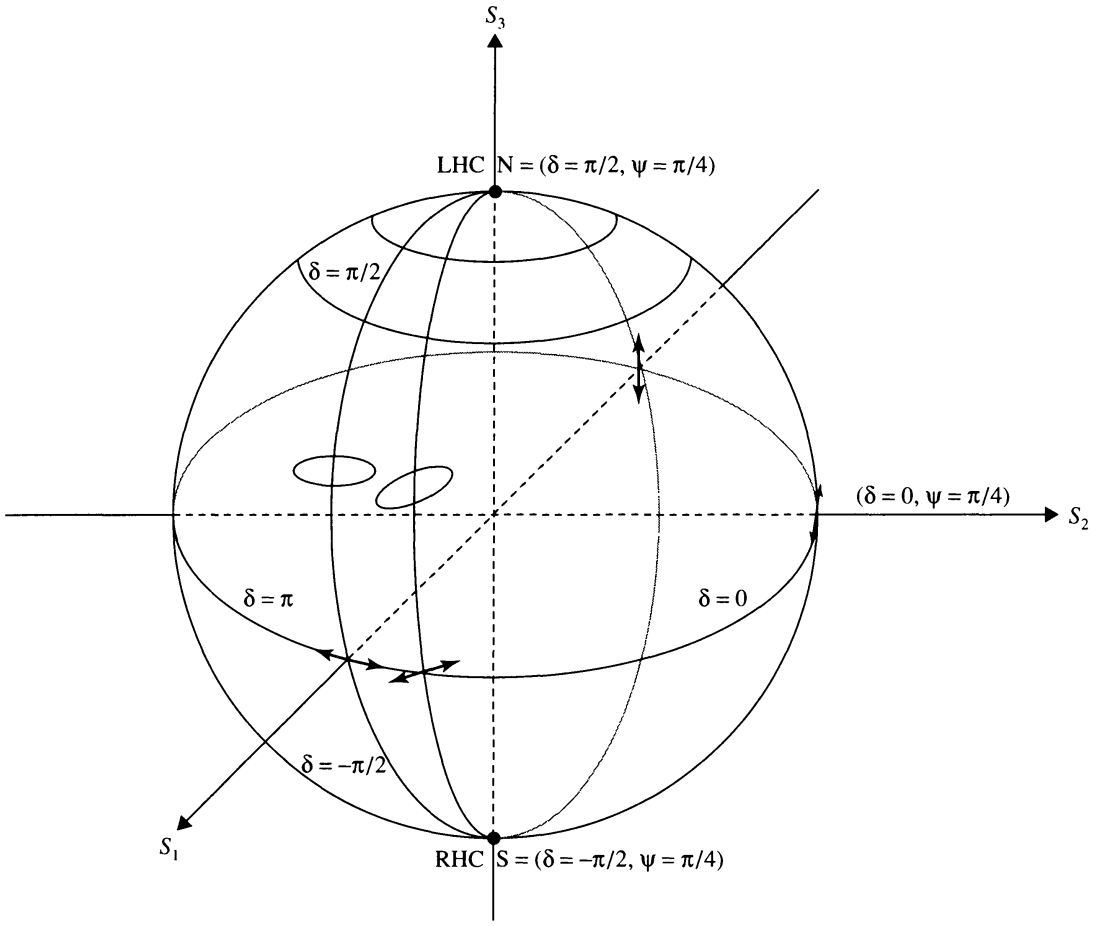


Figure 1.10 Poincaré sphere (showing polarization states at various points).

sphere (see Figure 1.10). Each point on the surface of the sphere represents a unique polarization state. For example, the north pole (0, 0, 1) corresponds to a left-handed circularly (LHC) polarized state, whereas the south pole (0, 0, −1) corresponds to a right-handed circularly (RHC) polarized state. Point (1, 0, 0) corresponds to a linear polarization state parallel to the horizontal direction, whereas point (−1, 0, 0) corresponds to a linear polarization state parallel to the vertical direction. In fact, all points on the equator correspond to a unique linear polarization state. The rest of the points correspond to elliptical polarization states. It is interesting to note that any pair of antipodal points on the Poincaré sphere (two points on the opposite side of the center of the sphere) correspond to states with orthogonal polarization.

According to Equations (1.6-18) and (1.6-19), we have

$$\tan 2\phi = S_2/S_1 \quad \text{and} \quad \sin 2\theta = -S_3 \tag{1.6-37}$$

where ϕ is the inclination angle of the polarization ellipse and θ is the ellipticity angle defined as $\tan^{-1}e$. Generally, S_2/S_1 = constant represents a vertical plane containing the poles. Since both S_1 and S_2 are confined on the surface of the sphere for polarized light, S_2/S_1 = constant actually represents a meridian—a half-circle connecting the north and south poles. According to Equation (1.6-37), ϕ is a constant on the meridian. Thus each meridian represents a class of elliptical polarization states with the same inclination angle ϕ, but with different ellipticities. In addition, S_3 = constant represents a circle on the sphere parallel to the equatorial plane. According to Equation (1.6-37), θ is a constant on this circle (parallel or latitude). Thus each

parallel (latitude) represents a class of elliptical polarization states with the same ellipticity $e = \tan\theta$, but with different inclination angles. The Poincaré sphere is particularly useful in optical birefringent networks where wave plates are employed to change the polarization state of light. It is also particularly useful in the description of the evolution of polarization states in optical fibers when polarization mode dispersion (PMD) is present.

Consider two different points on the Poincaré sphere. Each point represents a polarization state. Let the Stokes vectors be written

$$\begin{aligned} \mathbf{S}_a &= (1, S_{a1}, S_{a2}, S_{a3}) \\ \mathbf{S}_b &= (1, S_{b1}, S_{b2}, S_{b3}) \end{aligned} \tag{1.6-38}$$

Using Equation (1.6-33), it can easily be shown that

$$\mathbf{S}_a \cdot \mathbf{S}_b = 2|\mathbf{J}_a^* \cdot \mathbf{J}_b|^2 \tag{1.6-39}$$

where $\mathbf{J}_a$ and $\mathbf{J}_b$ are the corresponding Jones vectors. For polarized light, the first component of the Stokes vectors is 1, so it is convenient to define three-component unit vectors, consisting of the three components S_1, S_2, and S_3 of the Stokes vectors as

$$\begin{aligned} \mathbf{s}_a &= (S_{a1}, S_{a2}, S_{a3}) \\ \mathbf{s}_b &= (S_{b1}, S_{b2}, S_{b3}) \end{aligned} \tag{1.6-40}$$

These three-component unit vectors are real vectors in Poincaré space. The tips of these unit vectors correspond to points on the Poincaré sphere. With this definition, Equation (1.6-39) can be written

$$|\mathbf{J}_a^* \cdot \mathbf{J}_b|^2 = \tfrac{1}{2}\mathbf{S}_a \cdot \mathbf{S}_b = \tfrac{1}{2}(1 + \mathbf{s}_a \cdot \mathbf{s}_b) \tag{1.6-41}$$

For a pair of antipodal points, $\mathbf{s}_a \cdot \mathbf{s}_b = -1$, the two polarization states form an orthogonal pair.

Stokes vectors and the Poincaré sphere will be employed later in this book to study polarization mode dispersion in optical fibers. It will be shown in Section 7.5 that the polarization transformation by a wave plate can easily be represented by a simple rotation on the Poincaré sphere. Let P be the input polarization state, P' be the output polarization state, and R be the polarization state of the slow mode of the wave plate. The output polarization state P' is obtained by rotating P through an angle Γ around the axis OR, where O is the origin of the sphere and Γ is the phase retardation of the wave plate (see Figure 7.9).

1.7 ELECTROMAGNETIC PROPAGATION IN ANISOTROPIC MEDIA (CRYSTALS)

In anisotropic media such as lithium niobate, quartz, nematic liquid crystals, and calcite, the propagation of electromagnetic radiation is determined by the dielectric tensor ε_{ij} that links the displacement vector and the electric field vector,

$$D_i = \varepsilon_{ij} E_j \tag{1.7-1}$$

where the convention of summation over repeated indices is observed. In nonmagnetic and transparent materials, this tensor is real and symmetric (see Problem 1.37):

$$\varepsilon_{ij} = \varepsilon_{ji} \tag{1.7-2}$$

The magnitude of these nine tensor elements depends, of course, on the choice of the x, y, and z axes relative to the crystal structure. Because of its real and symmetric nature, it is

always possible to find three mutually orthogonal axes in such a way that the off-diagonal elements vanish, leaving

$$\varepsilon = \varepsilon_0 \begin{bmatrix} n_x^2 & 0 & 0 \\ 0 & n_y^2 & 0 \\ 0 & 0 & n_z^2 \end{bmatrix} = \begin{bmatrix} \varepsilon_x & 0 & 0 \\ 0 & \varepsilon_y & 0 \\ 0 & 0 & \varepsilon_z \end{bmatrix} \tag{1.7-3}$$

where ε_x, ε_y, and ε_z are the principal dielectric constants and n_x, n_y, and n_z are the principal indices of refraction. These directions (x, y, and z) are called the principal dielectric axes of the crystal. According to Equations (1.7-1) and (1.7-3), a plane wave propagating along the z axis can have two phase velocities, depending on its state of polarization. Specifically, the phase velocity is c/n_x for x-polarized light and c/n_y for y-polarized light. Generally, there are two normal modes of polarization for each direction of propagation.

It is important to note that the dielectric tensor (or dielectric constants) is a function of the frequency (or wavelength) of the electromagnetic field. This is known as dispersion. In the regime of optical waves, the frequencies are in the range of 10^{14}/s. We often use the refractive indices to describe the propagation of waves in optical media.

Plane Waves in Homogeneous Media and Normal Surface

To study such propagation along a general direction, we assume a monochromatic plane wave with an electric field vector

$$\mathbf{E} \exp[i(\omega t - \mathbf{k} \cdot \mathbf{r})] \tag{1.7-4}$$

and a magnetic field vector

$$\mathbf{H} \exp[i(\omega t - \mathbf{k} \cdot \mathbf{r})] \tag{1.7-5}$$

where $\mathbf{k}$ is the wavevector $\mathbf{k} = (\omega/c)n\mathbf{s}$, with $\mathbf{s}$ as a unit vector in the direction of propagation. The phase velocity c/n, or equivalently the refractive index n, is to be determined. Substituting $\mathbf{E}$ and $\mathbf{H}$ from Equations (1.7-4) and (1.7-5), respectively, into Maxwell's equations (1.1-1), (1.1-2) and (1.7-1) gives

$$\mathbf{k} \times \mathbf{E} = \omega\mu\mathbf{H} \tag{1.7-6}$$

$$\mathbf{k} \times \mathbf{H} = -\omega\varepsilon\mathbf{E} = -\omega\mathbf{D} \tag{1.7-7}$$

By eliminating $\mathbf{H}$ from Equations (1.7-6) and (1.7-7), we obtain

$$\mathbf{k} \times (\mathbf{k} \times \mathbf{E}) + \omega^2\mu\varepsilon\mathbf{E} = 0 \tag{1.7-8}$$

This equation will now be used to solve for the eigenvectors $\mathbf{E}$ and the corresponding eigenvalues n.

In the principal coordinate system, the dielectric tensor ε is given by Equation (1.7-3). Equation (1.7-8) can be written

$$\begin{bmatrix} \omega^2\mu\varepsilon_x - k_y^2 - k_z^2 & k_x k_y & k_x k_z \\ k_y k_x & \omega^2\mu\varepsilon_y - k_x^2 - k_z^2 & k_y k_z \\ k_z k_x & k_z k_y & \omega^2\mu\varepsilon_z - k_x^2 - k_y^2 \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = 0 \tag{1.7-9}$$

where we recall that $\varepsilon_x = \varepsilon_0 n_x^2$, $\varepsilon_y = \varepsilon_0 n_y^2$, and $\varepsilon_z = \varepsilon_0 n_z^2$.

For nontrivial solutions to exist, the determinant of the matrix in Equation (1.7-9) must vanish. This leads to a relation between ω and $\mathbf{k}$:

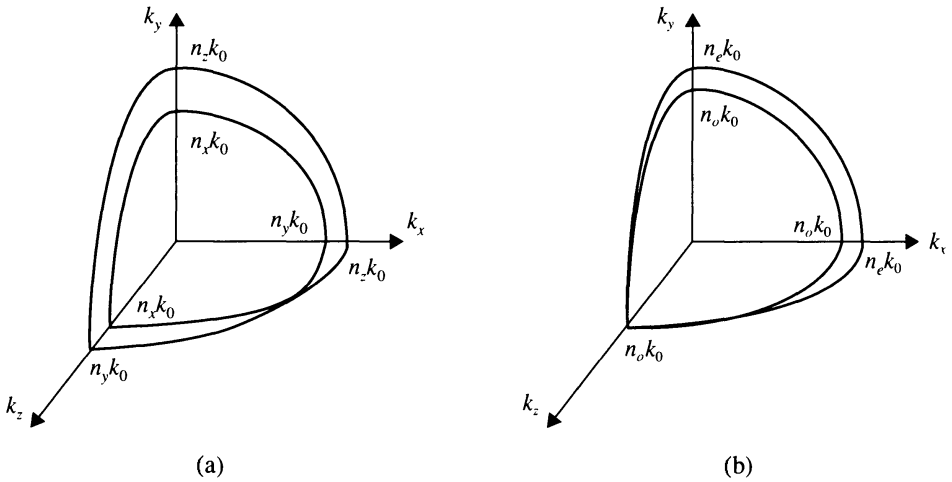


Figure 1.11 (a) One octant of a normal surface in momentum space with $n_x < n_y < n_z$. k_x, k_y, and k_z are in units of $k_0 = \omega/c$. (b) One octant of a normal surface in momentum space with $n_o = n_x = n_y < n_z = n_e$, $k_0 = \omega/c$.

$$\det \begin{vmatrix} \omega^2\mu\varepsilon_x - k_y^2 - k_z^2 & k_x k_y & k_x k_z \\ k_y k_x & \omega^2\mu\varepsilon_y - k_x^2 - k_z^2 & k_y k_z \\ k_z k_x & k_z k_y & \omega^2\mu\varepsilon_z - k_x^2 - k_y^2 \end{vmatrix} = 0 \tag{1.7-10}$$

At a given frequency ω, this equation represents a three-dimensional surface in **k** space (momentum space). This surface, known as the *normal surface*, consists of two shells. These two shells, in general, have four points in common. The two lines that go through the origin and these points are known as the optic axes. Figure 1.11 shows one octant of a general normal surface. Given a direction of propagation, there are in general two k values that are the intersections of the direction of propagation **s** and the normal surface. These two k values correspond to two different phase velocities (ω/k) of the waves propagating along the chosen direction. The directions of the electric field vector associated with these propagations can also be obtained from Equation (1.7-9) and are given by

$$\begin{bmatrix} \dfrac{k_x}{k^2 - \omega^2\mu\varepsilon_x} \\ \dfrac{k_y}{k^2 - \omega^2\mu\varepsilon_y} \\ \dfrac{k_z}{k^2 - \omega^2\mu\varepsilon_z} \end{bmatrix} \tag{1.7-11}$$

provided that the denominators do not vanish. In the case when $n_o = n_x = n_y < n_z = n_e$, the normal surface consists of a sphere and an ellipsoid of revolution with the z axis being the axis of symmetry.

For propagation in the direction of the optic axes, there is only one value of k and consequently only one phase velocity. There are, however, two independent directions of polarization. Equations (1.7-10) and (1.7-11) are often written in terms of the direction cosines of the wavevector. By using the relation $\mathbf{k} = (\omega/c)n\mathbf{s}$ for the plane wave given by Equation (1.7-4), Equations (1.7-10) and (1.7-11) can be written as

$$\frac{s_x^2}{n^2 - n_x^2} + \frac{s_y^2}{n^2 - n_y^2} + \frac{s_z^2}{n^2 - n_z^2} = \frac{1}{n^2} \tag{1.7-12}$$

and

$$\begin{bmatrix} \dfrac{s_x}{n^2 - n_x^2} \\ \dfrac{s_y}{n^2 - n_y^2} \\ \dfrac{s_z}{n^2 - n_z^2} \end{bmatrix} \tag{1.7-13}$$

respectively, where we have used $\varepsilon_x = \varepsilon_0 n_x^2$, $\varepsilon_y = \varepsilon_0 n_y^2$, and $\varepsilon_z = \varepsilon_0 n_z^2$.

Equation (1.7-12) is known as *Fresnel's equation of wave normals* and can be solved for the eigenvalues of index of refraction, and Equation (1.7-13) gives the directions of polarization. Equation (1.7-12) is a quadratic equation in n^2. Therefore, for each direction of propagation (a set of s_x, s_y, s_z), two solutions for n^2 can be obtained from Equation (1.7-12). To complete the solution of the problem, we use the values of n^2, one at a time, in Equation (1.7-13). This gives us the polarizations (electric field vectors) of these waves. It can be seen that in a nonabsorbing medium these normal modes are linearly polarized since all the components are real in Equation (1.7-13). Let $\mathbf{E}_1$ and $\mathbf{E}_2$ be the electric field vectors and $\mathbf{D}_1$ and $\mathbf{D}_2$ be the displacement vectors of the linearly polarized normal modes associated with n_1^2 and n_2^2, respectively. Maxwell's equation $\nabla \cdot \mathbf{D} = 0$ requires that $\mathbf{D}_1$ and $\mathbf{D}_2$ are orthogonal to $\mathbf{s}$. Since $\mathbf{D}_1 \cdot \mathbf{D}_2 = 0$ (the proof of this orthogonal relation is left as a problem for students), the three vectors $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{s}$ form an orthogonal triad. According to Equations (1.7-6) and (1.7-7), $\mathbf{D}$ and $\mathbf{H}$ are both perpendicular to the direction of propagation $\mathbf{s}$. Consequently, the direction of energy flow as given by the Poynting vector $\mathbf{E} \times \mathbf{H}$ is, in general, not collinear with the direction of propagation $\mathbf{s}$. Since $\mathbf{D}$, $\mathbf{E}$, and $\mathbf{k}$ are all orthogonal to $\mathbf{H}$, they must lie in the same plane.

Orthogonality of Normal Modes (Eigenmodes)

It can be shown that $\mathbf{D}$, $\mathbf{E}$, and $\mathbf{s}$ all lie in the same plane. In addition, these field vectors satisfy the following relations (see Problem 1.27):

$$\begin{aligned} \mathbf{D}_1 \cdot \mathbf{D}_2 &= 0 \\ \mathbf{D}_1 \cdot \mathbf{E}_2 &= 0 \\ \mathbf{D}_2 \cdot \mathbf{E}_1 &= 0 \\ \mathbf{s} \cdot \mathbf{D}_1 = \mathbf{s} \cdot \mathbf{D}_2 &= 0 \end{aligned} \tag{1.7-14}$$

The electric field vectors $\mathbf{E}_1$ and $\mathbf{E}_2$ are, in general, not orthogonal. The general orthogonality relation of the eigenmodes of propagation is often written

$$\mathbf{s} \cdot (\mathbf{E}_1 \times \mathbf{H}_2) = 0 \tag{1.7-15}$$

This latter relation shows that the power flow in a lossless anisotropic medium along the direction of propagation is the sum of the power carried by each mode individually.

To summarize: Along an arbitrary direction of propagation $\mathbf{s}$, there can exist two independent plane wave, linearly polarized propagation modes. These modes have phase velocities $\pm(c/n_1)$ and $\pm(c/n_2)$, where n_1^2 and n_2^2 are the two solutions of Fresnel's equation (1.7-12). The electric field vectors of these two normal modes are given by Equation (1.7-11) or (1.7-13).

Although Equation (1.7-11) provides a general explicit expression for the polarization states (**E** vector) of the normal modes in a general anisotropic medium, one must be careful in applying the equation for propagation along the principal axes or principal planes when the denominators in Equation (1.7-11) can become zero. For propagation along these special directions, it is often easier to obtain the normal modes directly from the wave equation (1.7-9).

Classification of Media

We have shown above that the normal surface contains a good deal of information about the wave propagation in anisotropic media. The normal surface is uniquely determined by the principal indices of refraction n_x, n_y, n_z. In the general case when the three principal indices n_x, n_y, n_z are all different, there are two optical axes. In this case, the medium is said to be biaxial. In many optical electronic materials (e.g., $LiNbO_3$ crystals, nematic liquid crystals) it happens that two of the principal indices are equal, in which case the equation for the normal surface (Equation (1.7-10) or (1.7-12)) can be factored according to

$$\left(\frac{k_x^2 + k_y^2}{n_e^2} + \frac{k_z^2}{n_o^2} - \frac{\omega^2}{c^2}\right)\left(\frac{k^2}{n_o^2} - \frac{\omega^2}{c^2}\right) = 0 \tag{1.7-16}$$

where $n_o^2 = \varepsilon_x/\varepsilon_0 = \varepsilon_y/\varepsilon_0$ and $n_e^2 = \varepsilon_z/\varepsilon_0$.

The normal surface in this case consists of a sphere and an ellipsoid of revolution (see Figure 1.11b). These two sheets of the normal surface touch at two points on the z axis. The z axis is therefore the only optic axis, and the medium is said to be uniaxial. If all three principal indices are equal, the two sheets of normal surface degenerate to a single sphere, and the medium is optically isotropic.

In a biaxial medium, the principal coordinate axes can be labeled in such a way that the three principal indices are in the following order:

$$n_x < n_y < n_z \tag{1.7-17}$$

In this convention, the optical axes lie in the xz plane. Cross sections of the normal surfaces with the xz plane are shown in Figure 1.12a. In a uniaxial medium, the index of refraction that corresponds to the two equal elements, $n_o^2 = \varepsilon_x/\varepsilon_0 = \varepsilon_y/\varepsilon_0$, is called the ordinary index n_o; the other index, corresponding to ε_z, is called the extraordinary index n_e. If $n_o < n_e$, the medium is said to be positive, whereas if $n_o > n_e$, it is said to be negative. Most liquid crystals with rodlike molecules are positive uniaxial media with $n_o < n_e$. Intersections of the normal surfaces with the xz plane are again shown in Figures 1.12b and 1.12c. The optic axis corresponds to the principal axis, which has a unique index of refraction. Table 1.3 lists some examples of solid crystals with their indices of refraction.

The Index Ellipsoid

The surface of constant energy density U_e in **D** space can be written

$$\frac{D_x^2}{\varepsilon_x} + \frac{D_y^2}{\varepsilon_y} + \frac{D_z^2}{\varepsilon_z} = 2U_e \tag{1.7-18}$$

where ε_x, ε_y, and ε_z are the principal dielectric constants. If we replace $\mathbf{D}/\sqrt{2U_e}$ by **r** and use the principal refractive indices $n_i^2 = \varepsilon_i/\varepsilon_0$ ($i = x, y, z$), the last equation can be written

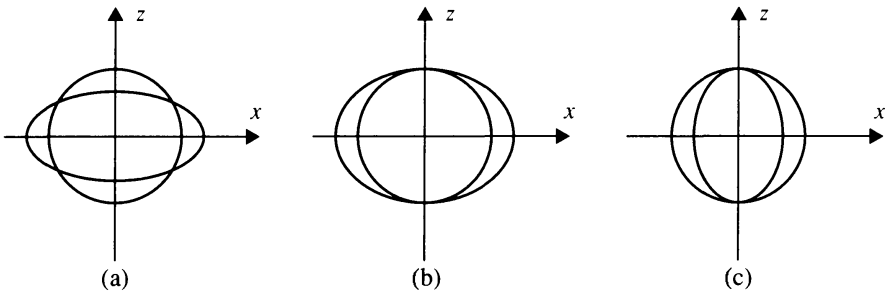


Figure 1.12 Intersection of the normal surface with the xz plane for (a) biaxial media (with $n_x < n_y < n_z$), (b) uniaxial media with positive birefringence ($n_o < n_e$), and (c) uniaxial media with negative birefringence ($n_e < n_o$).

TABLE 1.3 Refractive Indices[a] of Some Typical Solid Crystals

Isotropic	Fluorite		1.392	
	Sodium chloride, NaCl		1.544	
	Diamond, C		2.417	
	CdTe		2.69	
	GaAs		3.40	
	Ge		3.40	
	InP		3.61	
	GaP		3.73	
Uniaxial		n_o		n_e
Positive	MgF_2	1.378		1.390
	Quartz, SiO_2	1.544		1.553
	Beryllium oxide, BeO	1.717		1.732
	$La_3Ga_5SiO_{14}$	1.90		1.91
	ZnO	1.94		1.96
	SnO_2	2.01		2.10
	YVO_4	1.96		2.16
	$LiTaO_3$	2.183		2.188
	ZnS	2.354		2.358
	Rutile, TiO_2	2.616		2.903
Negative	KDP, KH_2PO_4	1.507		1.467
	ADP, $(NH_4)\,H_2PO_4$	1.522		1.478
	Beryl, $Be_3Al_2(SiO_3)_6$	1.598		1.590
	Sodium nitrate, $NaNO_3$	1.587		1.366
	Calcite, $CaCO_3$	1.658		1.486
	β-BaB_2O_4 (BBO)	1.67		1.55
	Sapphire, Al_2O_3	1.768		1.760
	Lithium niobate, $LiNbO_3$	2.300		2.208
	$PbMoO_3$	2.40		2.27
	Proustite, Ag_3AsS_3	3.019		2.739
Biaxial		n_x	n_y	n_z
	Gypsum	1.520	1.523	1.530
	Feldspar	1.522	1.526	1.530
	Mica	1.552	1.582	1.588
	Topaz	1.619	1.620	1.627
	Sodium nitrite, $NaNO_2$	1.344	1.411	1.651
	$YAlO_3$	1.923	1.938	1.947
	SbSI	2.7	1.7	3.8

[a] The refractive indices of most materials depend on the wavelength (dispersion). The listed numbers are typical values.

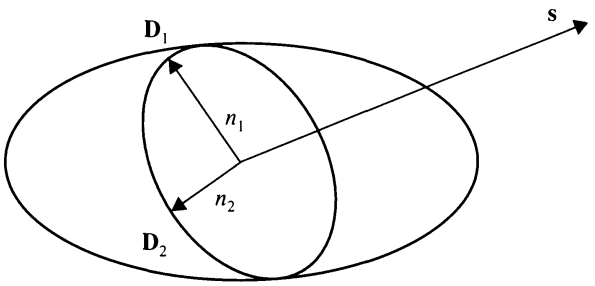


Figure 1.13 Method of index of ellipsoid. The inner ellipse is the intersection of the index ellipsoid with the plane that is normal to s and passes through the center of the ellipsoid.

$$\frac{x^2}{n_x^2} + \frac{y^2}{n_y^2} + \frac{z^2}{n_z^2} = 1 \tag{1.7-19}$$

This is the equation of a general ellipsoid with major axes parallel to the principal axes of the crystal, whose respective lengths are $2n_x$, $2n_y$, and $2n_z$. The ellipsoid is known as the *index ellipsoid* or, sometimes, as the *optical indicatrix*. The index ellipsoid is used mainly to find the **D** vectors and the two corresponding indices of refraction of the normal modes of propagation along a given direction of propagation **s** in a crystal. This is done by means of the following prescription: Find the intersection ellipse between the index ellipsoid and a plane through the origin that is normal to the direction of propagation **s**. The two axes of the intersection ellipse are equal in length to $2n_1$ and $2n_2$, where n_1 and n_2 are the two indices of refraction associated with the normal modes of propagation, that is, the solutions of Equation (1.7-12). These two axes are parallel, respectively, to the **D** vectors of the normal modes of propagation (see Figure 1.13). A formal proof of such a procedure is given in References [2, 3].

1.8 PLANE WAVES IN UNIAXIALLY ANISOTROPIC MEDIA—PHASE RETARDATION

Many of the optical electronic materials (e.g., $LiNbO_3$ crystal) are optically uniaxial. In addition, most liquid crystal displays involve the use of nematic liquid crystals. A planar nematic slab is a good example of a homogeneous uniaxial liquid crystal. Thus the propagation of optical waves in uniaxially anisotropic media deserves special attention. For the purpose of discussion, we rewrite the normal surface of a uniaxial medium as

$$\left(\frac{k_x^2 + k_y^2}{n_e^2} + \frac{k_z^2}{n_o^2} - \frac{\omega^2}{c^2}\right)\left(\frac{k^2}{n_o^2} - \frac{\omega^2}{c^2}\right) = 0 \tag{1.8-1}$$

We note that the normal surface consists of two parts. The sphere gives the relation between ω and **k** of the ordinary (O) wave. The ellipsoid of revolution gives the similar relation for the extraordinary (E) wave. These two surfaces touch at two points on the z axis. The eigen refractive indices associated with these two modes of propagation are given by

$$O \text{ wave:} \quad n = n_o \tag{1.8-2}$$

$$E \text{ wave:} \quad \frac{1}{n^2} = \frac{\cos^2\theta}{n_o^2} + \frac{\sin^2\theta}{n_e^2} \tag{1.8-3}$$

where θ is the angle between the direction of propagation and the c axis (the crystal z axis). For propagation along the optic axis (c axis), the eigen refractive indices of both modes are n_o, according to Equations (1.8-2) and (1.8-3).

The electric field vector of the O wave cannot be obtained directly from Equation (1.7-13) because of the vanishing denominators. It can easily be obtained from Equation (1.7-9). By using $\varepsilon_x = \varepsilon_y = \varepsilon_0 n_o^2$, $\varepsilon_z = \varepsilon_0 n_e^2$, and $\mathbf{k}_o = (\omega/c) n_o \mathbf{s}$, Equation (1.7-9) can be written

$$\begin{bmatrix} s_x^2 & s_x s_y & s_x s_z \\ s_y s_x & s_y^2 & s_y s_z \\ s_z s_x & s_z s_y & (n_e/n_o)^2 - (s_x^2 + s_y^2) \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = 0 \tag{1.8-4}$$

A simple inspection of this equation yields the following direction of polarization:

$$O \text{ wave:} \quad \mathbf{E} = \begin{bmatrix} s_y \\ -s_x \\ 0 \end{bmatrix} \tag{1.8-5}$$

where we recall that s_x, s_y, s_z are directional cosines of the direction of propagation. The electric field vector of the E wave can be obtained from Equation (1.7-13) and is given by

$$E \text{ wave:} \quad \mathbf{E} = \begin{bmatrix} \dfrac{s_x}{n^2 - n_o^2} \\ \dfrac{s_y}{n^2 - n_o^2} \\ \dfrac{s_z}{n^2 - n_e^2} \end{bmatrix} \tag{1.8-6}$$

where n is given by Equation (1.8-3). The corresponding wavevector is $\mathbf{k}_e = (\omega/c) n \mathbf{s}$.

Notice that the electric field vector of the O wave is perpendicular to the plane formed by the wavevector $\mathbf{k}_o$ and the c axis, whereas the electric field vector of the E wave is not exactly perpendicular to the wavevector $\mathbf{k}_e$. However, the deviation from 90° is very small. This small angle between the field vectors $\mathbf{E}$ and $\mathbf{D}$ is also the angle between the phase velocity and the group velocity (see Problem 1.30). Therefore, for practical purposes, we may assume that the electric field is transverse to the direction of propagation. The displacement vectors $\mathbf{D}$ of the normal modes are exactly perpendicular to the wavevectors $\mathbf{k}_o$ and $\mathbf{k}_e$, respectively, and can be written

$$O \text{ wave:} \quad \mathbf{D}_o = \frac{\mathbf{k}_o \times \mathbf{c}}{|\mathbf{k}_o \times \mathbf{c}|} \tag{1.8-7}$$

$$E \text{ wave:} \quad \mathbf{D}_e = \frac{\mathbf{D}_o \times \mathbf{k}_e}{|\mathbf{D}_o \times \mathbf{k}_e|} \tag{1.8-8}$$

where $\mathbf{c}$ is a unit vector parallel to the c axis of the crystal.

Let (θ, ϕ) be the angle of propagation in spherical coordinates. The unit vector $\mathbf{s}$ can be written

$$\mathbf{s} = \begin{bmatrix} \sin\theta \cos\phi \\ \sin\theta \sin\phi \\ \cos\theta \end{bmatrix} \tag{1.8-9}$$

Using Equation (1.8-9), the normal modes for **E** can be written

$$O \text{ wave:} \quad \mathbf{E}_o = \begin{bmatrix} \sin\phi \\ -\cos\phi \\ 0 \end{bmatrix} \tag{1.8-10}$$

$$E \text{ wave:} \quad \mathbf{E}_e = \begin{bmatrix} n_e^2 \cos\theta \cos\phi \\ n_e^2 \cos\theta \sin\phi \\ -n_o^2 \sin\theta \end{bmatrix} \tag{1.8-11}$$

We note that these two modes are mutually orthogonal. The normal modes for **D** can be written

$$O \text{ wave:} \quad \mathbf{D}_o = \begin{bmatrix} \sin\phi \\ -\cos\phi \\ 0 \end{bmatrix} \tag{1.8-12}$$

$$E \text{ wave:} \quad \mathbf{D}_e = \begin{bmatrix} \cos\theta \cos\phi \\ \cos\theta \sin\phi \\ -\sin\theta \end{bmatrix} \tag{1.8-13}$$

These two **D** vectors and **s** are also mutually orthogonal. The above results can also be obtained by using the index ellipsoid.

If inside the uniaxial medium a polarized light is generated that is to propagate along a direction **s**, the displacement vector of this light can always be written as a linear combination of these two normal modes; that is,

$$\mathbf{D} = C_o\mathbf{D}_o \exp(-i\mathbf{k}_o \cdot \mathbf{r}) + C_e\mathbf{D}_e \exp(-i\mathbf{k}_e \cdot \mathbf{r}) \tag{1.8-14}$$

where C_o and C_e are constants and $\mathbf{k}_o$ and $\mathbf{k}_e$ are the wavevectors that are, in general, different (Equation (1.8-2)). As the light propagates inside the medium, a phase retardation between these two components is built up due to the difference in their phase velocities. Such a phase retardation between the two components leads to a new polarization state. Thus birefringent plates can be used to alter the polarization state of light. For a parallel plate with thickness d, the phase retardation can be written

$$\Gamma = (k_{ez} - k_{oz})d \tag{1.8-15}$$

where k_{ez} and k_{oz} are the z components of the wavevectors and the z axis is perpendicular to the surface of the plates.

Optical Rotatory Power and Faraday Rotation

In addition to linear birefringence in crystals, optical electronic materials (e.g., quartz crystal) exist in which the normal modes of propagation are circularly polarized. These media are said to be circularly birefringent or optically active (OA). When a beam of linearly polarized light traverses through these materials, the plane of polarization is rotated by an angle that is proportional to the distance of propagation. It was first observed in 1811 that quartz crystal rotates the plane of polarization of a beam of linearly polarized light traversing through it in the direction of the optical axis. The plane of polarization rotates 21.7 degrees per mm for sodium light ($\lambda = 589$ nm) at 20 °C. Both right-handed and left-handed rotations exist in different specimens. The optical activity of liquids was first observed in sugar solutions in 1815.

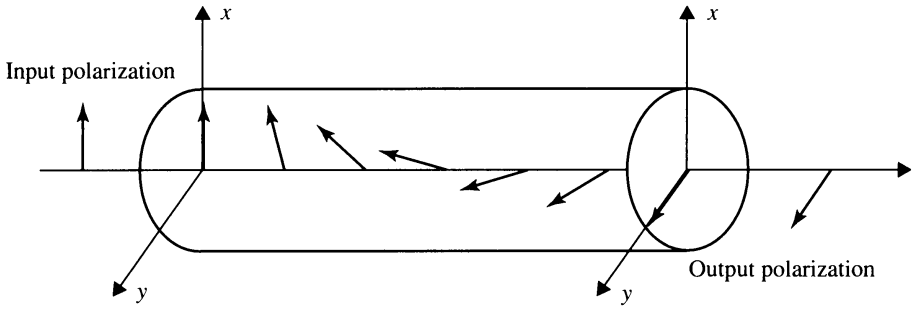


Figure 1.14 Rotation of the plane of polarization by an optically active medium.

It is known that fused quartz (silica) is optically inactive. The optical rotatory power in solids is due to the crystalline structure and the arrangement of molecules in the crystal. In liquids, the molecular orientations are random. Thus the cause of optical rotatory power in liquids must lie in the structure of the molecules themselves. Virtually all molecules in optically active liquids possess an asymmetric atom of carbon, nitrogen, or sulfur. In addition, all these liquids have a twin substance (isomer) with an opposite rotatory power. The existence of mirror-image isomers is a result of the presence of one or more asymmetric carbon atoms in the compound. Thus, these molecular structures can have left- and right-hand (chiral) forms. These forms are conventionally designated as dextro (D) and levo (L) because they compare to each other structurally as do the right and left hands when the molecule is compared with its mirror-image isomer. Optical rotatory power in liquids is due to the asymmetric structural arrangement (e.g., helical structure) of atoms in the randomly oriented molecules.

Figure 1.14 illustrates the rotation of the plane of polarization by an optically active medium. The amount of rotation is proportional to the path length of light in the medium. Conventionally, the rotatory power of a medium is given in degrees per mm; that is, the specific rotatory power is defined as the amount of rotation per unit length.

The sense of rotation in optically active materials bears a fixed relation to the propagation wavevector (**k**) of the beam of light, so if the beam of light is made to traverse the medium once in each of two opposite directions, as in the case of reflection from the right end face in Figure 1.14, the net rotation is zero. A substance is called dextrorotatory (or right-handed) if the sense of rotation of the plane of polarization is counterclockwise as viewed by an observer facing the approaching light beam. If the sense of rotation is clockwise, the substance is called levorotatory (or left-handed). Quartz occurs in both right-handed and left-handed crystalline forms. Many other substances are now known to exhibit optical activity; these include cinnabar, sodium chlorate, turpentine, sugar, strychnine sulfate, tellurium, selenium, and silver thiogallate ($AgGaS_2$). Many liquid crystal materials and organic compounds also exhibit optical rotatory power. The specific rotatory powers of some optically active media are given in Table 1.4.

Fresnel first recognized in 1825 that the optical activity arises from circular double refraction, in which the eigenwaves of propagation (i.e., the independent plane wave solutions of Maxwell's equations) are right and left circularly polarized waves. Let n_r and n_l be the refractive indices associated with these two waves, and assume that the waves are propagating in the $+z$ direction. It can be shown that the specific rotatory power is given by [2, 3]

$$\rho = \frac{\pi}{\lambda}(n_l - n_r) \tag{1.8-16}$$

TABLE 1.4 Optical Rotatory Powers of Some Solids

	λ (Å)	ρ (degree/mm)
Quartz	4000	49
	4550	37
	5000	31
	5500	26
	6000	22
	6500	17
$AgGaS_2$	4850	950
	4900	700
	4950	600
	5000	500
	5050	430
Se	7500	180
	10,000	30
Te	(6 μm)	40
	(10 μm)	15
TeO_2	3698	587
	4382	271
	5300	143
	6328	87
	10,000	30

The optical rotation is right-handed (counterclockwise) if $n_r < n_l$. Thus the plane of polarization turns in the same sense as the circularly polarized wave, which travels with the greater phase velocity. It can be shown that the polarization ellipse of a beam of elliptically polarized light will rotate the same angle $\omega z(n_l - n_r)/2c$ with its shape remaining unchanged (see Problem 1.18).

Faraday Rotation

In 1845, Faraday discovered that isotropic substances rotate the plane of polarization of linearly polarized light when placed in a strong magnetic field. When the field is reversed, the rotation is reversed. This is known as the Faraday effect. The rotation is proportional to the length of the optical path, and to the component of the magnetic field along the direction of propagation of the light. In an optically active medium, the direction of rotation bears a fixed relation to the direction of propagation, so that if a beam of light is reflected back on itself, the net rotation is zero. Such a medium is said to be reciprocal. In the Faraday effect, however, the rotation bears a fixed relation to the magnetic field **B**, so that reflection back on itself doubles the rotation. Such a phenomenon is known as nonreciprocal.

The specific rotation (i.e., rotation per unit length) of a Faraday cell is often written

$$\rho = VB \tag{1.8-17}$$

where B is the component of magnetic field along the direction of propagation and V is a constant known as the Verdet constant.

The Faraday effect originates from the effect of the static magnetic field on the motion of electrons via the Lorentz force. Faraday rotation has been observed in many solids, liquids, and even gases. A few values of the Verdet constant are given in Table 1.5.

TABLE 1.5 Value of the Verdet Constant at $\lambda = 5893$ Å

Substance	T (°C)	V (deg/G-mm)[a]
Water	20	2.18×10^{-5}
Fluorite		1.5×10^{-6}
Diamond		2.0×10^{-5}
Glass (crown)	18	2.68×10^{-5}
Glass (flint)		5.28×10^{-5}
Carbon disulfide (CS_2)	20	7.05×10^{-5}
Phosphorus	33	2.21×10^{-4}
Sodium chloride		6.0×10^{-5}
MBBA	20	6.67×10^{-5}

[a] 1 gauss = 10^{-4} tesla.

From the atomic point of view, the Faraday effect is related to the Zeeman effect. As a result of interaction between the orbiting electrons and the magnetic field, each electron energy level is split into several sublevels. By virtue of conservation of angular momentum, RHC polarized light and LHC polarized light that propagate along the direction of the magnetic field will interact with different sets of the sublevels. Thus the medium exhibits circular birefringence in the presence of the magnetic field, leading to a rotation of the polarization vector.

Faraday Isolator and Optical Circulator

Faraday effect plays an important role in nonreciprocal optical devices such as isolators and circulators. These nonreciprocal optical devices now form an integral part of most optical communication networks employing semiconductor lasers and amplifiers. Most of these lasers are extremely sensitive to even small amounts of reflected light, which cause instabilities in their output power and frequency characteristics. Optical isolators are made of Faraday cells in conjunction with polarizers that virtually remove all reflected light. Figure 1.15 shows the principle of operation of an optical isolator.

In an optical isolator as shown in Figure 1.15, a beam of polarized light will undergo a 45° rotation of the direction of propagation. Upon reflection, the same beam of light will undergo an additional 45° rotation, resulting in a net rotation of 90°. Using polarizers, it is possible to eliminate reflected light. As we know, reflected light can cause stability problems when the light enters the laser cavity. Optical isolators play an important role in optical networks to ensure stable operation of lasers. Figures 1.15c and 1.15d show an optical arrangement of an isolator for a beam of unpolarized light. A beam of unpolarized light is split into two polarized components by a polarizing beam splitter (PBS). These two polarized components are rotated by 90° by a combination of a 45° Faraday cell (FC) and a 45° optical rotator (OA). After rotation, these two components are recombined at an output PBS. For reflected light as shown in Figure 1.15d, the net rotation through the combination of FC and OA is zero. As a result, the reflected light is directed to a new output port.

1.9 JONES MATRIX METHOD

In Section 1.8, we discussed the propagation of plane waves in homogeneous and anisotropic media. Any plane wave propagation can easily be described in terms of a superposition (or linear combination) of the propagation of two orthogonally polarized components. Many

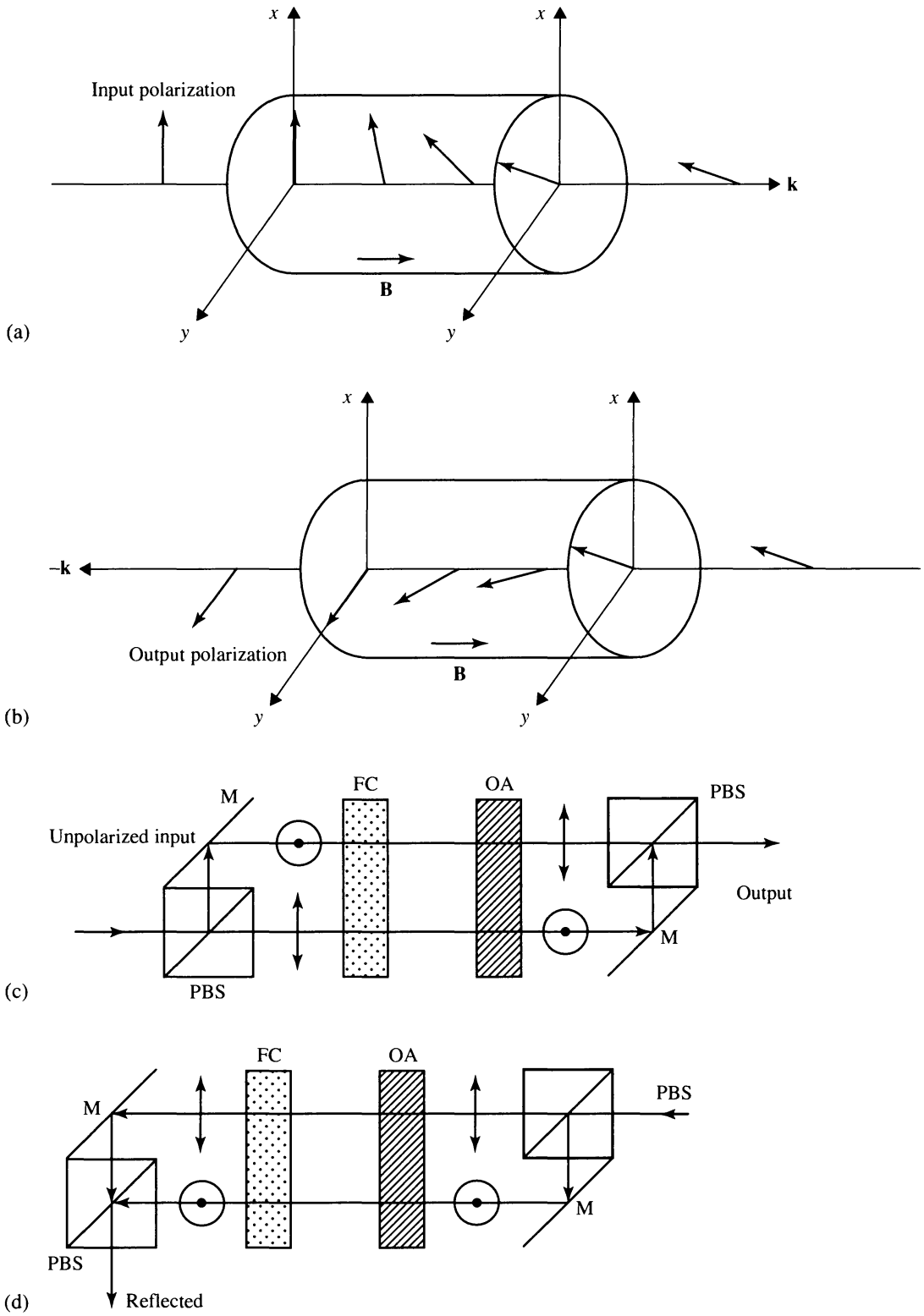


Figure 1.15 (a) A beam of linearly polarized light is rotated 45° by a Faraday rotator. (b) The retroreflected beam is rotated by an additional 45° by the Faraday rotator. (c) Schematic drawing of an optical circulator that also functions as an isolator. (d) Schematic drawing of the same circulator showing the reflected light is directed at a new output port. M = mirror, PBS = polarizing beam splitter, FC = Faraday cell, OA = optical rotator.

optical systems consist of a sequence of birefringent media. In the case of an imperfect optical fiber (e.g., silica fiber with an elliptical core), the birefringence may depend on the position along the fiber. In a situation like this, a systematic method is needed to treat the propagation of light through a series of birefringent elements. In this section, we describe such a powerful technique, known as the Jones matrix method [1].

We have shown in Section 1.8 that light propagation in a birefringent medium consists of a linear superposition of two normal modes. These normal modes have well-defined phase velocities and directions of polarization. The birefringent media may be either uniaxial or biaxial. However, most commonly used materials in optical electronics are uniaxial. In a uniaxial medium, these normal modes are the ordinary and the extraordinary waves. The directions of polarization for these normal modes are mutually orthogonal and are called the "slow" and "fast" axes of the medium for that direction of propagation. In traditional birefringent optics, retardation plates are usually cut in such a way that the c axis lies in the plane of the plate surfaces. In this case the propagation direction of normally incident light is perpendicular to the c axis.

Retardation plates (also called wave plates) and birefringent fibers are polarization-state converters, or transformers. The polarization state of a light beam can be converted to any other polarization state by means of a suitable retardation plate. In formulating the Jones matrix method, we assume that there is no reflection of light from either surface of the plate and the light is totally transmitted through the plate surfaces. In practice, there is reflection, though most retardation plates are coated so as to reduce the surface reflection loss. The Fresnel reflections at the plate surfaces not only decrease the transmitted intensity but also produce a fine structure (ripple) of the spectral transmittance because of multiple-reflection interference (Fabry–Perot effect). Referring to Figure 1.16, we consider an incident beam of light with a polarization state described by the Jones vector

$$\mathbf{V} = \begin{bmatrix} V_x \\ V_y \end{bmatrix} \tag{1.9-1}$$

where V_x and V_y are two complex numbers. The x and y axes are fixed laboratory axes. To determine how the light propagates in the retardation plate, we need to decompose the light into a linear combination of the "fast" and "slow" normal modes of the medium. This is done by the coordinate transformation

$$\begin{bmatrix} V_s \\ V_f \end{bmatrix} = \begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} \equiv R(\psi) \begin{bmatrix} V_x \\ V_y \end{bmatrix} \tag{1.9-2}$$

where $R(\psi)$ is called the coordinate rotation matrix and V_s and V_f are the slow and fast components, respectively, of the polarization vector $\mathbf{V}$. The slow and fast axes are fixed in the medium. These two components are normal modes of the retardation plate and will propagate with their own phase velocities and polarizations. The azimuth ψ is defined as the angle between the sf coordinate and the xy coordinate (see Figure 1.16) with the z axis as the axis of the coordinate rotation. Because of the difference in phase velocity, one component is retarded relative to the other. This retardation changes the polarization state of the emerging beam. Table 1.6 lists the Jones vectors of various polarization states.

Let n_s and n_f be the refractive indices associated with the propagation of the slow and fast components, respectively. The polarization state of the emerging beam in the medium sf coordinate system is given by

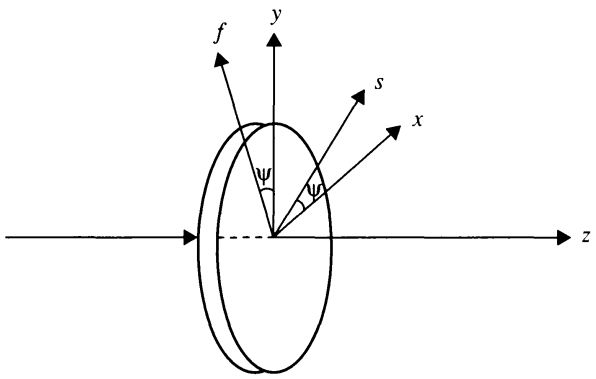
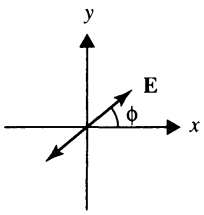
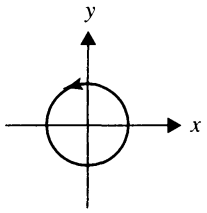
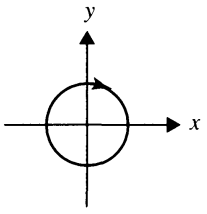
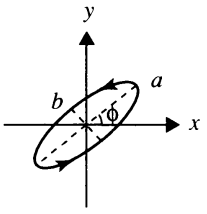
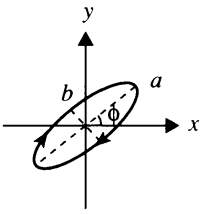


Figure 1.16 A wave plate (or retardation plate) with azimuth angle ψ. The s axis is in the direction of polarization of the slow mode, whereas the f axis is in the direction of polarization of the fast mode. The azimuth ψ is defined as the angle between the s axis and the x axis.

TABLE 1.6 Jones Vectors

Polarization State	Jones Vector
	$\begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix}$
	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -i \end{bmatrix}$
	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ i \end{bmatrix}$
	$\begin{bmatrix} a\cos\phi + ib\sin\phi \\ a\sin\phi - ib\cos\phi \end{bmatrix}$
	$\begin{bmatrix} a\cos\phi - ib\sin\phi \\ a\sin\phi + ib\cos\phi \end{bmatrix}$

$$\begin{bmatrix} V_s' \\ V_f' \end{bmatrix} = \begin{bmatrix} \exp\left(-in_s \dfrac{2\pi}{\lambda} d\right) & 0 \\ 0 & \exp\left(-in_f \dfrac{2\pi}{\lambda} d\right) \end{bmatrix} \begin{bmatrix} V_s \\ V_f \end{bmatrix} \tag{1.9-3}$$

where d is the thickness of the plate and λ is the wavelength of the light beam. The phase retardation is the difference of the exponents in Equation (1.9-3) and is defined as

$$\Gamma = \frac{2\pi}{\lambda}(n_s - n_f)d \tag{1.9-4}$$

Notice that the phase retardation Γ is a measure of the relative change in phase as a result of the propagation, not the absolute change. The birefringence of a typical retardation plate is small; that is, $|n_s - n_f| \ll n_s, n_f$. Consequently, the absolute change in phase caused by the plate may be hundreds of times greater than the phase retardati on. Let ϕ be the mean absolute phase change,

$$\phi = \tfrac{1}{2}(n_s + n_f)\frac{2\pi}{\lambda}d \tag{1.9-5}$$

Then Equation (1.9-3) can be written in terms of ϕ and Γ as

$$\begin{bmatrix} V_s' \\ V_f' \end{bmatrix} = e^{-i\phi} \begin{bmatrix} e^{-i\Gamma/2} & 0 \\ 0 & e^{i\Gamma/2} \end{bmatrix} \begin{bmatrix} V_s \\ V_f \end{bmatrix} \tag{1.9-6}$$

The Jones vector of the polarization state of the emerging beam in the xy coordinate is found by transforming back from the sf coordinate system:

$$\begin{bmatrix} V_x' \\ V_y' \end{bmatrix} = \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} V_s' \\ V_f' \end{bmatrix} \tag{1.9-7}$$

By combining Equations (1.9-1), (1.9-6), and (1.9-7), we can write the transformation due to the retardation plate as

$$\begin{bmatrix} V_x' \\ V_y' \end{bmatrix} = R(-\psi)W_0R(\psi) \begin{bmatrix} V_x \\ V_y \end{bmatrix} \tag{1.9-8}$$

where $R(\psi)$ is the coordinate rotation matrix and W_0 is the Jones matrix for the retardation plate. These are given, respectively, by

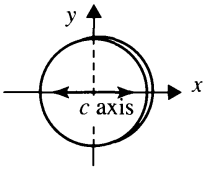
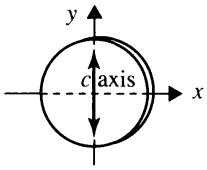
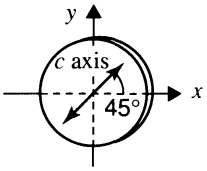
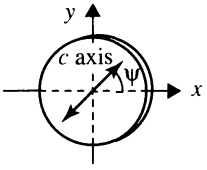
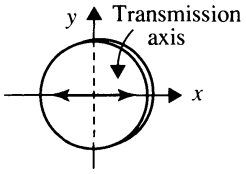
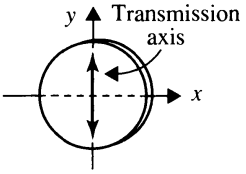
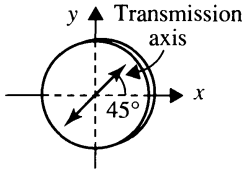
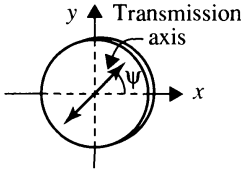
$$R(\psi) = \begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix} \tag{1.9-9}$$

and

$$W_0 = e^{-i\phi} \begin{bmatrix} e^{-i\Gamma/2} & 0 \\ 0 & e^{i\Gamma/2} \end{bmatrix} \tag{1.9-10}$$

The phase factor $e^{-i\phi}$ can be neglected if interference effects due to multiple reflections are not important, or not observable. A retardation plate is characterized by its phase retardation Γ and its azimuth angle ψ and is represented by the product of three matrices (Equation (1.9-8)):

TABLE 1.7 Jones Matrices

Optical Element	Jones Matrices
Wave plates	$\Gamma \equiv \dfrac{2\pi}{\lambda}(n_e - n_o)d$
	$\begin{bmatrix} e^{-i\Gamma/2} & 0 \\ 0 & e^{i\Gamma/2} \end{bmatrix}$
	$\begin{bmatrix} e^{i\Gamma/2} & 0 \\ 0 & e^{-i\Gamma/2} \end{bmatrix}$
	$\begin{bmatrix} \cos(\Gamma/2) & -i\sin(\Gamma/2) \\ -i\sin(\Gamma/2) & \cos(\Gamma/2) \end{bmatrix}$
	$R(-\psi)\begin{bmatrix} e^{-i\Gamma/2} & 0 \\ 0 & e^{i\Gamma/2} \end{bmatrix}R(\psi)$ $= \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix}\begin{bmatrix} e^{-i\Gamma/2} & 0 \\ 0 & e^{i\Gamma/2} \end{bmatrix}\begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix}$ $= \begin{bmatrix} e^{-i\Gamma/2}\cos^2\psi + e^{i\Gamma/2}\sin^2\psi & -i\sin(\Gamma/2)\sin(2\psi) \\ -i\sin(\Gamma/2)\sin(2\psi) & e^{-i\Gamma/2}\sin^2\psi + e^{i\Gamma/2}\cos^2\psi \end{bmatrix}$
Polarizers	
	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$
	$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$
	$R(-\psi)\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}R(\psi)$ $= \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix}$ $= \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix}\begin{bmatrix} \cos\psi & \sin\psi \\ 0 & 0 \end{bmatrix} = \begin{vmatrix} \cos^2\psi & \cos\psi\sin\psi \\ \sin\psi\cos\psi & \sin^2\psi \end{vmatrix}$

$$W = R(-\psi)W_0R(\psi) \tag{1.9-11}$$

Note that the Jones matrix of a wave plate is a unitary matrix; that is,

$$W^{\dagger}W = 1$$

where the dagger (†) means Hermitian conjugate. The passage of a beam of polarized light through a wave plate is mathematically described as a unitary transformation. Many physical properties are invariant under unitary transformations; these include the orthogonal relationship between the Jones vectors and the magnitude of the Jones vectors. Thus if the polarization states of two beams are mutually orthogonal, they will remain orthogonal after passing through an arbitrary wave plate.

The Jones matrix of an ideal homogeneous linear sheet polarizer oriented with its transmission axis parallel to the laboratory x axis is

$$P_0 = e^{-i\phi}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{1.9-12}$$

where ϕ is the absolute phase accumulated due to the finite optical thickness of the polarizer. A sheet polarizer is made of an anisotropic medium in which one of the normal modes suffers attenuation due to material absorption. In this case, the y component is absorbed by the sheet polarizer. The Jones matrix of a polarizer rotated by an angle ψ about the z axis is given by

$$P = R(-\psi)P_0R(\psi) \tag{1.9-13}$$

Thus, if we neglect the absolute phase ϕ, the Jones matrix representations of the polarizers transmitting light with electric field vectors parallel to the x and y axes, respectively, are given by

$$P_x = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad P_y = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

To find the effect of a train of retardation plates and polarizers on the polarization state of a polarized light beam, we write down the Jones vector of the incident beam, and then write down the Jones matrices of the various elements. The Jones vector of the emerging beam is obtained by carrying out the matrix multiplication in sequence. Table 1.7 lists the Jones matrices of wave plates and polarizers at selected orientations.

EXAMPLE: A HALF-WAVE RETARDATION PLATE

A half-wave plate has a phase retardation of $\Gamma = \pi$. According, to Equation (1.9-4), an a-cut (surface is perpendicular to a axis) uniaxial plate will act as a half-wave plate provided the thickness is $d = \lambda/(2|n_e - n_o|)$. We will determine the effect of a half-wave plate on the polarization state of a transmitted light beam. The azimuth angle of the wave plate is taken as 45°, and the incident beam is vertically polarized. The Jones vector for the incident beam can be written

$$\mathbf{V} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{1.9-14}$$

and the Jones matrix for the half-wave plate is obtained by using Equations (1.9-9) to (1.9-11):

$$W = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}\begin{bmatrix}-i & 0\\ 0 & i\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix} = \begin{bmatrix}0 & -i\\ -i & 0\end{bmatrix} \tag{1.9-15}$$

The Jones vector for the emerging beam is obtained by multiplying Equations (1.9-15) and (1.9-14); the result is

$$\mathbf{V}' = \begin{bmatrix}-i\\ 0\end{bmatrix} = -i\begin{bmatrix}1\\ 0\end{bmatrix} \tag{1.9-16}$$

This represents a beam of horizontally polarized light. The effect of the half-wave plate is to rotate the polarization by 90°. It can be shown that for a general azimuth angle ψ, the half-wave plate will rotate the polarization by an angle 2ψ (see Problem 1.19). In other words, linearly polarized light remains linearly polarized, except that the plane of polarization is rotated by an angle of 2ψ.

When the beam of incident light is circularly polarized, a half-wave plate will convert a beam of right-handed circularly polarized light into a beam of left-handed circularly polarized light and vice versa, regardless of the azimuth angle. The proof is left as an exercise (see Problem 1.19). Figure 1.17 illustrates the effect of a half-wave plate.

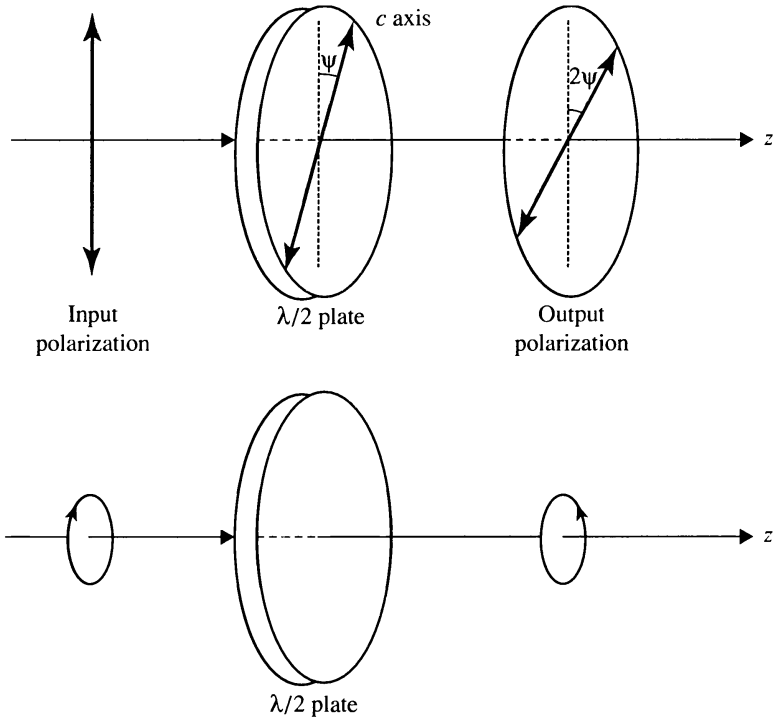


Figure 1.17 The effect of a half-wave plate on the polarization state of a beam.

EXAMPLE: A QUARTER-WAVE PLATE

A quarter-wave plate has a phase retardation of $\Gamma = \pi/2$. If the plate is made of an a-cut (or b-cut) uniaxially anisotropic medium, the thickness is $d = \lambda/(4|n_e - n_o|)$ (or odd multiples thereof). Suppose again that the azimuth angle of the plate is $\psi = 45°$ and the incident beam is vertically polarized. The Jones vector for the incident beam is given again by Equation (1.9-14). The Jones matrix for this quarter-wave plate, according to Equation (1.9-11), is

$$W = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}\begin{bmatrix}e^{-i\pi/4} & 0\\ 0 & e^{i\pi/4}\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & -i\\ -i & 1\end{bmatrix} \tag{1.9-17}$$

The Jones vector for the emerging beam is obtained by multiplying Equations (1.9-17) and (1.9-14) and is given by

$$\mathbf{V}' = \frac{1}{\sqrt{2}}\begin{bmatrix}-i\\ 1\end{bmatrix} = \frac{-i}{\sqrt{2}}\begin{bmatrix}1\\ i\end{bmatrix} \tag{1.9-18}$$

This represents a beam of left-handed circularly polarized light. The effect of a 45° oriented quarter-wave plate is to convert a beam of vertically polarized light into a beam of left-handed circularly polarized light. If the incident beam is horizontally polarized, the emerging beam will be right-handed circularly polarized. In general, a quarter-wave plate can convert a linearly polarized light into an elliptically polarized light and vice versa. The effect of the quarter-wave plate is illustrated in Figure 1.18.

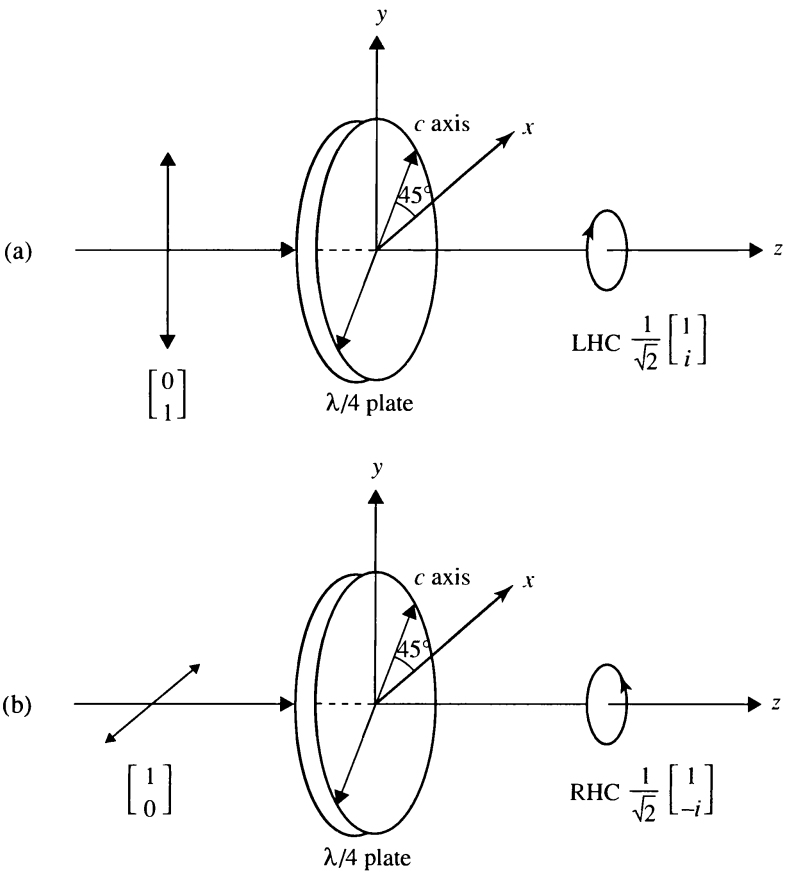


Figure 1.18 The effect of a quarter-wave plate (assuming $\Gamma = k(n_e - n_o)d = \pi/2$) on the polarization state of a linearly polarized beam. LHE = left-handed elliptically, RHE = right-handed elliptically.

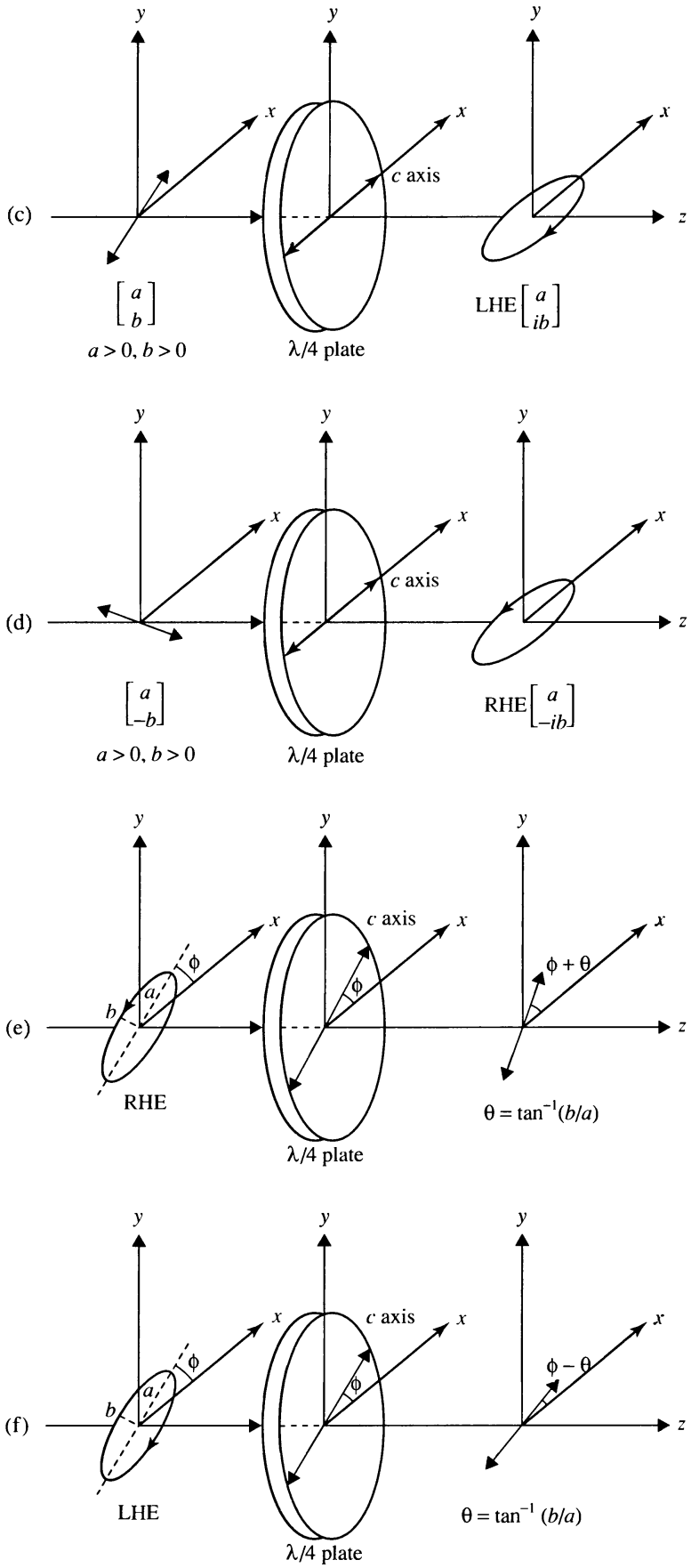


Figure 1.18 *(cont'd)*

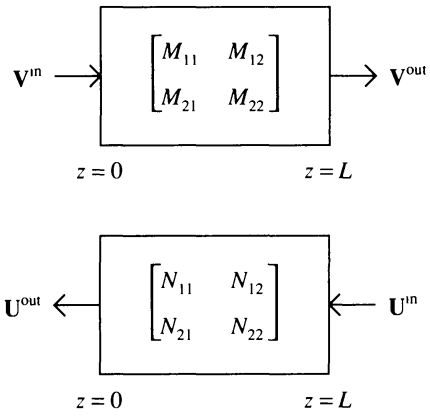


Figure 1.19 Schematic drawing of input–output relationships and definition of Jones matrices M and N.

General Properties of the Jones Matrix

Here, we consider some general properties of the Jones matrix. Specifically, we consider the Jones matrix under the transformation of retroreflection, mirror reflection, and time reversal and then we discuss the principle of reciprocity. We also consider the unimodular and unitary nature of the Jones matrix. The results are extremely useful in understanding the transmission properties of birefringent networks. We consider a general birefringent system (between $z = 0$ and $z = L$), which consists of a series of anisotropic wave plates (e.g., wave plates, birefringent fibers, LC cells) with arbitrary orientations of the optical axes. Referring to Figure 1.19, we consider the situation of incidence from both ends of a general birefringent system. For the purpose of discussion, we define two Jones matrices as shown in the figure.

1. Incidence from the left side ($z = 0$): The relationship between the input Jones vector $\mathbf{V}^{in}$ at ($z = 0$) and the output Jones vector $\mathbf{V}^{out}$ at ($z = L$) is written

$$\begin{bmatrix} V_x^{out} \\ V_y^{out} \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} V_x^{in} \\ V_y^{in} \end{bmatrix} \tag{1.9-19}$$

   where M is the Jones matrix for incidence from the left side.

2. Incidence from the right side ($z = L$): The relationship between the input Jones vector $\mathbf{U}^{in}$ at ($z = L$) and the output Jones vector $\mathbf{U}^{out}$ at ($z = 0$) is written

$$\begin{bmatrix} U_x^{out} \\ U_y^{out} \end{bmatrix} = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} U_x^{in} \\ U_y^{in} \end{bmatrix} \tag{1.9-20}$$

   where N is the Jones matrix for incidence from the right side. Here we assume that the beam of light retraces the path defined by the incidence from $z = 0$ in Case 1. In other words, Case 2 can be obtained by a retroreflection of the beam of light in Case 1.

Based on the above definitions, the following are fundamental properties of these Jones matrices.

Time Reversal Symmetry

Consider the situation when the time is reversed. The output beam will retrace the beam path, propagate through the birefringent system, and become the phase conjugate of the input

beam. In this case, $\mathbf{U}^{in} = (\mathbf{V}^{out})^*$. As a result of the time reversal symmetry, $\mathbf{U}^{out} = (\mathbf{V}^{in})^*$. This leads to

$$NM^* = 1 \tag{1.9-21}$$

Principle of Reciprocity

Based on the fundamental principle of reciprocity in physics and the definition of the Jones matrices in Equations (1.9-19) and (1.9-20), we have

$$\begin{aligned} N_{11} &= M_{11} \\ N_{22} &= M_{22} \\ N_{21} &= M_{12} \\ N_{12} &= M_{21} \end{aligned} \tag{1.9-22}$$

In other words, N is the transpose of M,

$$N = \tilde{M} \tag{1.9-23}$$

where the notation ~ indicates a transpose operation. This property is particularly useful in optical systems that involve the use of a reflector (mirror).

Using Equations (1.9-21), (1.9-22), and (1.9-23), it can be shown that both N and M are unitary matrices; that is,

$$\begin{aligned} M^\dagger M &= 1 \\ N^\dagger N &= 1 \end{aligned} \tag{1.9-24}$$

where $N^\dagger$ and $M^\dagger$ are the Hermitian conjugates of N and M, respectively. This proves the unitary nature of the Jones matrix. Thus, if the Jones matrix M is written

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{1.9-25}$$

where A, B, C, and D are the matrix elements, then the inverse of M can be written

$$M^{-1} = M^\dagger = \begin{bmatrix} A^* & C^* \\ B^* & D^* \end{bmatrix} \tag{1.9-26}$$

Furthermore, the Jones matrix is unimodular based on the definition in Equation (1.9-11), provided all the birefringent plates are lossless. In other words,

$$\det(M) = AD - BC = 1 \tag{1.9-27}$$

Thus the inverse of the Jones matrix can also be written

$$M^{-1} = \begin{bmatrix} D & -B \\ -C & A \end{bmatrix} \tag{1.9-28}$$

Using Equations (1.9-26) and (1.9-28), we obtain the following relationships for the matrix elements:

$$\begin{aligned} C &= -B^* \\ D &= A^* \end{aligned} \tag{1.9-29}$$

These relationships are very useful in simplifying calculations involving use of the Jones matrix method. Based on the relationships in Equation (1.9-29), the Jones matrix can be written

$$M = \begin{bmatrix} A & B \\ -B^* & A^* \end{bmatrix} \tag{1.9-30}$$

We note that the collection of all Jones matrices forms a mathematical group. In other words, the relationships remain true after the multiplication of any two Jones matrices.

Intensity Transmission Spectrum

So far our development of the Jones matrix method was concerned with the polarization state of the light beam. In many cases, we need to determine the transmitted intensity. A narrowband filter, for example, transmits radiation only in a small spectral regime and rejects (or absorbs) radiation at other wavelengths. To change the intensity of the transmitted beam, an analyzer is usually required. An analyzer is basically a polarizer. It is called an analyzer simply because of its location in the optical system. In most birefringent optical systems, a polarizer is placed in front of the system in order to "prepare" a beam of polarized light. A second polarizer (analyzer) is placed at the output to analyze the polarization state of the emerging beam. Because the phase retardation of each wave plate is wavelength dependent, the polarization state of the emerging beam depends on the wavelength of the light. A polarizer at the rear will cause the overall transmitted intensity to be wavelength dependent.

The Jones vector representation of a beam of light contains information about not only the polarization state but also the intensity of light. Let us now consider the light beam after it passes through the polarizer. Its electric vector can be written as a Jones vector

$$\mathbf{E} = \begin{bmatrix} E_x \\ E_y \end{bmatrix} \tag{1.9-31}$$

where E_x, E_y are the components in the xy coordinate. The intensity is calculated as follows:

$$I = \mathbf{E}^\dagger \cdot \mathbf{E} = |E_x|^2 + |E_y|^2 \tag{1.9-32}$$

where the dagger indicates the Hermitian conjugate. If the Jones vector of the emerging beam after it passes through the analyzer is written

$$\mathbf{E}' = \begin{bmatrix} E_x' \\ E_y' \end{bmatrix} \tag{1.9-33}$$

the transmittance of the birefringent optical system is calculated as

$$T = \frac{|E_x'|^2 + |E_y'|^2}{|E_x|^2 + |E_y|^2} \tag{1.9-34}$$

EXAMPLE: A BIREFRINGENT PLATE SANDWICHED BETWEEN PARALLEL POLARIZERS

Referring to Figure 1.20, we consider a birefringent plate sandwiched between a pair of parallel polarizers. The plate is oriented so that the "slow" and "fast" axes are at 45° with respect to the transmission axes of polarizers. Let the birefringence be $n_e - n_o$ and the plate thickness be d. The phase retardation is then given by

$$\Gamma = \frac{2\pi}{\lambda}(n_e - n_o)d \tag{1.9-35}$$

and the corresponding Jones matrix is, according to Equation (1.9-11) (or Table 1.7),

$$W = \begin{bmatrix} \cos(\Gamma/2) & -i\sin(\Gamma/2) \\ -i\sin(\Gamma/2) & \cos(\Gamma/2) \end{bmatrix} \tag{1.9-36}$$

Let the incident beam be unpolarized, so that after it passes through the front polarizer, the electric field vector can be represented by the following Jones vector:

$$\frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{1.9-37}$$

where we assume that the intensity of the incident beam is unity and only half of the intensity passes through the polarizer where the transmission axis is parallel to the y axis. The Jones vector representation of the electric field vector of the transmitted beam is obtained as follows:

$$\mathbf{E}' = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \cos(\Gamma/2) & -i\sin(\Gamma/2) \\ -i\sin(\Gamma/2) & \cos(\Gamma/2) \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ \cos(\Gamma/2) \end{bmatrix} \tag{1.9-38}$$

The transmitted beam is vertically (y) polarized with an intensity given by

$$I = \tfrac{1}{2}\cos^2(\Gamma/2) = \tfrac{1}{2}\cos^2[\pi(n_e - n_o)d/\lambda] \tag{1.9-39}$$

It can be seen from Equation (1.9-39) that the transmitted intensity is a sinusoidal function of the wavenumber and peaks at $\lambda = (n_e - n_o)d, (n_e - n_o)d/2, (n_e - n_o)d/3, \ldots$. These wavelengths correspond to $\Gamma = 2\pi, 4\pi, 6\pi, \ldots$. In other words, maximum transmission occurs when the plate is an integral number of full waves. The wavenumber separation between transmission maxima increases with decreasing plate thickness.

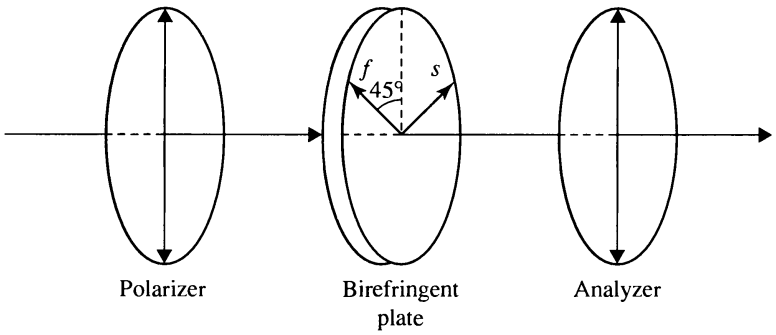


Figure 1.20 A birefringent plate sandwiched between a pair of parallel polarizers.

EXAMPLE: A BIREFRINGENT PLATE SANDWICHED BETWEEN A PAIR OF CROSSED POLARIZERS

If we rotate the analyzer shown in Figure 1.20 by 90°, then the input and output polarizers are crossed. The transmitted beam for this case is obtained as follows:

$$\mathbf{E}' = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \cos(\Gamma/2) & -i\sin(\Gamma/2) \\ -i\sin(\Gamma/2) & \cos(\Gamma/2) \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \frac{-i}{\sqrt{2}} \begin{bmatrix} \sin(\Gamma/2) \\ 0 \end{bmatrix} \tag{1.9-40}$$

The transmitted beam is horizontally (x) polarized with an intensity given by

$$I = \tfrac{1}{2}\sin^2(\Gamma/2) = \tfrac{1}{2}\sin^2[\pi(n_e - n_o)d/\lambda] \tag{1.9-41}$$

This is again a sinusoidal function of the wavenumber. The transmission spectrum consists of a series of maxima at $\lambda = 2(n_e - n_o)d$, $2(n_e - n_o)d/3, \ldots$. These wavelengths correspond to phase retardations of $\pi, 3\pi, 5\pi, \ldots$, that is, when the wave plate becomes a half-wave plate or an odd integral multiple of a half-wave plate.

EXAMPLE: A BIREFRINGENT PLATE SANDWICHED BETWEEN A PAIR OF POLARIZERS

If we rotate the analyzer shown in Figure 1.20 so that the transmission axis of the analyzer forms an angle ψ with respect to the x axis, then the input and output polarizers are neither parallel nor crossed. The transmitted beam for this case is obtained as follows:

$$\mathbf{E}' = \begin{bmatrix} \cos^2\psi & \cos\psi\sin\psi \\ \sin\psi\cos\psi & \sin^2\psi \end{bmatrix} \begin{bmatrix} \cos(\Gamma/2) & -i\sin(\Gamma/2) \\ -i\sin(\Gamma/2) & \cos(\Gamma/2) \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \frac{-i\cos\psi\sin(\Gamma/2) + \sin\psi\cos(\Gamma/2)}{\sqrt{2}} \begin{bmatrix} \cos\psi \\ \sin\psi \end{bmatrix} \tag{1.9-42}$$

The transmitted beam is polarized in the same direction as the transmission axis of the analyzer with an intensity

$$I = \tfrac{1}{2}\cos^2\psi\sin^2(\Gamma/2) + \tfrac{1}{2}\sin^2\psi\cos^2(\Gamma/2) \tag{1.9-43}$$

where

$$\Gamma = \frac{2\pi(n_e - n_o)d}{\lambda}$$

We note that Equation (1.9-43) reduces to Equation (1.9-39) when $\psi = \pi/2$ (i.e., when the analyzer is parallel to the polarizer) and Equation (1.9-43) reduces to Equation (1.9-41) when $\psi = 0$ (i.e., when the analyzer is perpendicular to the polarizer).

1.10 ELEMENTARY THEORY OF COHERENCE

It is known that the principle of superposition applies to the electromagnetic field. This principle states that the total field due to all sources is the sum of the field due to each source. The reason why this is true is that Maxwell's equations, which govern the electromagnetic field, are linear differential equations, provided both ε and μ are independent of the field strength. Consider now two linearly polarized plane waves of the same frequency ω. Let the electric fields be

$$\begin{aligned} \mathbf{E}_1 &= \mathbf{A}_1 \exp[i(\omega t - \mathbf{k}_1 \cdot \mathbf{r} + \phi_1)] \\ \mathbf{E}_2 &= \mathbf{A}_2 \exp[i(\omega t - \mathbf{k}_2 \cdot \mathbf{r} + \phi_2)] \end{aligned} \tag{1.10-1}$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are the amplitudes and are real vectors because of the assumption of linear polarization states. The real quantities ϕ_1 and ϕ_2 are the phases, and $\mathbf{k}_1$ and $\mathbf{k}_2$ are the wavevectors. We assume that the amplitudes $\mathbf{A}_1$ and $\mathbf{A}_2$ are constants. The sources of these two planes waves are said to be mutually coherent if the phase difference $\phi_1 - \phi_2$ is constant. On the other hand, if the quantity $\phi_1 - \phi_2$ varies with time in a random fashion, the sources of these two plane waves are said to be mutually incoherent. To define the degree of mutual coherence of these two waves (1.10-1), we can examine the interference pattern formed by these two waves. We have shown earlier that the time-averaged intensity of electromagnetic radiation at any point in space is proportional to the square of the amplitude of the electric field. Thus, aside from a constant factor, the time-averaged intensity distribution of the interference pattern formed by the superposition of these two waves can be written

$$I = |\mathbf{E}_1 + \mathbf{E}_2|^2 = I_1 + I_2 + 2\mathbf{A}_1 \cdot \mathbf{A}_2 \cos(\mathbf{K} \cdot \mathbf{r} - \phi) \tag{1.10-2}$$

where $I_1 = |\mathbf{E}_1|^2$, $I_2 = |\mathbf{E}_2|^2$, $\mathbf{K} = \mathbf{k}_1 - \mathbf{k}_2$, and $\phi = \phi_1 - \phi_2$. The first two terms are the time-averaged intensities of the two waves, respectively. The third term is the interference term and contains information about the degree of mutual coherence. Note that Equation (1.10-2) is a time-averaged result, averaging over a period of $2\pi/\omega$. If we limit ourselves to the case of quasi-monochromatic plane waves, all the phases ϕ, ϕ_1, and ϕ_2 are considered constant during such a small time interval $2\pi/\omega$, which is on the order of 10^{-15} s for visible light. If these two waves are of the same polarization state, then the intensity of the interference pattern can be written

$$I = |\mathbf{E}_1 + \mathbf{E}_2|^2 = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\mathbf{K} \cdot \mathbf{r} - \phi) \tag{1.10-3}$$

If the two waves are mutually coherent (i.e., ϕ = constant), the interference pattern is stationary with a spatial period of

$$\Lambda = \frac{2\pi}{|\mathbf{K}|} = \frac{\lambda}{2\sin(\theta/2)} \tag{1.10-4}$$

where θ is the angle between the two wavevectors, and λ is the wavelength of the light (see Figure 1.21).

If the two waves are not mutually coherent, the phase ϕ varies with time. As a result, the time-averaged intensity in Equation (1.10-3) fluctuates rapidly with time. If ϕ varies significantly over a time period τ_D, the characteristic time constant associated with the detection process, the fluctuation is too fast for the detector to respond, and the detected intensity is averaged over a time interval τ_D. As a result of this time averaging, the visibility of the fringes decreases. For mutually incoherent sources, the phase ϕ varies with time in a

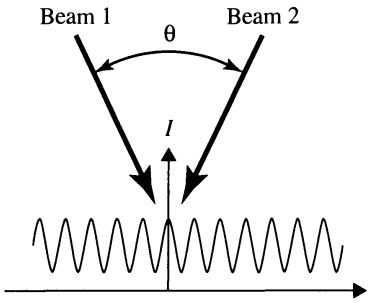


Figure 1.21 Intensity distribution of a fringe pattern due to the interference of two beams.

random fashion between 0 and 2π, and the averaging of the third term in Equation (1.10-3) is obviously zero. This leads to the disappearance of the fringe pattern. Thus, for many practical purposes, the degree of mutual coherence between the two waves (1.10-1) can be defined as

$$\gamma_{12} = \left(\frac{\langle\langle \mathbf{E}_1^* \cdot \mathbf{E}_2 \rangle\rangle}{\langle\langle \mathbf{E}_1^* \cdot \mathbf{E}_1 \rangle\rangle^{1/2} \langle\langle \mathbf{E}_2^* \cdot \mathbf{E}_2 \rangle\rangle^{1/2}} \right)_{\mathbf{r}=0} = \langle\langle e^{i(\phi_2 - \phi_1)} \rangle\rangle = \frac{1}{\tau_D} \int_0^{\tau_D} e^{i(\phi_2 - \phi_1)}\, dt \qquad (1.10\text{-}5)$$

where the double brackets indicate time averaging over a time interval τ_D, the characteristic time constant of the detection system. The parameter is, in general, complex and can be written

$$\gamma_{12} = |\gamma_{12}| \exp(i\alpha) \qquad (1.10\text{-}6)$$

where α is a real constant.

We now evaluate the detection of the time-averaged intensity. From Equation (1.10-2), we obtain

$$\langle\langle I \rangle\rangle = |\mathbf{E}_1 + \mathbf{E}_2|^2 = I_1 + I_2 + 2\mathbf{A}_1 \cdot \mathbf{A}_2 \langle\langle \cos(\mathbf{K} \cdot \mathbf{r} - \phi) \rangle\rangle \qquad (1.10\text{-}7)$$

where we assumed $\mathbf{A}_1$ and $\mathbf{A}_2$ are constants.

Using $\cos(\mathbf{K} \cdot \mathbf{r} - \phi) = \cos(\mathbf{K} \cdot \mathbf{r}) \cos\phi + \sin(\mathbf{K} \cdot \mathbf{r}) \sin\phi$, and Equation (1.10-5), we obtain

$$\langle\langle I \rangle\rangle = |\mathbf{E}_1 + \mathbf{E}_2|^2 = I_1 + I_2 + 2\mathbf{A}_1 \cdot \mathbf{A}_2 |\gamma_{12}| \cos(\mathbf{K} \cdot \mathbf{r} - \alpha) \qquad (1.10\text{-}8)$$

Thus an interference fringe pattern will result if $|\gamma_{12}|$, called the degree of mutual coherence, has a value other than zero. In other words, a periodic variation of intensity exists provided the mutual coherence of the two beams is nonzero. By definition (1.10-5), the magnitude of $|\gamma_{12}|$ is always between zero and one. In terms of $|\gamma_{12}|$, we have the following types of mutual coherence:

Complete mutual coherence:	$\lvert\gamma_{12}\rvert = 1$
Partial coherence:	$0 < \lvert\gamma_{12}\rvert < 1$
Complete mutual incoherence:	$\lvert\gamma_{12}\rvert = 0$

Fringe Visibility

In a pattern of interference fringes, the intensity varies between two limits $I_{\max}$ and $I_{\min}$. From the above discussion, we note that these two limits are governed by

$$I_{max} = I_1 + I_2 + 2|\mathbf{A}_1 \cdot \mathbf{A}_2||\gamma_{12}| \tag{1.10-9}$$

$$I_{min} = I_1 + I_2 - 2|\mathbf{A}_1 \cdot \mathbf{A}_2||\gamma_{12}| \tag{1.10-10}$$

The fringe visibility V is defined as the ratio

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{1.10-11}$$

For beams with the same polarization state (i.e., $\mathbf{A}_1$ is parallel to $\mathbf{A}_2$), it follows that

$$V = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}|\gamma_{12}| \tag{1.10-12}$$

The fringe visibility V is always between 0 and 1. In particular, if $I_1 = I_2$, then

$$V = |\gamma_{12}| \tag{1.10-13}$$

In other words, the interference fringe visibility is equal to the degree of mutual coherence. In the case of complete mutual coherence ($|\gamma_{12}| = 1$), the interference fringes have the maximum visibility of unity; whereas for beams of complete incoherence, $|\gamma_{12}| = 0$, the fringe visibility is zero—in other words, there are no interference fringes at all.

Coherence Time and Coherence Length

In addition to the mutual coherence between two beams, we can define the coherence of electromagnetic radiation itself. Consider a quasi-monochromatic wave of the form

$$E = A(t) \exp[i(\omega t - \mathbf{k}_1 \cdot \mathbf{r})] \tag{1.10-14}$$

where $A(t)$ is a time-varying complex amplitude. In the two-beam interference discussed earlier, the two electric fields E_1 and E_2 may originate from the same source given by Equation (1.10-14). This happens in many interferometers, including the Michelson interferometer, the Mach–Zehnder interferometer, and Young's double-slit interferometer. In these interference experiments, the two beams of light are only different in their optical paths. Referring to Figure 1.21, we assume that the two beams are

$$E_1 = A(t) \exp[i(\omega t - \mathbf{k}_1 \cdot \mathbf{r})] \tag{1.10-15}$$

$$E_2 = A(t + \tau) \exp[i(\omega t + \omega\tau - \mathbf{k}_2 \cdot \mathbf{r})] \tag{1.10-16}$$

where τ is the time delay. We can see that these two waves originate from the same source (1.10-14). The degree of mutual coherence between these two waves is a measure of the self-coherence of the wave (1.10-14). Thus self-coherence $\gamma(\tau)$ is defined as

$$\gamma(\tau) = \frac{\langle\langle E^*(t)\, E(t + \tau)\rangle\rangle}{\langle\langle E^*(t)\, E(t)\rangle\rangle^{1/2} \langle\langle E^*(t + \tau)\, E(t + \tau)\rangle\rangle^{1/2}} \tag{1.10-17}$$

Assume that all quantities are stationary; that is, the time average is independent of the choice of the origin. Then the self-coherence function $\gamma(\tau)$ becomes

$$\gamma(\tau) = \frac{\langle\langle E^*(t)\, E(t + \tau)\rangle\rangle}{\langle\langle E^*(t)\, E(t)\rangle\rangle} \tag{1.10-18}$$

The self-coherence function defined earlier also satisfies the condition

$$0 \le |\gamma(\tau)| \le 1 \tag{1.10-19}$$

For monochromatic plane waves, $A(t)$ is a constant amplitude, and the self-coherence function is unity for any τ. Such a wave is called coherent radiation. For most polychromatic waves, $|\gamma(\tau)| = 1$ only occurs when $\tau = 0$, and $|\gamma(\tau)|$ approaches zero when τ becomes large. Such waves are called partially coherent radiation. If $|\gamma(\tau)|$ is a monotonically decreasing function of τ and, in addition, if $|\gamma(\tau)|$ drops to virtually zero for $\tau > \tau_c$, then we say that τ_c is the coherence time of the wave. Thus, in the two-beam interference experiment, if the path difference between the two beams does not exceed the value

$$l_c = c\tau_c \tag{1.10-20}$$

interference fringes appear. The quantity l_c is known as the coherence length of the wave in Equation (1.10-14).

For many practical purposes, the time interval for averaging in Equation (1.10-17) or (1.10-18) can be taken as infinity. Thus the self-coherence function can be considered as the normalized autocorrelation function of the electromagnetic field. A basic theorem of stochastic theory, the so-called Wiener–Khintchine theorem, states that the power spectrum (or spectral density) of the electromagnetic field and the autocorrelation function form a Fourier transform pair. Therefore $\gamma(\tau)$, as given by Equation (1.10-18), is the normalized Fourier transform of the power spectrum of the electromagnetic field. Let $\Delta\omega$ be the width of the power spectral distribution of the field. Then, according to the theory of Fourier transform, we have

$$\tau_c = \frac{2\pi}{\Delta\omega} = \frac{1}{\Delta f} \tag{1.10-21}$$

which states that the coherence time of quasi-monochromatic radiation is given by the inverse of its spectral linewidth Δf.

PROBLEMS

1.1 Conservation of charges requires that the charge density at any point in space be related to the current density in that neighborhood by a continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0$$

Derive this equation from Maxwell's equation. If we integrate this equation over a volume element, the first term represents the change of total electric charge per unit time within the volume element; the second term represents the amount of charge flowing out of the volume element per unit time.

1.2 Derive the boundary conditions for the normal components of **B** and **D**, by using an application of the Gauss divergence theorem to Maxwell's equations.

1.3 Derive the boundary conditions that apply to the tangential components of **E** and **H** by using an application of the Stokes theorem to Maxwell's equations.

1.4 Let $\mathbf{E} = \mathbf{E}_0 \exp(i\omega t)$ and $\mathbf{H} = \mathbf{H}_0 \exp(i\omega t)$ be solutions to Maxwell's equations.

(a) Show that $\mathbf{E}^*$ and $\mathbf{H}^*$ also satisfy Maxwell's equations. Note that $\mathbf{E}^*$, $\mathbf{H}^*$ and **E**, **H** are, in fact, the same field since only the real parts have physical meaning.

(b) Show that the conjugate waves

$$\mathbf{E}_c = \mathbf{E}_0^* \exp(i\omega t) \quad \text{and} \quad \mathbf{H}_c = \mathbf{H}_0^* \exp(i\omega t)$$

also satisfy Maxwell's equations, provided the medium is lossless (i.e., ε and μ are real tensors.)

1.5 The electromagnetic-field momentum density and Maxwell's stress tensor are given by

$$\mathbf{P} = \mu\varepsilon(\mathbf{E} \times \mathbf{H})$$

$$T_{ij} = \varepsilon E_i E_j + \mu H_i H_j - \tfrac{1}{2}\delta_{ij}(\varepsilon E^2 + \mu H^2)$$

Derive the hydrodynamical equation of motion

$$\frac{\partial \mathbf{P}}{\partial t} = \nabla \cdot \mathbf{T} - \mathbf{F}$$

where **F** is the Lorentz force exerted on a distribution of charges and currents by the electromagnetic field, $\mathbf{F} = \rho\mathbf{E} + \mathbf{J} \times \mathbf{B}$.

1.6 Consider a charge particle with a charge q under the influence of an electric field E.

(a) Show that the work done per unit time by the field on the particle can be written work/time $= \mathbf{E} \cdot q\mathbf{v} = \mathbf{E} \cdot q\, d\mathbf{x}/dt$, where $\mathbf{v} = d\mathbf{x}/dt$ is the velocity of the charged particle.

(b) Consider a material medium that consists of a large number of charged particles. Show that the total work done per unit time is

$$\frac{dW}{dt} = \mathbf{E} \cdot \sum_i q_i \frac{d\mathbf{x}_i}{dt}$$

where i stands for the ith particle, and the summation is over all charged particles.

(c) The polarization **P** is defined as the dipole moment per unit volume. Show that

$$\frac{dW}{dt}\frac{1}{\text{Volume}} = \mathbf{E} \cdot \frac{d\mathbf{P}}{dt}$$

1.7 Consider the propagation of an optical pulse in a homogeneous medium of length L.

(a) Show that the time of propagation is

$$\tau = L\left(\frac{n}{c} + \frac{\omega}{c}\frac{dn}{d\omega}\right) = L\left(\frac{n}{c} - \frac{\lambda}{c}\frac{dn}{d\lambda}\right)$$

(b) Show that the group velocity dispersion defined as

$$D = \frac{d}{d\lambda}\left(\frac{\tau}{L}\right)$$

can be written

$$D = -\frac{1}{c\lambda}\left(\lambda^2 \frac{d^2n}{d\lambda^2}\right)$$

Note that the term in parentheses is dimensionless. The parameter D has a dimension of s/m-m. For practical application, D is often expressed in units of ps/nm-km.

Consider the superposition of two waves. Let

$$v_1 = \frac{\omega_1}{k_1}, \quad v_2 = \frac{\omega_2}{k_2}, \quad v_g = \frac{\omega_2 - \omega_1}{k_2 - k_1}$$

Show that

$$(v_2 - v_g)(v_1 - v_g) > 0$$

In other words, the group velocity is either greater than both v_1 and v_2, or less than both v_1 and v_2. In normally dispersive media, the group velocity is always less than both v_1 and v_2.

1.8 Using Equation (1.5-20), plot the following in the spectral regime from $\lambda = 0.5$ to 2.0 μm.

(a) Plot n versus λ.

(b) Plot group velocity (in units of c) versus λ.

(c) Plot D (in units of ps/nm-km) versus λ.

1.9 Derive the polarization ellipse, Equation (1.6-12).

1.10 Derive the major and minor axes, and the inclination angle of the polarization ellipse, Equations (1.6-14) and (1.6-15), respectively.

1.11 Show that the end point of the electric vector of an elliptically polarized light beam will revolve in a clockwise direction if $\sin\delta > 0$ and in a counterclockwise direction if $\sin\delta < 0$.

1.12

(a) Find a polarization state that is orthogonal to the polarization state

$$\mathbf{J}(\psi, \delta) = \begin{bmatrix} \cos\psi \\ e^{i\delta}\sin\psi \end{bmatrix}$$

Answer: $\begin{bmatrix} \sin\psi \\ e^{i(\pi+\delta)}\cos\psi \end{bmatrix}$

(b) Show that the major axes of the ellipses of two mutually orthogonal polarization states are perpendicular to each other and the senses of revolution are opposite.

1.13 Derive the inclination angle and the ellipticity angle of the polarization ellipse, Equations (1.6-18) and (1.6-19), respectively.

1.14 Consider two monochromatic plane waves of the form

$$\mathbf{E}_a(z, t) = \text{Re}[\mathbf{A}e^{i(\omega t - kz)}] \quad \text{and} \quad \mathbf{E}_b(z, t) = \text{Re}[\mathbf{B}e^{i(\omega t - kz)}]$$

The polarization states of these two waves are orthogonal; that is, $\mathbf{A}^* \cdot \mathbf{B} = 0$.

(a) Let δ_a and δ_b be the phase angles defined in Equation (1.6-5). Show that

$$\delta_a - \delta_b = \pm\pi$$

(b) Since δ_a and δ_b are both in the range $-\pi < \delta \leq \pi$, show that

$$\delta_a\delta_b \leq 0$$

(c) Let χ_a and χ_b be the complex numbers representing the polarization states of these two waves. Show that

$$\chi_a^* \chi_b = -1$$

(d) Show that the major axes of the polarization ellipses are mutually orthogonal and the ellipticities are of the same magnitude with opposite signs.

1.15 Show that any polarization state can be converted into a linearly polarized state by using a quarter-wave plate. Describe your approach.

1.16 Draw the end point of the electric field vector in space for right-handed circularly polarized light at a given point in time (say, $t = 0$). Show that the locus is a left-handed helix. (This has been the traditional way of defining the handedness of circular polarization states. The advantage of this convention is that the handedness of the space helix is independent of the direction of viewing.)

(a) Show that a linear combination of two beams of circularly polarized light with equal amplitude and opposite handedness is a beam of plane polarized light. What determines the plane of polarization of the resultant beam?

(b) Show that the combination of two beams of elliptically polarized light is in general another beam of elliptically polarized light. Write down the conditions for which the resultant beam shall be plane polarized or circularly polarized.

1.17

(a) Show that it is impossible to distinguish between a beam of unpolarized light and a beam of circularly polarized light by using linear polarizers. Generally, wave plates and polarizers are needed to determine the degree of polarization.

(b) Given a beam of completely polarized light (say, elliptically polarized), show that the ellipticity and the orientation of the principal axes can be measured by using a single linear polarizer.

1.18 Use the following basis

$$\mathbf{E}_1 = (a, ib) \quad \text{and} \quad \mathbf{E}_2 = (b, -ia)$$

where a and b are real.

(a) Let $\mathbf{E} = (\cos\psi, \sin\psi) = c_1\mathbf{E}_1 + c_2\mathbf{E}_2$. Find c_1 and c_2.

(b) Show that a beam of elliptically polarized light will maintain its polarization ellipse after transmitting through an optically active medium. The angle of rotation of the major axis of the polarization ellipse is the same as that of a beam of linearly polarized light.

1.19 A half-wave plate has a phase retardation of $\Gamma = \pi$. Assume that the plate is oriented so that the azimuth angle (i.e., the angle between the x axis and the slow axis of the plate) is ψ.

(a) Find the polarization state of the transmitted beam, assuming that the incident beam is linearly polarized in the y direction.

(b) Show that a half-wave plate will convert right-handed circularly polarized light into left-handed circularly polarized light, and vice versa, regardless of the azimuth angle of the plate.

(c) E7 is a nematic liquid crystal with $n_o = 1.52$ and $n_e = 1.75$ at $\lambda = 577$ nm. Find the half-wave plate thickness at this wavelength, assuming the plate is made in such a way that the surfaces are parallel to the directors (i.e., a-plate).

1.20 A quarter-wave plate has a phase retardation of $\Gamma = \pi/2$. Assume that the plate is oriented in a direction with azimuth angle ψ.

(a) Find the polarization state of the transmitted beam, assuming that the incident beam is linearly polarized in the y direction.

(b) If the polarization state resulting from (a) is represented by a complex number on the complex plane, show that the locus of these points as ψ varies from 0 to $\pi/2$ is a branch of a hyperbola. Obtain the equation of the hyperbola.

(c) ZLI-1646 is a nematic liquid crystal with $n_o = 1.478$ and $n_e = 1.558$ at $\lambda = 589$ nm. Find the thickness of an a-plate at this wavelength.

1.21 A wave plate is characterized by its phase retardation Γ and azimuth angle ψ.

(a) Find the polarization state of the emerging beam, assuming that the incident beam is polarized in the y direction.

(b) Use a complex number to represent the resulting polarization state obtained in (a).

(c) The polarization state of the emerging beam is represented by a point in the complex plane. Show that the transformed polarization state can be anywhere in the complex plane, provided Γ can be varied from 0 to 2π and ψ can be varied from 0 to $\pi/2$. Physically, this means that any polarization state can be produced from linearly polarized light, provided a proper wave plate is available.

(d) Show that the locus of these points in the complex plane obtained by rotating a wave plate from $\psi = 0$ to $\psi = \pi/2$ is a hyperbola. Derive the equation of this hyperbola.

(e) Show that the Jones matrix W of a wave plate is unitary; that is,

$$W^{\dagger}W = 1$$

where the dagger indicates Hermitian conjugation.

(f) Let $\mathbf{V}_1'$ and $\mathbf{V}_2'$ be the transformed Jones vectors from $\mathbf{V}_1$ and $\mathbf{V}_2$, respectively. Show that if $\mathbf{V}_1$ and $\mathbf{V}_2$ are orthogonal, so are $\mathbf{V}_1'$ and $\mathbf{V}_2'$.

1.22 An ideal polarizer can be considered as a projection operator that acts on the incident polarization state and projects the polarization vector along the transmission axis of the polarizer.

(a) If we neglect the absolute phase factor in Equation (1.9-12), show that

$$P_o^2 = P_0 \quad \text{and} \quad P^2 = P$$

Operators satisfying these conditions are called projection operators in linear algebra.

(b) Show that if $\mathbf{E}_1$ is the amplitude of the electric field, the amplitude of the beam after it passes through the polarizer is given by

$$\mathbf{p}(\mathbf{p} \cdot \mathbf{E}_1)$$

where $\mathbf{p}$ is the unit vector along the transmission axis of the polarizer.

(c) If the incident beam is vertically polarized (i.e., $\mathbf{E}_1 = \hat{\mathbf{y}}E_0$), the polarizer transmission axis is in the x direction (i.e., $\mathbf{p} = \hat{\mathbf{x}}$). The transmitted beam has zero amplitude, since $\hat{\mathbf{x}} \cdot \hat{\mathbf{y}} = 0$. However, if a second polarizer is placed in front of the first polarizer and is oriented at 45° with respect to it, the transmitted amplitude is not zero. Find this amplitude.

(d) Consider a series of N polarizers with the first one oriented at $\psi_1 = \pi/2N$, the second one at $\psi_2 = 2(\pi/2N)$, the third one at $\psi_3 = 3(\pi/2N)$, . . . , and the Nth one at $N(\pi/2N)$. Let the incident beam be horizontally polarized. Show that the transmitted beam is vertically polarized with an amplitude of

$$[\cos(\pi/2N)]^N$$

Evaluate the amplitude for $N = 1, 2, 3, \ldots, 10$. Show that in the limit of $N \to \infty$, the amplitude becomes one. In other words, a series of polarizers oriented like a fan can rotate the polarization of the light without attenuation.

1.23 In solar physics, the distribution of hydrogen in the solar corona is measured by photographing at the wavelength of the H_α line ($\lambda = 6563$ Å). To enhance the signal-to-noise ratio, a filter of extremely narrow bandwidth (~ 1 Å) is required. The polarization filter devised by Lyot and Öhman consists of a set of birefringent plates separated by parallel polarizers. The plate thicknesses are in geometric progression, that is, d, $2d$, $4d$, $8d$, ... All the plates are oriented at an azimuth angle of 45° [4–7].

(a) Show that if n_o and n_e are the refractive indices of the plates, the transmittance of the whole stack of N plates is given by

$$T = \tfrac{1}{2}\cos^2 x \cos^2 2x \cos^2 4x \cdots \cos^2 2^{N-1}x$$

with

$$x = \frac{\pi d(n_e - n_o)}{\lambda} = \frac{\pi d(n_e - n_o)\nu}{c}$$

(b) Show that the transmission can be written

$$T = \frac{1}{2}\left(\frac{\sin 2^N x}{2^N \sin x}\right)^2$$

(c) Show that the transmission bandwidth (FWHM, full width at half-maximum) of the whole system is governed by that of the bands of the thickest plate, that is,

$$\Delta\nu_{1/2} \sim \frac{c}{2^N d(n_e - n_o)}$$

and the free spectral range $\Delta\nu$ is governed by that of the bands of the thinnest plate, that is,

$$\Delta\nu \sim \frac{c}{d(n_e - n_o)}$$

The finesse F of the system, defined as $\Delta\nu/\Delta\nu_{1/2}$, is then

$$F \sim 2^N$$

(d) Design a filter with a bandwidth of 1 Å at the H_α line, using quartz as the birefringent material. Assume that $n_o = 1.5416$ and $n_e = 1.5506$ at $\lambda = 6563$ Å. Find the required thickness of the thickest plate.

(e) Show that according to (b) the bandwidth (FWHM) is given by

$$\Delta\nu_{1/2} = 0.886\frac{c}{2^N d(n_e - n_o)}$$

1.24 In the examples in Section 1.9, we assumed that the c axis of the a-plate is oriented at 45° to the transmission axis of the polarizer. Show that if the orientation is at a general angle θ, the transmission can be written

$$T = \tfrac{1}{2}\sin^2 2\theta \sin^2(\Gamma/2)$$

which reduces to Equation (1.9-41) when $\theta = 45°$.

1.25 It can be shown that a general birefringent network (e.g., a sequence of wave plates) is equivalent to a wave plate and a polarization rotator. A polarization rotator will rotate the polarization ellipse of an input beam by an angle without changing the shape of the ellipse. The Jones matrix of a polarization rotator can be written

$$A(\rho) = \begin{bmatrix} \cos\rho & -\sin\rho \\ \sin\rho & \cos\rho \end{bmatrix}$$

where ρ is the angle of rotation. We note that a polarization rotation of ρ is equivalent to a coordinate rotation of $-\rho$.

(a) Find the Jones matrix of a birefringent wave plate with phase retardation Γ and azimuth angle ψ followed by a polarization rotator. Show that the Jones matrix is unitary.

(b) The Jones matrix of a general birefringent network is unitary. Based on the result in (a), an arbitrary birefringent network is equivalent to a phase retardation wave plate followed by a polarization rotator. Let the Jones matrix of the general birefringent network be written

$$M = \begin{bmatrix} a + ib & c + id \\ -c + id & a - ib \end{bmatrix}$$

where a, b, c, and d are real. Show that the phase retardation Γ, azimuth angle ψ, and rotation angle ρ of the equivalent network (a wave plate and a rotator) can be written

$$\cos^2(\Gamma/2) = a^2 + c^2, \quad \sin^2(\Gamma/2) = b^2 + d^2$$

$$\tan(\rho + 2\psi) = d/b, \quad \tan\rho = -c/a$$

Based on the above results in (a) and (b), an input beam of linearly polarized light with electric field vector oriented along the principal axes of the equivalent wave plate (ψ or $\psi + \pi/2$) will be transformed into an output beam of linearly polarized light.

1.26 Given a general birefringent network, there exist two linearly polarized input states such that the corresponding output polarization states are linear. These two input polarization states are mutually orthogonal. So are the two output polarization states.

(a) A beam of linearly polarized light is transformed into a beam of elliptically polarized light by using a wave plate (uniaxial) of phase retardation Γ. Let θ be the angle between the polarization vector and the x axis, and let the c axis (slow axis) of the plate be parallel to the x axis. Show that as θ varies from 0 to π, the output polarization state traces out a great circle on the Poincaré sphere. This circle can be obtained by rotating the equator around the S_1 axis by an angle Γ. Show that as Γ varies from 0 to 2π, the output polarization state traces out a circle on the Poincaré sphere. This circle can be obtained by rotating the initial polarization state on the equator around the S_1 axis. For an arbitrary phase retardation Γ, the output polarization state can be obtained by rotating the initial polarization state on the equator around the S_1 axis by an angle Γ.

(b) If a polarization rotator (ρ) is placed after the wave plate in Problem 1.25(a), show that as θ varies from 0 to π, the output polarization state also traces out a great circle on the Poincaré sphere. This great circle is obtained by rotating the great circle in (a) around the polar axis by an angle 2ρ. Based on the result, any input polarization state of a beam of monochromatic light can be transformed into any output polarization state by using the combination of a single wave plate and a polarization rotator.

(c) Using the equivalent circuit in Problem 1.25, show that given a general birefringent network, there exist two linearly polarized input states such that the output polarization states are linear. These two polarization states are mutually orthogonal. Find the polarization states of the input and output beams.

The existence of these linearly polarized states can be proved geometrically as follows. Given an input beam of linearly polarized light, an output beam of elliptically polarized light is obtained. By varying the angle of polarization of the input linearly polarized state from 0 to π, the output polarization state will trace out a great circle on the Poincaré sphere. This great circle either is the equator itself or will intersect with the equator at two points. These two points correspond to output states of linear polarization.

1.27

(a) Derive the expression for the eigenpolarization of the electric field vector in Equations (1.7-11) and (1.7-13).

(b) Use the relation $\mathbf{D} = \varepsilon\mathbf{E}$ and obtain an expression for the corresponding eigenpolarization of the displacement vector $\mathbf{D}$.

(c) Let n_1 and n_2 be the solution of the Fresnel equation (1.7-12), and $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{D}_1$, and $\mathbf{D}_2$ the corresponding eigen-field-vectors. Evaluate $\mathbf{E}_1 \cdot \mathbf{E}_2$ and $\mathbf{D}_1 \cdot \mathbf{D}_2$ and show that $\mathbf{D}_1$ and $\mathbf{D}_2$ are always mutually orthogonal, whereas $\mathbf{E}_1$ and $\mathbf{E}_2$ are only mutually orthogonal in a uniaxial or an isotropic medium.

(d) Show that $\mathbf{E}_1 \cdot \mathbf{D}_2 = 0$ and $\mathbf{E}_2 \cdot \mathbf{D}_1 = 0$.

1.28

(a) Derive the Fresnel equation (1.7-12) directly from Equation (1.7-10).

(b) Show that the Fresnel equation (1.7-12) is a quadratic equation in n^2; that is,

$$An^4 + Bn^2 + C = 0$$

and obtain expressions for A, B, and C.

(c) Show that $B^2 - 4AC > 0$ for the case of pure dielectrics with real ε_x, ε_y, ε_z.

(d) Derive Equation (1.7-16) from Equation (1.7-10) for the case of uniaxial crystals.

(e) Show that in an isotropic medium, Equation (1.7-10) reduces to

$$k^2 - \frac{\varepsilon}{\varepsilon_0}\left(\frac{\omega}{c}\right)^2 = 0$$

1.29 Show that the group velocity for a wave packet propagation in an anisotropic medium also represents the transport of energy, that is, $\mathbf{v}_g = \mathbf{v}_e$. Show that the equality $\mathbf{v}_g = \mathbf{v}_e$ also holds for complex field amplitudes $\mathbf{E}$ and $\mathbf{H}$.

1.30

(a) Derive an expression for the group velocity of the extraordinary wave in a uniaxial crystal as a function of the polar angle θ of the propagation vector.

(b) Derive an expression for the angle α between the phase velocity and the group velocity. This angle is also the angle between the field vectors $\mathbf{E}$ and $\mathbf{D}$.

(c) Show that $\alpha = 0$ when $\theta = 0, \pi/2$. Find the angle θ at which α is maximized, and obtain an expression for α_{max}. Calculate this angle α_{max} for ZLI-1646 with $n_o = 1.478$ and $n_e = 1.558$.

(d) Show that for $n_o \cong n_e$ the maximum angular separation α_{max} occurs at $\theta \cong 45°$, and show that α_{max} is proportional to $|n_o - n_e|$.

1.31

(a) Show that the normal surface Equation (1.7-10) can also be written

$$\frac{k_x^2}{k^2 - \omega^2\mu\varepsilon_x} + \frac{k_y^2}{k^2 - \omega^2\mu\varepsilon_y} + \frac{k_z^2}{k^2 - \omega^2\mu\varepsilon_z} = 1$$

where $k^2 = k_x^2 + k_y^2 + k_z^2$.

(b) Find the normal vector to the normal surface by taking the gradient of the above equation and show that the normal vector is perpendicular to the eigenvectors in Equation (1.7-11). This proves that the eigenvectors for $\mathbf{E}$ are tangent to the normal surface.

1.32 The phenomenon of double refraction in an anisotropic crystal may be utilized to produce polarized light. Consider a light beam incident on a plane boundary from the inside of a calcite crystal ($n_o = 1.658$, $n_e = 1.486$). Suppose that the c axis of the crystal is normal to the plane of incidence.

(a) Find the range of the internal angle of incidence such that the ordinary wave is totally reflected. The transmitted wave is thus completely polarized.

(b) Use the basic principle described in (a) to design a calcite Glan prism shown in Figure P1.32.

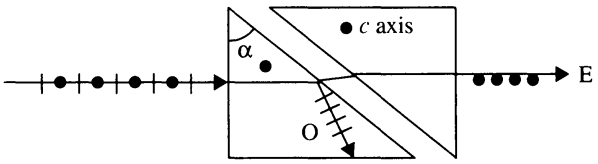


Figure P1.32

Find the range of the apex angle α.

1.33 Dichroic polarizers are materials whose absorption properties and reflection properties are strongly dependent on the direction of vibration of the electric field. If the two absorption coefficients are very different, a thin sheet of the material will be sufficient to transform unpolarized light into a nearly linearly polarized light.

(a) Let α_1 and α_2 be the two absorption coefficients corresponding to two independent polarizations. Derive an expression for the ratio of the two transmitted components as a function of the thickness of the medium.

(b) Show that, strictly speaking, the normal modes of propagation are no longer linearly polarized in the presence of absorption.

1.34 The wave equation (1.7-8) can be written

$$\mathbf{s} \times \left(\mathbf{s} \times \frac{\varepsilon_0}{\varepsilon}\mathbf{D}\right) = -\frac{1}{n^2}\mathbf{D}$$

where $\mathbf{s}$ is a unit vector in the direction of propagation. Let $\mathbf{D}_1$, $\mathbf{D}_2$ be the normalized eigenvectors with eigenvalues $1/n_1^2$, $1/n_2^2$, respectively. We assume that $\varepsilon_0/\varepsilon$ is a Hermitian tensor.

(a) Show that

$$\left(\frac{\varepsilon_0}{\varepsilon}\right)_{11} \equiv \mathbf{D}_1^* \cdot \frac{\varepsilon_0}{\varepsilon}\mathbf{D}_1 \equiv \frac{1}{n_1^2},$$

$$\left(\frac{\varepsilon_0}{\varepsilon}\right)_{22} \equiv \mathbf{D}_2^* \cdot \frac{\varepsilon_0}{\varepsilon}\mathbf{D}_2 \equiv \frac{1}{n_2^2},$$

$$\left(\frac{\varepsilon_0}{\varepsilon}\right)_{12} \equiv \mathbf{D}_1^* \cdot \frac{\varepsilon_0}{\varepsilon}\mathbf{D}_2 = 0$$

(b) Show that $\mathbf{D}_1^* \cdot \mathbf{D}_2 = 0$.

1.35

(a) Show that the power flow in the direction of propagation is given by

$$\mathbf{S} \cdot \mathbf{s} = \frac{1}{2}\frac{c}{\varepsilon_0}\left(\frac{1}{n_1^3}|\mathbf{D}_1|^2 + \frac{1}{n_2^3}|\mathbf{D}_2|^2\right)$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ are the amplitudes of the displacement fields of the eigenmodes and n_1 and n_2 are the corresponding refractive indices.

(b) Show that

$$\mathbf{S} \cdot \mathbf{s} = \frac{1}{2\mu c}(n_1|\mathbf{A}_1|^2 + n_2|\mathbf{A}_2|^2)$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are the transverse parts of the electric field vectors.

(c) Show that the total power flow along the direction of propagation is a constant of integration; that is,

$$\frac{d}{d\xi}(n_1|\mathbf{A}_1|^2 + n_2|\mathbf{A}_2|^2) = 0$$

where ξ is the distance along the direction of propagation $\mathbf{s}$ (i.e., $\xi = \mathbf{s} \cdot \mathbf{r}$).

1.36 Consider a beam of circularly polarized light with a finite transverse dimension.

(a) Let the transverse dimension be much greater than the wavelength λ. Show that the electric field and magnetic field vector can be written approximately

$$E(x, y, z, t) \approx E_0(x, y)(\hat{\mathbf{x}} - i\hat{\mathbf{y}}) - \frac{i}{k}\left(\frac{\partial E_0}{\partial x} - i\frac{\partial E_0}{\partial y}\right)\hat{\mathbf{z}}e^{i(\omega t - kz)}$$

$$H(x, y, z, t) \approx i\frac{k}{\omega\mu}E(x, y, z, t)$$

(b) Using the angular momentum defined as $\mathbf{L} = \mathbf{r} \times \mathbf{P}$, where $\mathbf{P}$ is the linear momentum defined in Problem 1.5, calculate the time-averaged component of the angular momentum along the direction of propagation ($+z$). Show that this component of the angular momentum is $\hbar$ provided the energy of the electromagnetic wave is normalized to $\hbar\omega$.

(c) Show that the transverse components of the angular momentum vanish.

1.37 According to Problem 1.6, the work done per unit volume by applying an electric field in a dielectric medium is $W = \int \mathbf{E} \cdot d\mathbf{P}$, where $\mathbf{P}$ is the polarization. In a linear medium $P_i = \varepsilon_0\chi_{ij}E_j$. Carry out the integration from (0, 0) to (E_{10}, E_{20}) in a two-dimensional case by using two different paths.

(a) Path A: First from (0, 0) to $(E_{10}, 0)$, and then from $(E_{10}, 0)$ to (E_{10}, E_{20}). Show that the integration yields the following result:

$$W = \tfrac{1}{2}\varepsilon_0\chi_{11}E_{10}^2 + \tfrac{1}{2}\varepsilon_0\chi_{22}E_{20}^2 + \varepsilon_0\chi_{12}E_{10}E_{20}$$

(b) Path B: First from (0, 0) to $(0, E_{20})$, and then from $(0, E_{20})$ to (E_{10}, E_{20}). Show that the integration yields the following result:

$$W = \tfrac{1}{2}\varepsilon_0\chi_{11}E_{10}^2 + \tfrac{1}{2}\varepsilon_0\chi_{22}E_{20}^2 + \varepsilon_0\chi_{21}E_{10}E_{20}$$

In a lossless medium, the integration should be independent of the path. This leads to $\chi_{12} = \chi_{21}$. A similar analysis in the three-dimensional case will yield $\chi_{ij} = \chi_{ji}$.

REFERENCES

1. Jones, R. C., New calculus for the treatment of optical systems. *J. Opt. Soc. Am.* **31**:488 (1941).
2. Yariv, A., and P. Yeh, *Optical Waves in Crystals*. Wiley, New York, 1984.
3. Yeh, P., *Optical Waves in Layered Media*. Wiley, New York, 1988.
4. Lyot, B., Optical apparatus with wide field using interference of polarized light. *C. R. Acad. Sci. (Paris)* **197**:1593 (1933).
5. Lyot, B., Filter monochromatique polarisant et ses applications en physique solaire. *Ann. Astrophys.* **7**:31 (1944).
6. Öhman, Y., A new monochromator. *Nature* **41**:157 (1938).
7. Öhman, Y., On some new birefringent filter for solar research. *Ark. Astron.* **2**:165 (1958).

CHAPTER 2

RAYS AND OPTICAL BEAMS

2.0 INTRODUCTION

In the last chapter we described plane waves in homogeneous media. In this chapter we consider the propagation of rays and beams through a variety of media that are often employed in optical electronic systems. The media include homogeneous and inhomogeneous materials, thin lenses, gradient index (GRIN) lenses, dielectric interfaces, and mirrors. In geometrical optics, the propagation of optical waves can be described approximately by using the concept of rays. This is valid provided the beam diameter is much larger than the wavelength, and the diffraction can be neglected. The rays travel in straight lines in homogeneous media and obey Fermat's principle of least time in inhomogeneous media. A good understanding of the propagation of rays makes it possible to trace their trajectories when they are passing through various optical media. We find that the passage of rays through these optical elements can be described by simple 2×2 matrices. In this chapter, we also describe the propagation of Gaussian beams in various optical media, as well as the integral representation of the propagation of a general paraxial beam of light. It will be shown that simple ray matrices can be employed to describe the propagation of spherical waves and Gaussian beams, such as those characteristic of the output of lasers. Furthermore, the ray matrices are also useful for the propagation of a general paraxial beam through various optical media.

2.1 RAY MATRICES

Referring to Figure 2.1, we consider the passing of a paraxial ray through a thin lens of focal length f. In most optical systems that involve lenses made of homogeneous optical materials with spherical surfaces, the axis of symmetry (z axis) is often called the optical axis. Paraxial rays are defined as those rays whose angular deviation from the optical axis is small enough that the sine and tangent of the angle can be approximated by the angle itself. For such a system with cylindrical symmetry, we consider a class of rays whose trajectories can be conveniently described by $r(z)$, where r is the distance measured from the optical axis. These are known as meridional rays whose trajectories are confined in a plane through the optical axis.

For a thin lens, the output ray and the input ray are related by the following simple equations:

$$r_{\text{out}} = r_{\text{in}} \tag{2.1-1}$$

$$r'_{\text{out}} = r'_{\text{in}} - \frac{r_{\text{in}}}{f} \tag{2.1-2}$$

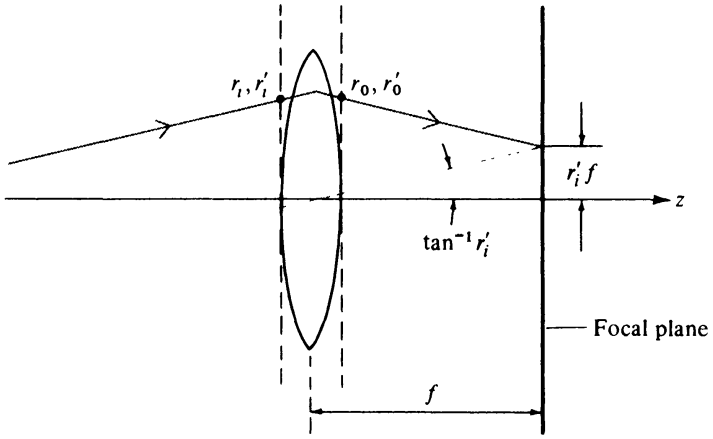


Figure 2.1 Deflection of a ray by a thin lens.

where r_{in} and r_{out} are the positions of the ray measured from the optical axis, and r'_{in} and r'_{out} are the slopes of the ray relative to the optical axis. In the paraxial approximation, the slopes are actually the ray angles measured relative to the optical axis. We note that the ray position is unchanged, reflecting the definition of a thin lens. The slope of the output ray is changed due to the presence of the thin lens. Equation (2.1-2) can easily be derived by considering the focusing nature of the thin lens. In other words, parallel rays intersect at the focal plane after passing through a thin lens.

Equations (2.1-1) and (2.1-2) can be represented conveniently by a simple 2×2 matrix equation

$$\begin{bmatrix} r_{out} \\ r'_{out} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\dfrac{1}{f} & 1 \end{bmatrix} \begin{bmatrix} r_{in} \\ r'_{in} \end{bmatrix} \tag{2.1-3}$$

where $f > 0$ for a convergent lens and is negative for a divergent one. The ray matrices for a number of optical elements are shown in Table 2.1.

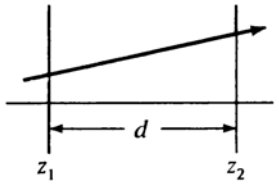
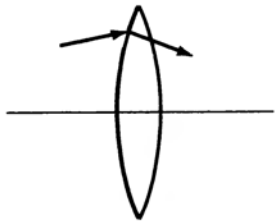
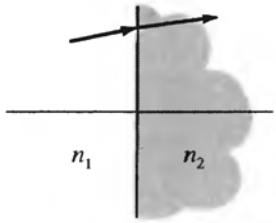
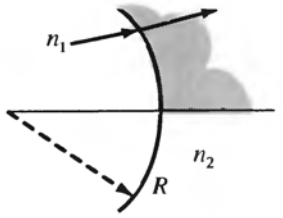
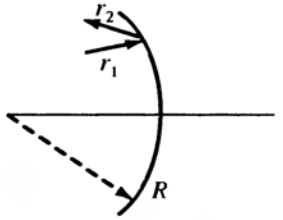
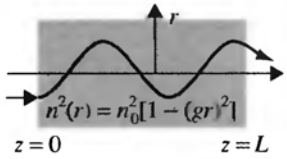
According to Table 2.1, propagation through a homogeneous medium with length d can be described as

$$\begin{bmatrix} r_{out} \\ r'_{out} \end{bmatrix} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_{in} \\ r'_{in} \end{bmatrix} \tag{2.1-4}$$

In the matrix approach, each ray is represented by a ray vector, while the optical elements are represented by ray matrices. The ray vector is a two-element column vector that consists of the ray distance from the axis $r(z)$ and the ray angle $r'(z)$, at any position z on the axis. In an optical system consisting of a series of optical elements as shown in Table 2.1, the output ray vector can easily be obtained by multiplying the ray matrices in sequence with the input ray vector.

To illustrate this, we consider the propagation of a ray through a periodic lens system, which consists of lenses of focal lengths f_1 and f_2 separated by a distance d as shown in Figure 2.2. Such an optical system will be shown later in this book to be formally equivalent to the problem of Gaussian beam propagation inside an optical resonator with mirrors of radii of curvature $R_1 = 2f_1$ and $R_2 = 2f_2$ that are separated by a distance d.

TABLE 2.1 Ray Matrices for Some Optical Elements and Media.

(1) Straight Section: Length d		$\begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}$
(2) Thin Lens: Focal length f ($f > 0$, converging; $f < 0$, diverging)		$\begin{bmatrix} 1 & 0 \\ \frac{-1}{f} & 1 \end{bmatrix}$
(3) Dielectric Interface: Refractive indices n_1, n_2		$\begin{bmatrix} 1 & 0 \\ 0 & \frac{n_1}{n_2} \end{bmatrix}$
(4) Spherical Dielectric Interface: Radius R		$\begin{bmatrix} 1 & 0 \\ \frac{n_2 - n_1}{n_2 R} & \frac{n_1}{n_2} \end{bmatrix}$
(5) Spherical Mirror: Radius of curvature R		$\begin{bmatrix} 1 & 0 \\ \frac{-2}{R} & 1 \end{bmatrix}$
(6) A medium with a quadratic index profile		$\begin{bmatrix} \cos(gz) & \sin(gz)/g \\ -g\sin(gz) & \cos(gz) \end{bmatrix}$

The section between planes s and $s + 1$ can be considered as a unit cell of the optical system. The matrix relating the ray vector at the output of a unit cell to that at the input is the product of the ray matrices of the elements in sequence. In other words,

$$\begin{bmatrix} r_{s+1} \\ r'_{s+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/f_1 & 1 \end{bmatrix} \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/f_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_s \\ r'_s \end{bmatrix} \tag{2.1-5}$$

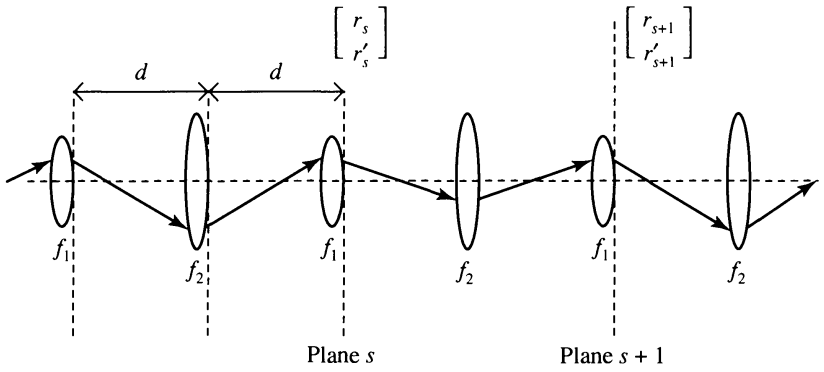


Figure 2.2 A periodic lens waveguide consisting of lenses with focal lengths f_1 and f_2.

where we note the rightmost ray matrix corresponds to the free space between plane s and the lens with focal length f_2 in the unit cell. By carrying out the matrix multiplication in Equation (2.1-5), we obtain

$$\begin{bmatrix} r_{s+1} \\ r'_{s+1} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} r_s \\ r'_s \end{bmatrix} \tag{2.1-6}$$

or equivalently,

$$\begin{aligned} r_{s+1} &= Ar_s + Br'_s \\ r'_{s+1} &= Cr_s + Dr'_s \end{aligned} \tag{2.1-7}$$

where A, B, C, and D are the matrix elements given by

$$\begin{aligned} A &= 1 - \frac{d}{f_2} \\ B &= d\left(2 - \frac{d}{f_2}\right) \\ C &= -\left[\frac{1}{f_1} + \frac{1}{f_2}\left(1 - \frac{d}{f_1}\right)\right] \\ D &= -\left[\frac{d}{f_1} - \left(1 - \frac{d}{f_1}\right)\left(1 - \frac{d}{f_2}\right)\right] \end{aligned} \tag{2.1-8}$$

It is important to note that the ray matrix for the unit cell is a unimodular matrix. In other words,

$$AD - BC = 1 \tag{2.1-9}$$

The result (Equation (2.1-6)) can now be applied to the propagation through N unit cells, say, between planes 0 and N. We obtain

$$\begin{bmatrix} r_N \\ r'_N \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^N \begin{bmatrix} r_0 \\ r'_0 \end{bmatrix} \tag{2.1-10}$$

The Nth power of a unimodular matrix can be simplified by the following matrix identity (Chebyshev's identity, see Problem 2.18):

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^N = \begin{bmatrix} AU_{N-1} - U_{N-2} & BU_{N-1} \\ CU_{N-1} & DU_{N-1} - U_{N-2} \end{bmatrix} \tag{2.1-11}$$

where

$$U_N = \frac{\sin(N+1)\theta}{\sin\theta} \tag{2.1-12}$$

with

$$\theta = \cos^{-1}\left(\frac{A+D}{2}\right) \tag{2.1-13}$$

If a ray is launched at plane 0 with finite r_0 and r_0', the ray will remain confined by the lens system along the optical axis provided the ray parameters r_N and r_N' remain finite as N approaches infinity. By examining Equations (2.1-10)–(2.1-13), we find that confined propagation of a ray requires a real θ, or equivalently,

$$\left|\frac{A+D}{2}\right| \le 1 \tag{2.1-14}$$

In terms of the system parameters (Equation (2.1-8)), the condition for confined propagation becomes

$$-1 \le 1 - \frac{d}{f_1} - \frac{d}{f_2} + \frac{d^2}{2f_1 f_2} \le 1 \tag{2.1-15}$$

or equivalently,

$$0 \le \left(1 - \frac{d}{2f_1}\right)\left(1 - \frac{d}{2f_2}\right) \le 1 \tag{2.1-16}$$

If, on the other hand, the confinement condition Equation (2.1-14) is violated, then the parameter θ becomes a purely imaginary number, $\theta = iq$, where q is a real number. In this case, U_N becomes

$$U_N = \frac{\sinh(N+1)q}{\sinh q} \tag{2.1-17}$$

and thus the ray parameters r_N and r_N' will grow exponentially as a function of N. This leads to unconfined propagation.

Identical Lens Waveguide—A Special Case

Consider the special case when $f_1 = f_2 = f$ in Figure 2.2. This is the simplest lens waveguide. The confinement condition Equation (2.1-16) becomes

$$\left(1 - \frac{d}{2f}\right)^2 \le 1 \tag{2.1-18}$$

Since d is always positive, the condition is satisfied only when $f > 0$. In other words, in the simplest lens waveguide, confinement is possible only when the lenses are convergent lenses. The confinement condition can further be simplified as

$$0 \le d \le 4f \tag{2.1-19}$$

The same problem can also be solved directly. In this case, the ray parameters from one plane to the next one are given by

$$\begin{bmatrix} r_{s+1} \\ r'_{s+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_s \\ r'_s \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} r_s \\ r'_s \end{bmatrix} \tag{2.1-20}$$

By carrying out the matrix multiplication in Equation (2.1-20), we obtain

$$\begin{aligned} A &= 1 \\ B &= d \\ C &= -1/f \\ D &= 1 - d/f \end{aligned} \tag{2.1-21}$$

The confinement condition is given by, according to Equation (2.1-14)

$$-1 \le 1 - \frac{d}{2f} \le 1 \tag{2.1-22}$$

which is equivalent to Equations (2.1-18) and (2.1-19).

The beam radius (or ray location) at the mth lens can be written, according to Equations (2.1-10)–(2.1-12)

$$r_m = r_{\max} \sin(m\theta + \alpha) \tag{2.1-23}$$

with

$$\cos \theta = 1 - d/2f \tag{2.1-24}$$

Using Equations (2.1-10)–(2.1-12) and the explicit expressions of A, B, C, and D, we obtain

$$(r_{\max})^2 = \frac{4f}{4f - d}(r_0^2 + dr_0 r'_0 + dfr'^2_0) \tag{2.1-25}$$

$$\tan \alpha = \frac{\sqrt{4f/d - 1}}{(1 + 2fr'_0/r_0)} \tag{2.1-26}$$

where m corresponds to the plane immediately to the right of the mth lens. The derivation of the last two equations is left as an exercise.

The stability criteria can be demonstrated experimentally by tracing the behavior of a laser beam as it propagates down a sequence of lenses spaced uniformly. One can easily note the rapid escape of the beam once condition (2.1-19) or (2.1-22) is violated.

Rays in Optical Resonators

Consider the bouncing of rays between two curved mirrors that form an optical resonator. Since the reflection at a mirror with a radius of curvature R is equivalent, except for the folding of the path, to passage through a lens with a focal length of $f = R/2$, we can use the ray matrix formulation to describe the propagation of a ray inside the resonator, which consists of two mirrors of radii R_1 and R_2. Figure 2.3 shows such an optical resonator and its equivalent optical lens system.

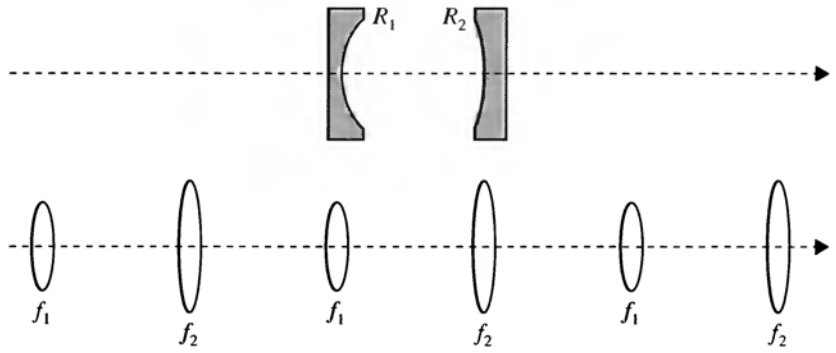


Figure 2.3 An optical resonator and its equivalent optical lens system. R_1 and R_2 are radii of curvature of the mirrors: $f_1 = R_1/2$ and $f_2 = R_2/2$.

For an optical resonator with two concave mirrors, the confinement condition becomes

$$0 \le \left(1 - \frac{d}{R_1}\right)\left(1 - \frac{d}{R_2}\right) \le 1 \tag{2.1-27}$$

where d is the separation between the mirrors, and R_1 and R_2 are the radii of curvature of the mirrors.

2.2 SKEW RAYS AND REENTRANT RAYS

Skew Rays

So far, we have only considered a particular class of rays (the meridional rays) in which each ray trajectory is confined in a plane that passes through the optical axis. In general, a ray can propagate with a three-dimensional trajectory along the optical axis (e.g., a helical trajectory). These are the so-called skew rays. For these rays, the x and y coordinates are independent variables. A complete description of the ray requires $x(z)$, $x'(z)$, $y(z)$, and $y'(z)$. In the lens waveguide described in the previous section, the propagation of the rays (from plane 0 to plane N) can be written

$$\begin{bmatrix} x_N \\ x'_N \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^N \begin{bmatrix} x_0 \\ x'_0 \end{bmatrix} \tag{2.2-1}$$

$$\begin{bmatrix} y_N \\ y'_N \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^N \begin{bmatrix} y_0 \\ y'_0 \end{bmatrix} \tag{2.2-2}$$

We note that both components (x and y) of the ray parameters obey the same ray matrix for propagation in the lens system. Using Equations (2.1-11)–(2.1-13), the ray position at the Nth plane can be written

$$x_N = x_{\max} \sin(N\theta + \alpha_x) \tag{2.2-3}$$

$$y_N = y_{\max} \sin(N\theta + \alpha_y) \tag{2.2-4}$$

where N is an integer, and $x_{\max}$, $y_{\max}$, α_x, and α_y are constants. These constants depend on the initial ray parameters at the 0th plane. In general, the two phases, α_x and α_y are different.

According to Equations (2.2-3) and (2.2-4), the locus of the points (x_N, y_N) lies on an ellipse for confined propagation when θ is a real number. These are the skew rays. The ray coordinates are confined in a rectangle of $2x_{\max}$ by $2y_{\max}$. In the special case when $\alpha_x = \alpha_y$, the ray trajectory becomes confined in a meridional plane through the optical axis. This corresponds to a meridional ray. We note that θ is a system parameter that is independent of the ray coordinates.

Reentrant Rays

Based on Equations (2.2-3) and (2.2-4), the ray coordinates in a periodic lens waveguide are sinusoidal functions of $N\theta$. Under the appropriate conditions (system parameters), the ray may reproduce itself after certain periods of propagation. If the system parameters are such that $\theta/2\pi$ is a rational number, then an integer N can be found such that

$$N\theta = 2M\pi \tag{2.2-5}$$

where M is also an integer. When this happens, the ray matrix after N periods becomes, according to Equations (2.1-11) and (2.1-12),

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^N = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and thus the ray retraces its trajectory after N periods.

EXAMPLE: REENTRANT RAYS IN SYMMETRIC CONFOCAL RESONATORS

Consider an optical resonator made of two identical concave mirrors of radius of curvature R. If the separation d equals the radius of curvature R ($d = R$), then these two mirrors share a common focal point. Such resonators are called confocal resonators. For such resonators, $f_1 = f_2 = R/2$ in the equivalent lens waveguide, the system parameter θ is, according to Equation (2.1-8),

$$\theta = \cos^{-1}\left(\frac{A+D}{2}\right) = \cos^{-1}(-1) = \pi$$

The rays inside the resonator will thus repeat after two periods in the equivalent lens waveguide. Since each period in the lens waveguide is equivalent to a round-trip in the optical resonator, the ray trajectory will repeat after two round-trips in the resonator. It is important to note that the periodic retracing of the ray trajectory is independent of the initial launching condition (r_0, r_0').

2.3 RAYS IN LENSLIKE MEDIA

We consider in this section the propagation of a ray in an optically inhomogeneous medium, particularly a lenslike medium. In the realm of geometrical optics, a light ray travels between two points along a trajectory such that the time taken is the least. This is known as *Fermat's principle of least time*. In terms of calculus of variation, this is written

$$\delta \int_{P_1}^{P_2} \frac{n}{c}\, ds = 0 \tag{2.3-1}$$

where P_1 and P_2 are the two points, c is speed of light in vacuum, n is the index of refraction as a function of position, and ds is a differential element along the ray path. In other words, the optical length $\int n\,ds$ of the path traversed by a ray between two points is stationary, that is, either a maximum or a minimum.

We represent the ray trajectory in parametric form as

$$x = x(t), \quad y = y(t), \quad z = z(t) \tag{2.3-2}$$

where t is an arbitrary parameter. Equation (2.3-1) can then be written

$$\delta \int_{t_1}^{t_2} n(x, y, z)\sqrt{x'^2 + y'^2 + z'^2}\, dt = 0 \tag{2.3-3}$$

where x', y', and z' are derivatives with respect to the parameter t, and we have ignored the constant c. If we define a functional

$$F(x, x', y, y', z, z', t) = n(x, y, z)\sqrt{x'^2 + y'^2 + z'^2} \tag{2.3-4}$$

then the Euler–Lagrange equations following the condition of Equation (2.3-1) can be written

$$\frac{\partial F}{\partial x} = \frac{d}{dt}\left(\frac{\partial F}{\partial x'}\right), \quad \frac{\partial F}{\partial y} = \frac{d}{dt}\left(\frac{\partial F}{\partial y'}\right), \quad \frac{\partial F}{\partial z} = \frac{d}{dt}\left(\frac{\partial F}{\partial z'}\right) \tag{2.3-5}$$

In other words, the ray trajectory with the stationary optical length must satisfy the Euler–Lagrange equations (2.3-5). Using

$$ds = dt\sqrt{x'^2 + y'^2 + z'^2} \tag{2.3-6}$$

the above Euler–Lagrange equations can be written

$$\frac{\partial n}{\partial x} = \frac{d}{ds}\left(n\frac{dx}{ds}\right), \quad \frac{\partial n}{\partial y} = \frac{d}{ds}\left(n\frac{dy}{ds}\right), \quad \frac{\partial n}{\partial z} = \frac{d}{ds}\left(n\frac{dz}{ds}\right) \tag{2.3-7}$$

These three equations can be written as the following vector equation:

$$\frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) = \nabla n \tag{2.3-8}$$

where we recall that s is the arc length along the ray trajectory measured from P_1 and $\mathbf{r}$ is the vector representation of (x, y, z). This is the ray equation. The left side is a measure of the bending of the ray, while the right side is the gradient of the index of refraction. Thus the ray always bends toward the high-index region as it propagates. The ray equation has successfully been employed to explain many atmospheric phenomena, including mirage and green flash. In this chapter, we will investigate the propagation of a ray in a quadratic index medium (or lenslike medium).

We now consider the propagation of a ray in a lenslike medium whose index of refraction n varies according to

$$n^2(x, y) = n_0^2[1 - g^2(x^2 + y^2)] \quad \text{with} \quad g^2(x^2 + y^2) \ll 1 \tag{2.3-9}$$

where n_0 is the refractive index at the axis of symmetry and g is a real constant characteristic of the medium. Since the phase delay of a wave propagating through a section dz of a medium with an index of refraction n is $(2\pi\, dz/\lambda)n$, it follows directly that a thin slab of the medium described by Equation (2.3-9) will act as a thin lens, introducing a phase shift proportional to r^2. The medium has a maximum index of refraction at the axis of symmetry ($r = 0$).

The index of refraction decreases in the radial direction away from the axis. This type of medium is also known as gradient index (GRIN) medium, or quadratic index medium. This equation can also be viewed as consisting of the lowest terms in the Taylor series expansion of $n(x, y)$ for a continuous optical medium with cylindrical symmetry. Using cylindrical coordinates with

$$r = \sqrt{x^2 + y^2}$$

Equation (2.3-9) can be written

$$n^2(r) = n_0^2[1 - (gr)^2] \tag{2.3-10}$$

For paraxial rays, we may replace d/ds by d/dz in Equation (2.3-8). Using Equation (2.3-9) for the refractive index, we obtain

$$\frac{d^2\mathbf{r}}{dz^2} + g^2\mathbf{r} = 0 \quad \text{with} \quad \mathbf{r} = (x, y) \tag{2.3-11}$$

This is a differential equation for the ray trajectory in a lenslike medium. For meridional rays, the above equation can be written

$$\frac{d^2r}{d^2z} + g^2r = 0 \quad \text{with} \quad r = \sqrt{x^2 + y^2} \tag{2.3-12}$$

If at the input plane $z = 0$, the ray has a position r_0 and slope r_0', the ray trajectory can be written, according to Equation (2.3-12),

$$\begin{aligned} r(z) &= r_0 \cos(gz) + r_0' \sin(gz)/g \\ r'(z) &= -r_0 g \sin(gz) + r_0' \cos(gz) \end{aligned} \tag{2.3-13}$$

The ray trajectory can also be written as a matrix equation,

$$\begin{bmatrix} r \\ r' \end{bmatrix} = \begin{bmatrix} \cos(gz) & \sin(gz)/g \\ -g\sin(gz) & \cos(gz) \end{bmatrix} \begin{bmatrix} r_0 \\ r_0' \end{bmatrix} \tag{2.3-14}$$

According to Equation (2.3-13), the ray oscillates back and forth across the optical axis, as shown in Figure 2.4. This is in agreement with the fact that rays tend to bend toward the high-index region as they propagate. In fact, we note that the ray trajectory is a periodic function of z with a period

$$\Lambda = \frac{2\pi}{g} \tag{2.3-15}$$

This period is also known as the pitch.

As a result of the periodic behavior, sections of such a lenslike medium can be used for imaging purposes. These sections are called GRIN-rod lenses, or simply rod lenses. To illustrate this, let us consider the situation when a bundle of parallel rays are incident on such a medium at $z = 0$. All the incident rays are parallel to the z axis. The paths of these rays are represented by

$$r(z) = r_0 \cos(gz) \tag{2.3-16}$$

where r_0 is the point of incidence for each ray. After traversing a distance d such that

$$gd = \pi/2 \tag{2.3-17}$$

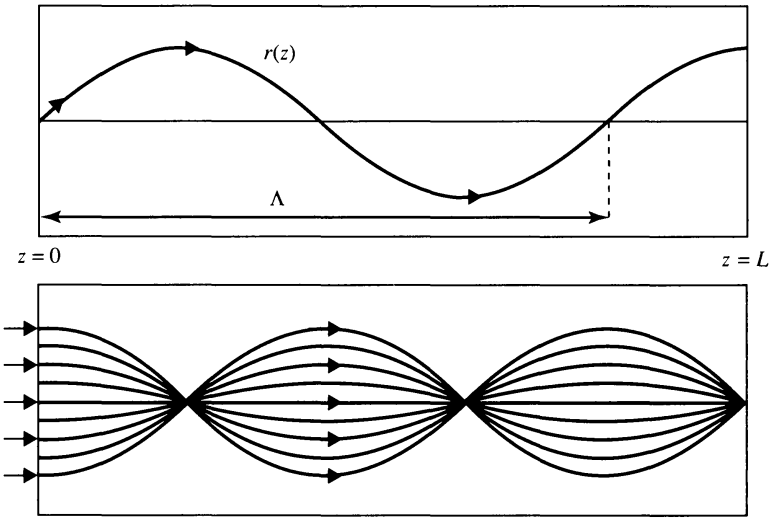


Figure 2.4 Path of rays in a medium with a quadratic index variation.

all the rays collapse into a point at the axis. If this is called a focal point, then the focal length is

$$f = \pi/2g = \Lambda/4 \tag{2.3-18}$$

where $\Lambda = 2\pi/g$ is the period. If the GRIN-rod is long enough, the bundle of rays will diverge and then converge again. In fact, the focusing will repeat itself indefinitely at an interval of π/g. Figure 2.4 illustrates the ray paths inside the medium.

If the length of the medium L is not an exact integral number of f, then the bundle of rays will converge upon emerging at $z = L$ to a common focus at a distance

$$h = \frac{1}{n_0 g \tan(gL)} \tag{2.3-19}$$

from the exit plane. The factor n_0 accounts for the refraction at the boundary, assuming the medium at $z > L$ posseses an index of $n = 1$, and a small angle of incidence. The derivation of Equation (2.3-19) is left as an exercise (Problem 2.3). If h in Equation (2.3-19) is negative, then the bundle of rays are actually divergent upon emerging at $z = L$. The diverging beam appears to originate from a point source located at $z = L + h$ (note that $h < 0$), inside the medium. In this case, the section of the medium acts like a negative lens.

There are physical situations that give rise to the quadratic index variation. These include the following:

1. Propagation of laser beams with Gaussian-like intensity profile in a slightly absorbing medium. As a result of the absorption, the temperature tends to increase more at the high-intensity region. This leads to a quadratic temperature profile. If the medium exhibits a thermo-optical coefficient of $dn/dT > 0$, then a lenslike medium is produced by the presence of the laser beam. A negative lens effect is produced if $dn/dT < 0$.
2. Propagation of laser beams with Gaussian intensity profile in Kerr media. As a result of the nonlinear response, the local index of refraction increases with the intensity, according to

$$n(x, y, z) = n_0 + n_2 I(x, y, z)$$

where n_0 is a constant and n_2 is known as the Kerr coefficient. For a Gaussian beam, the intensity has a quadratic profile near the axis of the beam. If $n_2 > 0$, then a lenslike medium is produced by the presence of the laser beam. A negative lens effect is produced if $n_2 < 0$. The Kerr effect will be discussed further in Chapter 8, as well as in Chapter 14.

3. Gradient index glass fibers produced by ion-exchange diffusion. The ion-exchange diffusions inside a hollow glass tube create an index distribution profile along the radial direction after the tube collapses into a solid fiber. These optical fibers can be employed for transmitting images in medical examinations.

The rays described in the preceding are all in meridional planes that contain the z axis. There are other types of rays that propagate out of the meridional planes. These are the so-called skew rays. Typical examples of skew rays take a helical path around the z axis (see examples in Problem 2.19).

2.4 WAVE EQUATION IN QUADRATIC INDEX MEDIA AND BEAMS

In geometrical optics, a bundle of parallel rays can be focused into a point by using a thin lens or a GRIN-rod lens. In reality, the focused spot has a finite size. Furthermore, the effect of diffraction is completely ignored in geometrical optics. In this section, we consider the propagation of optical beams by taking into account the wave nature of electromagnetic radiation. Particularly, we are interested in optical beams in which the flow of energy is predominantly along one direction. We are also interested in optical beams whose wavefronts are spherical. Later in this chapter, we will show that optical beams with spherical wavefronts can propagate in optical resonators made of spherical mirrors with minimum propagation loss.

In Chapter 1, we obtained the following equation for the propagation of an electromagnetic wave in isotropic media:

$$\nabla \times (\nabla \times \mathbf{E}) + \mu\varepsilon \frac{\partial^2}{\partial t^2} \mathbf{E} = 0 \tag{2.4-1}$$

In this section, we will find a solution of this equation in the form of an optical beam. The first term in the above equation can be expanded as

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} \tag{2.4-2}$$

Using Equation (1.1-3) and $\mathbf{D} = \varepsilon\mathbf{E}$, we obtain

$$\nabla^2 \mathbf{E} - \mu\varepsilon \frac{\partial^2}{\partial t^2} \mathbf{E} = -\nabla\left(\frac{1}{\varepsilon} \mathbf{E} \cdot \nabla\varepsilon\right) \tag{2.4-3}$$

The term on the right side of Equation (2.4-3) can be neglected provided the fractional change of ε in one wavelength is $\ll 1$. In this case, the wave equation for a harmonic field with a time dependence of $\exp(i\omega t)$ becomes

$$\nabla^2 \mathbf{E} + k^2(\mathbf{r})\mathbf{E} = 0 \tag{2.4-4}$$

where

$$k^2(\mathbf{r}) = \omega^2 \mu\varepsilon(\mathbf{r})\left(1 - i\frac{\sigma(\mathbf{r})}{\omega\varepsilon}\right) \tag{2.4-5}$$

thus allowing for a possible dependence of ε on position $\mathbf{r}$. We have also taken k as a complex number to allow for the possibility of losses ($\sigma > 0$) or gain ($\sigma < 0$) in the medium. According to the discussion in Chapter 1, the field amplitude of a plane wave $E_0 \exp[(i\omega t - kz)]$ with a complex wavenumber k will either grow or decay exponentially.

Equation (2.4-4) is known as the *Helmholtz equation*. The most widely encountered type of laser beam in optical electronics is one where the intensity distribution at planes normal to the propagation direction is Gaussian, and the wavefronts are spherical. The medium in which the laser beam will propagate often possesses a cylindrical symmetry with respect to the beam axis. In particular, we will consider a lenslike medium whose index of refraction varies according to

$$n^2(x, y) = \frac{\varepsilon(r)}{\varepsilon_0} = n_0^2[1 - g^2(x^2 + y^2)] \quad \text{with} \quad g^2(x^2 + y^2) \ll 1 \tag{2.4-6}$$

where n_0 is the refractive index at the axis of symmetry, and g is a real constant characteristic of the medium. The special case of $g = 0$ corresponds to a homogeneous medium.

We limit our derivation to the solution of the field with cylindrical symmetry, so that the Laplacian ∇^2 can be written

$$\nabla^2 = \nabla_t^2 + \frac{\partial^2}{\partial z^2} = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2} \tag{2.4-7}$$

The kind of propagation we are considering is that of a nearly plane wave in which the flow of energy is predominantly along a single (e.g., z axis) direction, so that we may limit our derivation to a single transverse field component of $\mathbf{E}$. Taking $\mathbf{E}$ as

$$\mathbf{E} = \hat{\mathbf{a}}\psi(r, z)e^{-ikz} \tag{2.4-8}$$

where $\hat{\mathbf{a}}$ is a unit vector in the xy plane representing the polarization state, we obtain from Equations (2.4-4) and (2.4-7) in a few steps

$$\nabla_t^2\psi - 2ik\psi' - k^2g^2r^2\psi = 0 \tag{2.4-9}$$

where $\psi' = \partial\psi/\partial z$, and where we assume that the variation of the field amplitude is slow enough that

$$\psi'' \ll k\psi', k^2\psi \tag{2.4-10}$$

This approximation is known as the *slowly varying amplitude* (SVA) approximation. The approximation is legitimate provided the transverse dimension of the beam is much larger than the wavelength (see Problem 2.22). Since we are looking for a beam amplitude with cylindrical symmetry, it is convenient to introduce two complex functions $P(z)$ and $q(z)$ such that ψ is in the form of

$$\psi = \exp\left[-i\left(P(z) + \frac{k}{2q(z)}r^2\right)\right] \tag{2.4-11}$$

This solution can be viewed as an optical wave with a phase of $P(z)$ and a spherical wavefront with a radius of curvature of $q(z)$. The problem at hand is to find $P(z)$ and $q(z)$ so that Equation (2.4-11) is a solution of Equation (2.4-9).

We now substitute ψ into Equation (2.4-9). Using Equation (2.4-7), we obtain

$$-\left(\frac{k}{q}\right)^2 r^2 - 2i\left(\frac{k}{q}\right) - k^2r^2\left(\frac{1}{q}\right)' - 2kP' - k^2g^2r^2 = 0 \tag{2.4-12}$$

where the prime indicates differentiation with respect to z. If Equation (2.4-12) is to hold for all r, the coefficient of the different powers of r must be equal to zero. This leads to

$$\left(\frac{1}{q}\right)^2 + \left(\frac{1}{q}\right)' + g^2 = 0 \quad \text{and} \quad P' = -\frac{i}{q} \tag{2.4-13}$$

The Helmholtz equation for ψ is thus reduced to Equation (2.4-13) for the functions $P(z)$ and $q(z)$. The first differential equation can be solved for $q(z)$. The other beam parameter $P(z)$ can then be integrated by using the second equation. In practice, a general solution for the nonlinear differential equation (2.4-13) is not available. In the next section, we consider the solution for the case of homogeneous media where $g = 0$.

2.5 GAUSSIAN BEAMS IN HOMOGENEOUS MEDIA

If the medium is homogeneous, we can, according to Equation (2.4-6), put $g = 0$, and Equation (2.4-13) becomes

$$\left(\frac{1}{q}\right)^2 + \left(\frac{1}{q}\right)' = 0 \quad \text{and} \quad P' = -\frac{i}{q} \tag{2.5-1}$$

where we recall that the prime indicates a differentiation with respect to z. We now introduce a function $u(z)$ by the relation

$$\frac{1}{q} = \frac{1}{u}\frac{du}{dz} \tag{2.5-2}$$

We obtain directly from Equation (2.5-1)

$$\frac{d^2u}{dz^2} = 0 \tag{2.5-3}$$

The general solution for Equation (2.5-3) can be written

$$u(z) = az + b \tag{2.5-4}$$

where a and b are arbitrary constants. The beam parameter $q(z)$ can then be written, according to Equation (2.5-2),

$$\frac{1}{q} = \frac{a}{az + b} \tag{2.5-5}$$

or equivalently,

$$q(z) = z + q_0 \tag{2.5-6}$$

where q_0 is a constant ($q_0 = q(0) = b/a$). The other beam parameter $P(z)$ can be obtained from Equations (2.5-1) and (2.5-6):

$$P' = -\frac{i}{q} = -\frac{i}{z + q_0} \tag{2.5-7}$$

so that, upon integration, one obtains

$$P(z) = -i \ln\left(1 + \frac{z}{q_0}\right) \tag{2.5-8}$$

where the arbitrary constant is taken to be zero. This is legitimate, as the constant will only modify the phase of the field solution (2.4-11). This is equivalent to a mere shift of the time origin.

Combining Equations (2.5-6) and (2.5-8) in Equation (2.4-11), we obtain the following cylindrically symmetric solution of the Helmholtz equation:

$$\psi = \frac{q_0}{q_0 + z} \exp\left(-i \frac{k}{2(q_0 + z)} r^2\right) \tag{2.5-9}$$

We recall that q_0 is an arbitrary complex constant. For a physically realizable beam with its energy concentrated around the axis, the field amplitude must approach zero as $r \to \infty$. This requires a choice of a complex q_0 with a positive imaginary part. We can lump the real part of q_0 with z by selecting a new origin of the z axis. This leads to a pure imaginary q_0. We can now reexpress q_0 in terms of a new constant ω_0:

$$q_0 = i \frac{\pi \omega_0^2 n}{\lambda} = i \frac{k \omega_0^2}{2} \equiv i z_0 \tag{2.5-10}$$

where n is the index of refraction of the medium and z_0 is the magnitude of q_0. To see the physical meaning of ω_0, we examine the field amplitude at $z = 0$,

$$\psi(z = 0) = \exp\left(-\frac{r^2}{\omega_0^2}\right) \tag{2.5-11}$$

We note the beam amplitude has a Gaussian distribution with a spot size of ω_0. So ω_0 is a measure of the beam spot at $z = 0$.

For $z \neq 0$, we can rewrite the beam amplitude as, according to Equations (2.5-9) and (2.5-10),

$$\psi = \frac{1}{1 - iz/z_0} \exp\left(-\frac{r^2}{\omega_0^2(1 - iz/z_0)}\right) \tag{2.5-12}$$

or equivalently,

$$\psi = \frac{1}{\sqrt{1 + (z/z_0)^2}} \exp\left(i \tan^{-1} \frac{z}{z_0}\right) \exp\left(-\frac{r^2}{\omega_0^2[1 + (z/z_0)^2]}(1 + iz/z_0)\right) \tag{2.5-13}$$

If we define the parameters

$$\omega^2(z) = \omega_0^2\left(1 + \frac{z^2}{z_0^2}\right) = \omega_0^2\left[1 + \left(\frac{\lambda z}{\pi \omega_0^2 n}\right)^2\right] \tag{2.5-14}$$

$$R(z) = z\left(1 + \frac{z_0^2}{z^2}\right) = z\left[1 + \left(\frac{\pi \omega_0^2 n}{\lambda z}\right)^2\right] \tag{2.5-15}$$

so that

$$\frac{1}{q(z)} = \frac{1}{z + iz_0} = \frac{1}{R(z)} - i \frac{\lambda}{\pi \omega^2(z) n} \tag{2.5-16}$$

and

$$\eta(z) = \tan^{-1}\left(\frac{z}{z_0}\right) = \tan^{-1}\left(\frac{\lambda z}{\pi\omega_0^2 n}\right) \tag{2.5-17}$$

then we can combine Equations (2.5-9) and (2.5-11)–(2.5-13) and, recalling that $E = \psi \exp(-ikz)$, obtain

$$E(x, y, z) = E_0 \frac{\omega_0}{\omega(z)} \exp\left(-i[kz - \eta(z)] - i\frac{kr^2}{2q(z)}\right)$$

or equivalently,

$$E(x, y, z) = E_0 \frac{\omega_0}{\omega(z)} \exp\left[-i[kz - \eta(z)] - r^2\left(\frac{1}{\omega^2(z)} + \frac{ik}{2R(z)}\right)\right] \tag{2.5-18}$$

where we recall that $k = 2\pi n/\lambda$. This is our basic result. We refer to it as the fundamental Gaussian-beam solution, since we have excluded the more complicated solutions of Equation (2.4-4) (i.e., those with azimuthal variation) by limiting ourselves to transverse dependence involving $r = \sqrt{x^2 + y^2}$ only. These higher-order modes will be discussed separately.

From Equation (2.5-18) the parameter $\omega(z)$, which evolves according to Equation (2.5-14), is the distance r at which the field amplitude is down by a factor $1/e$ compared to its value on the axis. We will consequently refer to it as the beam *spot size*. The parameter ω_0 is the minimum spot size. It is the beam spot size at the plane $z = 0$. The parameter R in Equation (2.5-18) is the radius of curvature of the very nearly spherical wavefronts at z. We can verify this statement by deriving the radius of curvature of the constant phase surfaces (wavefronts) or, more simply, by considering the form of a spherical wave emitted by a point radiator placed at $z = 0$. It is given by

$$\begin{aligned} E \propto \frac{1}{R} e^{-ikR} &= \frac{1}{R} \exp(-ik\sqrt{x^2 + y^2 + z^2}) \\ &\approx \frac{1}{R} \exp\left(-ikz - ik\frac{x^2 + y^2}{2R}\right), \quad x^2 + y^2 \ll z^2 \end{aligned} \tag{2.5-19}$$

since z is equal to R, the radius of curvature of the spherical wave. Comparing Equations (2.5-19) and (2.5-18), we identify R as the radius of curvature of the Gaussian beam, with the exception of the immediate vicinity of the plane $z = 0$. In arriving at the above conclusion, we ignored the phase factor $\eta(z)$, which is much smaller than kz for most practical situations. The complex parameter q is often referred to as the Gaussian beam parameter, or the complex radius of curvature of the Gaussian beam.

The convention regarding the sign of $R(z)$ is that it is negative if the center of curvature occurs at a position on the right side of z (i.e., $z' > z$) and vice versa. The form of the fundamental Gaussian beam is, according to Equation (2.5-18), uniquely determined once its minimum spot size ω_0 and its location—that is, the plane $z = 0$—are specified. The spot size $\omega(z)$ and radius of curvature R at any plane z are then found from Equations (2.5-14) and (2.5-15). Some of these characteristics are displayed in Figure 2.5. The hyperbolas shown in this figure correspond to the ray direction and are intersections of planes that include the z axis and the hyperboloids

$$x^2 + y^2 = \text{const. } \omega^2(z) \tag{2.5-20}$$

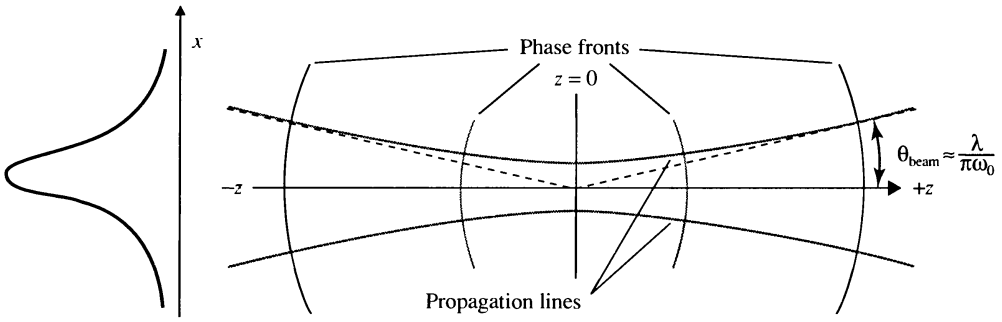


Figure 2.5 Propagating Gaussian beam.

These hyperbolas correspond to the local direction of energy propagation. The spherical surfaces shown have radii of curvature given by Equation (2.5-15). For large z the hyperboloids $x^2 + y^2 = \omega^2(z)$ are asymptotic to the cone

$$r = \sqrt{x^2 + y^2} = \frac{\lambda}{\pi\omega_0 n} z \tag{2.5-21}$$

whose half-apex angle, which we take as a measure of the angular beam spread, is

$$\theta_{\text{beam}} = \tan^{-1}\left(\frac{\lambda}{\pi\omega_0 n}\right) \cong \frac{\lambda}{\pi\omega_0 n} \quad \text{for} \quad \theta_{\text{beam}} \ll \pi \tag{2.5-22}$$

This last result is a rigorous manifestation of wave diffraction according to which a wave that is confined in the transverse direction to an aperture of radius ω_0 will spread (diffract) in the far field ($z \gg \pi\omega_0^2 n/\lambda$) according to Equation (2.5-22).

EXAMPLE: GAUSSIAN BEAM

The outputs of laser resonators are Gaussian beams. Consider a Gaussian beam with a beam spot size of $\omega_0 = 1$ mm at the laser output, and a wavelength of $\lambda = 1.06$ μm. According to Equation (2.5-10), the confocal parameter z_0 is

$$z_0 = \frac{\pi\omega_0^2 n}{\lambda} = \frac{\pi\omega_0^2}{\lambda} = 3 \text{ m}$$

Assume that the minimum spot size ω_0 is located at the output of the laser. The beam spot size at a distance of 10 m from the output becomes

$$\omega(z) = \omega_0\left(1 + \frac{z^2}{z_0^2}\right)^{1/2} = \omega_0\left[1 + \left(\frac{\lambda z}{\pi\omega_0^2}\right)^2\right]^{1/2} = 3.5 \text{ mm}$$

The radius of curvature of the wavefront at $z = 10$ m becomes

$$R(z) = z\left(1 + \frac{z_0^2}{z^2}\right) = z\left[1 + \left(\frac{\pi\omega_0^2}{\lambda z}\right)^2\right] = 10.9 \text{ m}$$

The beam divergence angle is given by

$$\theta_{\text{beam}} = \tan^{-1}\left(\frac{\lambda}{\pi\omega_0}\right) \cong \frac{\lambda}{\pi\omega_0} = 0.00034 \text{ radian} = 0.02°$$

2.6 FUNDAMENTAL GAUSSIAN BEAM IN A LENSLIKE MEDIUM—THE *ABCD* LAW

The Gaussian beam solution obtained in the previous section is a solution of the wave equation in homogeneous media. In this section, we return to the general case of a lenslike medium so that $g \neq 0$. The P and q functions of Equation (2.4-11) obey, according to Equation (2.4-13),

$$\left(\frac{1}{q}\right)^2 + \left(\frac{1}{q}\right)' + g^2 = 0 \quad \text{and} \quad P' = -\frac{i}{q} \tag{2.6-1}$$

If we introduce a function u defined by

$$\frac{1}{q} = \frac{1}{u}\frac{du}{dz} \tag{2.6-2}$$

we obtain from Equation (2.6-1)

$$\frac{d^2u}{dz^2} + g^2u = 0 \tag{2.6-3}$$

This is a differential equation for simple harmonic oscillators. The general solutions can be written

$$\begin{aligned} u(z) &= a \sin gz + b \cos gz \\ u'(z) &= ag \cos gz - bg \sin gz \end{aligned} \tag{2.6-4}$$

where a and b are arbitrary constants.

Using Equation (2.6-4) in Equation (2.6-2) and expressing the result in terms of an input value q_0, we obtain the following result for the complex beam parameter (complex radius of curvature) $q(z)$:

$$q(z) = \frac{\cos(gz)q_0 + \sin(gz)/g}{-\sin(gz)gq_0 + \cos(gz)} \tag{2.6-5}$$

This is the solution of Equation (2.6-1) that describes the evolution of the complex beam parameter $q(z)$, with an initial complex beam parameter of q_0 at $z = 0$. The physical significance of $q(z)$ in this case can be extracted from Equation (2.4-11). We focus on the part of $\psi(r, z)$ that involves r. Equation (2.4-11) can thus be written

$$\psi \propto e^{-ikr^2/2q(z)} \tag{2.6-6}$$

If we express the real and imaginary parts of $q(z)$ by means of

$$\frac{1}{q(z)} = \frac{1}{R(z)} - i\frac{\lambda}{\pi n \omega^2(z)} \tag{2.6-7}$$

we obtain

$$\psi \propto \exp\left(\frac{-r^2}{\omega^2(z)} - i\frac{kr^2}{2R(z)}\right) \tag{2.6-8}$$

so that $\omega(z)$ is the beam spot size and R its radius of curvature, as in the case of a homogeneous medium, which is described by Equation (2.5-18). For the special case of a homogeneous medium ($g = 0$), Equation (2.6-5) reduces to Equation (2.5-4).

Transformation of the Gaussian Beam—The *ABCD* Law

We have derived the transformation law of a Gaussian beam (Equation (2.6-5)) propagating through a lenslike medium that is characterized by g. We note first by comparing Equation (2.6-5) to Table 2.1(6) and to Equation (2.3-14) that the transformation of the Gaussian beam parameter (or the complex radius of curvature) can be described by

$$q_2 = \frac{Aq_1 + B}{Cq_1 + D} \tag{2.6-9}$$

where A, B, C, and D are the elements of the ray matrix that relate the ray (r, r') at plane 2 to the ray at plane 1. It follows immediately that the propagation through, or reflection from, any of the elements shown in Table 2.1 also obeys Equation (2.6-9), since these elements can all be viewed as special cases of a lenslike medium. This can also be seen by comparing Equation (2.6-3) to Equation (2.3-12). We note that the equation for the ray parameter $r(z)$ is identical to that of the function $u(z)$, defined in Equation (2.6-2). Equation (2.6-9) is known as a *bilinear transformation* (or *Möbius transformation*). For real matrix elements A, B, C, and D, it can be shown that q_2 is in the upper complex plane if q_1 is in the upper complex plane. In other words, a confined Gaussian beam remains a confined Gaussian beam provided the matrix elements are all real. For lenslike media, all matrix elements are real.

For future reference we note that by applying Equation (2.6-9) to a thin lens of focal length f we obtain from (2.6-9) and Table 2.1(2)

$$\frac{1}{q_2} = \frac{1}{q_1} - \frac{1}{f} \tag{2.6-10}$$

so that using Equation (2.6-7)

$$\begin{aligned} \omega_2 &= \omega_1 \\ \frac{1}{R_2} &= \frac{1}{R_1} - \frac{1}{f} \end{aligned} \tag{2.6-11}$$

These results apply, as well, to reflection from a mirror with a radius of curvature R if we replace f by $R/2$. Consider next the propagation of a Gaussian beam through two lenslike media that are adjacent to each other. The ray matrix describing the first one is (A_1, B_1, C_1, D_1) while that of the second one is (A_2, B_2, C_2, D_2). Taking the input beam parameter as q_1 and the output beam parameter as q_3 we have from Equation (2.6-9)

$$q_2 = \frac{A_1 q_1 + B_1}{C_1 q_1 + D_1}$$

for the beam parameter at the output of medium 1 and

$$q_3 = \frac{A_2 q_2 + B_2}{C_2 q_2 + D_2}$$

and after combining the last two equations,

$$q_3 = \frac{A_T q_1 + B_T}{C_T q_1 + D_T} \tag{2.6-12}$$

where A_T, B_T, C_T, D_T are the elements of the matrix relating the output plane (3) to the input plant (1); that is,

$$\begin{bmatrix} A_T & B_T \\ C_T & D_T \end{bmatrix} = \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \tag{2.6-13}$$

It follows by induction that Equation (2.6-9) applies to the propagation of a Gaussian beam through any arbitrary number of lenslike media and elements. The matrix with elements (A_T, B_T, C_T, D_T) is the ordered product of the matrices characterizing the individual members of the chain. The great power of the *ABCD* law is that it enables us to trace the Gaussian beam parameter $q(z)$ through a complicated sequence of lenslike elements. The beam radius $R(z)$ and spot size $\omega(z)$ at any plane z can be recovered through the use of Equation (2.6-7). The application of this method will be made clear by the following example.

EXAMPLE: GAUSSIAN BEAM FOCUSING

As an illustration of the application of the *ABCD* law, we consider the case of a Gaussian beam with a minimum beam spot size ω_{01} that is incident at a thin lens of focal length f, as shown in Figure 2.6. Let the waist of the incident beam be located at distance d_1 in front of the lens. We will find the location d_2 of the waist of the output beam and the beam spot size ω_{02} at that point.

At the input plane 1 (location of beam waist of incident beam), $\omega = \omega_{01}$ and $R_1 = \infty$, so that

$$\frac{1}{q_1} = \frac{1}{R_1} - i\frac{\lambda}{\pi\omega_{01}^2 n} = -i\frac{\lambda}{\pi\omega_{01}^2 n} \equiv \frac{1}{iz_1}$$

Similarly, at plane 2 (location of output beam waist), $\omega = \omega_{02}$ and $R_2 = \infty$, so the complex Gaussian beam parameter can be written

$$\frac{1}{q_2} = \frac{1}{R_2} - i\frac{\lambda}{\pi\omega_{02}^2 n} = -i\frac{\lambda}{\pi\omega_{02}^2 n} \equiv \frac{1}{iz_2}$$

These two beam parameters are related by the *ABCD* law:

$$q_2 = \frac{Aq_1 + B}{Cq_1 + D}$$

where A, B, C, and D are elements of the following matrix:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & d_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} \begin{bmatrix} 1 & d_1 \\ 0 & 1 \end{bmatrix}$$

By carrying out the matrix multiplications, we obtain

$$A = 1 - \frac{d_2}{f}, \quad B = d_1 + d_2 - \frac{d_1 d_2}{f}$$

$$C = -\frac{1}{f}, \quad D = 1 - \frac{d_1}{f}$$

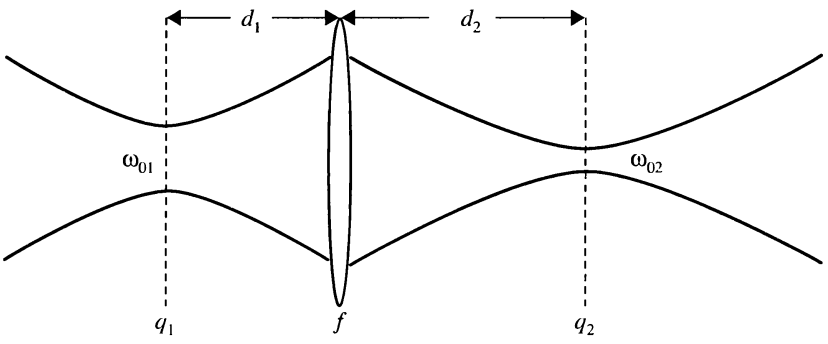


Figure 2.6 Focusing of a Gaussian beam.

Using the transformation formula, we have

$$iz_2 = \frac{Aiz_1 + B}{Ciz_1 + D} = \frac{(Aiz_1 + B)(-izC + D)}{C^2z_1^2 + D^2} = \frac{ACz_1^2 + BD + iz_1}{C^2z_1^2 + D^2}$$

where we have used the unimodular relationship $AD - BC = 1$. Since A, B, C, D, z_1, and z_2 are all real, we obtain

$$ACz_1^2 + BD = 0 \tag{2.6-14}$$

and

$$z_2 = \frac{z_1}{C^2z_1^2 + D^2} \tag{2.6-15}$$

Using $AD - BC = 1$ and eliminating z_1^2 from the above two equations, we obtain

$$z_2 = \frac{A}{D}z_1 = \frac{d_2 - f}{d_1 - f}z_1$$

or equivalently,

$$\omega_{02}^2 = \frac{d_2 - f}{d_1 - f}\omega_{01}^2 \tag{2.6-16}$$

The location of the output beam waist can be found by using Equation (2.6-14). We obtain

$$z_1^2 = -\frac{BD}{AC} = \frac{(d_1 - f)(d_1f + d_2f - d_1d_2)}{(d_2 - f)}$$

or equivalently,

$$(d_2 - f) = \frac{f^2}{z_1^2 + (d_1 - f)^2}(d_1 - f) \tag{2.6-17}$$

The minimum spot size of the output beam can now be written, according to Equations (2.6-16) and (2.6-17),

$$\omega_{02}^2 = \frac{f^2}{z_1^2 + (d_1 - f)^2}\omega_{01}^2 \tag{2.6-18}$$

The last two expressions are for the spot size and location of the beam waist.

We now examine some limiting cases of interest:

(a) For a point source with z_1 approaching zero, the image location is given by, according to Equation (2.6-17),

$$\frac{1}{d_1} + \frac{1}{d_2} = \frac{1}{f} \tag{2.6-19}$$

which is the familiar lens formula.

(b) For an input Gaussian beam of infinite spot size (plane wave with z_1 approaching infinity), the output beam waist is located at the focal point ($d_2 = f$).

(c) For an input Gaussian beam with its waist located at the front focal point ($d_1 = f$), the output beam waist is located at the rear focal point ($d_2 = f$). The output beam spot size at the waist is related to the input beam spot size by the following relationship:

$$z_1z_2 = f^2 \tag{2.6-20}$$

or equivalently,

$$\omega_{01}^2\omega_{02}^2 = \frac{\lambda^2}{\pi^2n^2}f^2 \tag{2.6-21}$$

(d) For an input Gaussian beam with its waist located at the lens ($d_1 = 0$), the output beam waist and location are given by

$$d_2 = \frac{f}{1 + (f/\pi\omega_{01}^2 n/\lambda)^2} = \frac{f}{1 + (f/z_1)^2} \tag{2.6-22}$$

and

$$\frac{\omega_{02}}{\omega_{01}} = \frac{f\lambda/\pi\omega_{01}^2 n}{\sqrt{1 + (f\lambda/\pi\omega_{01}^2 n)^2}} = \frac{f/z_1}{\sqrt{1 + (f/z_1)^2}} \tag{2.6-23}$$

The confocal beam parameter

$$z_1 \equiv \frac{\pi\omega_{01}^2 n}{\lambda}$$

is, according to Equation (2.5-14), the distance from the waist at which the input beam spot size increases by $\sqrt{2}$ and is a convenient measure of the convergence of the input beam. The smaller z_1, the "stronger" the convergence.

According to Equation (2.6-11), the radius of curvature of a beam is decreased by $1/f$, after passing through a thin lens, while the spot size remains unchanged. Using Equation (2.6-8), the input–output relationship can thus be written

$$E_{\text{out}}(x, y) = E_{\text{in}}(x, y) \exp\left(+i\frac{kr^2}{2f}\right) = E_{\text{in}}(x, y) \exp\left(+i\frac{k(x^2 + y^2)}{2f}\right) \tag{2.6-24}$$

where f is the focal length of the thin lens, and $k = 2\pi n/\lambda$.

2.7 GAUSSIAN BEAMS IN LENS WAVEGUIDE

As another example of the application of the *ABCD* law, we consider the propagation of a Gaussian beam through a sequence of thin lenses, as shown in Figure 2.2. The matrix relating a ray in plane $s = N$ to that in plane $s = 0$ is

$$\begin{bmatrix} A_T & B_T \\ C_T & D_T \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^N \tag{2.7-1}$$

where (A, B, C, D) are the elements of the matrix for propagation through a single two-lens, unit cell ($\Delta s = 1$) and is given by Equation (2.1-6). We can use the matrix identity of Equation (2.1-11) for the Nth power of a unimodular matrix to obtain

$$\begin{aligned} A_T &= \frac{A \sin(N\theta) - \sin[(N-1)\theta]}{\sin\theta} \\ B_T &= \frac{B \sin(N\theta)}{\sin\theta} \\ C_T &= \frac{C \sin(N\theta)}{\sin\theta} \\ D_T &= \frac{D \sin(N\theta) - \sin[(N-1)\theta]}{\sin\theta} \end{aligned} \tag{2.7-2}$$

where

$$\cos\theta = \tfrac{1}{2}(A + D) = \left(1 - \frac{d}{f_2} - \frac{d}{f_1} + \frac{d^2}{2f_1 f_2}\right) \tag{2.7-3}$$

and then use Equation (2.7-2) in (2.6-9) with the result

$$q_N = \frac{\{A\sin(N\theta) - \sin[(N-1)\theta]\}q_0 + B\sin(N\theta)}{C\sin(N\theta)q_0 + D\sin(N\theta) - \sin[(N-1)\theta]} \tag{2.7-4}$$

We now examine the behavior of q_N as N becomes infinity. For the case $|\cos\theta| > 1$, the sine functions on the right side will yield growing exponentials. This leads to a limiting value on the right side. It can be shown that this limiting value is a real number (see Problem 2.7). A Gaussian beam with a real beam parameter q is an unconfined beam. For the case $|\cos\theta| \le 1$, the right side remains oscillatory as a function of N. The bilinear transformation with real constants will maintain the confinement of the beam. The condition for the confinement of the Gaussian beam by the lens sequence is thus $|\cos\theta| \le 1$, or

$$0 \le \left(1 - \frac{d}{2f_1}\right)\left(1 - \frac{d}{2f_2}\right) \le 1 \tag{2.7-5}$$

that is, the same as condition (2.1-16) for stable ray propagation.

2.8 HIGH-ORDER GAUSSIAN BEAM MODES IN A HOMOGENEOUS MEDIUM

The Gaussian mode treated up to this point has a field variation that depends only on the axial distance z and the distance r from the axis. If we do not impose the condition $\partial/\partial\phi = 0$ (where ϕ is the azimuthal angle in a cylindrical coordinate system (r, ϕ, z)) and take $g = 0$, the wave equation (2.4-4) has solutions in the form of

$$\begin{aligned} E_{l,m}(x,y,z) &= E_0 \frac{\omega_0}{\omega(z)} H_l\left(\sqrt{2}\frac{x}{\omega(z)}\right) H_m\left(\sqrt{2}\frac{y}{\omega(z)}\right) \times \exp\left(-ik\frac{x^2+y^2}{2q(z)} - ikz + i(l+m+1)\eta\right) \\ &= E_0 \frac{\omega_0}{\omega(z)} H_l\left(\sqrt{2}\frac{x}{\omega(z)}\right) H_m\left(\sqrt{2}\frac{y}{\omega(z)}\right) \times \exp\left(-\frac{x^2+y^2}{\omega^2(z)} - ik\frac{x^2+y^2}{2R(z)} - ikz + i(l+m+1)\eta\right) \end{aligned} \tag{2.8-1}$$

where H_l is the Hermite polynomial of order l, and $\omega(z)$, $R(z)$, $q(z)$, and η are defined as in Equations (2.5-14)–(2.5-17) [1]. We note for future reference that the phase shift on the axis is

$$\begin{aligned} \theta &= kz - (l + m + 1)\tan^{-1}\left(\frac{z}{z_0}\right) \\ z_0 &= \frac{\pi\omega_0^2 n}{\lambda} \end{aligned} \tag{2.8-2}$$

The transverse variation of the electric field along x (or y) is seen to be of the form $H_l(\xi)\exp(-\xi^2/2)$, where $\xi = \sqrt{2}x/\omega$. This function has been studied extensively, since it also corresponds to the quantum mechanical wavefunction $u_l(\xi)$ of the harmonic oscillator [2]. The Hermite–Gaussian beams form a complete set of orthogonal functions. They satisfy the following orthogonal relationship:

$$\iint E^*_{l_1 m_1}(x, y, z)\, E_{l_2 m_2}(x, y, z)\, dx\, dy = \delta_{(l_1 m_1)(l_2 m_2)} \tag{2.8-3}$$

These Hermite–Gaussian functions can be used as a basis to expand any arbitrary paraxial optical beam $E(x, y, z)$ in the form

$$E(x, y, z) = \sum_l \sum_m c_{lm} E_{l,m}(x, y, z) \tag{2.8-4}$$

where c_{lm} are constants that can be obtained by using the orthogonal relationship.

An alternative but equally valid family of solutions to the paraxial wave equation can be written in cylindrical rather than rectangular coordinates. They are the so-called Laguerre–Gaussian beams of the form

$$E_{p,m}(r, \phi, z) = E_0 \frac{\omega_0}{\omega(z)} \left(\frac{\sqrt{2}r}{\omega(z)}\right)^{|m|} L_p^{|m|}\left(\frac{2r^2}{\omega^2(z)}\right) \exp\left[-ik\frac{r^2}{2q(z)} - ikz + im\phi + i(2p + |m| + 1)\eta\right]$$

$$= E_0 \frac{\omega_0}{\omega(z)} \left(\frac{\sqrt{2}r}{\omega(z)}\right)^{|m|} L_p^{|m|}\left(\frac{2r^2}{\omega^2(z)}\right) \exp\left[-\frac{r^2}{\omega^2(z)} - ik\frac{r^2}{2R(z)} - ikz + im\phi + i(2p + |m| + 1)\eta\right] \tag{2.8-5}$$

where L_p^m functions are the associated Laguerre polynomials [3]. In the above solutions, the integer $p \geq 0$ is the radial index and the integer m is the azimuthal index. The Laguerre–Gaussian beams are also mutually orthogonal and are particularly useful for problems with cylindrical symmetry. As a result of the cylindrical symmetry, the radial dependence of the Laguerre–Gaussian beams with $\exp(\pm im\phi)$ are identical. In addition, the intensity patterns of Laguerre–Gaussian beams are cylindrically symmetric (i.e., no ϕ dependence), according to Equation (2.8-5). We also note the fundamental Laguerre–Gaussian beam ($p = 0$, $m = 0$) is exactly the same as the fundamental Hermite–Gaussian beam ($l = 0$, $m = 0$).

Some low-order Hermite–Gaussian functions normalized to represent the same amount of total beam power are shown in Figure 2.7. Photographs of actual field patterns are shown in Figure 2.8. Note that the first four correspond to the intensity $|u_l(\xi)|^2$ plots ($l = 0, 1, 2, 3$) of Figure 2.7.

The following is a list of the first few Hermite polynomials and associated Laguerre polynomials.

$$H_0(x) = 1$$

$$H_1(x) = 2x$$

$$H_2(x) = 4x^2 - 2$$

$$H_3(x) = 8x^3 - 12x$$

$$H_4(x) = 16x^4 - 48x^2 + 12$$

$$L_0^m(x) = 1$$

$$L_1^m(x) = -x + (m + 1)$$

$$L_2^m(x) = \tfrac{1}{2}[x^2 - 2(m + 2)x + (m + 1)(m + 2)]$$

$$L_3^m(x) = \tfrac{1}{6}[-x^3 + 3(m + 3)x^2 - 3(m + 2)(m + 3)x + (m + 1)(m + 2)(m + 3)]$$

These polynomials are provided as a ready source of reference.

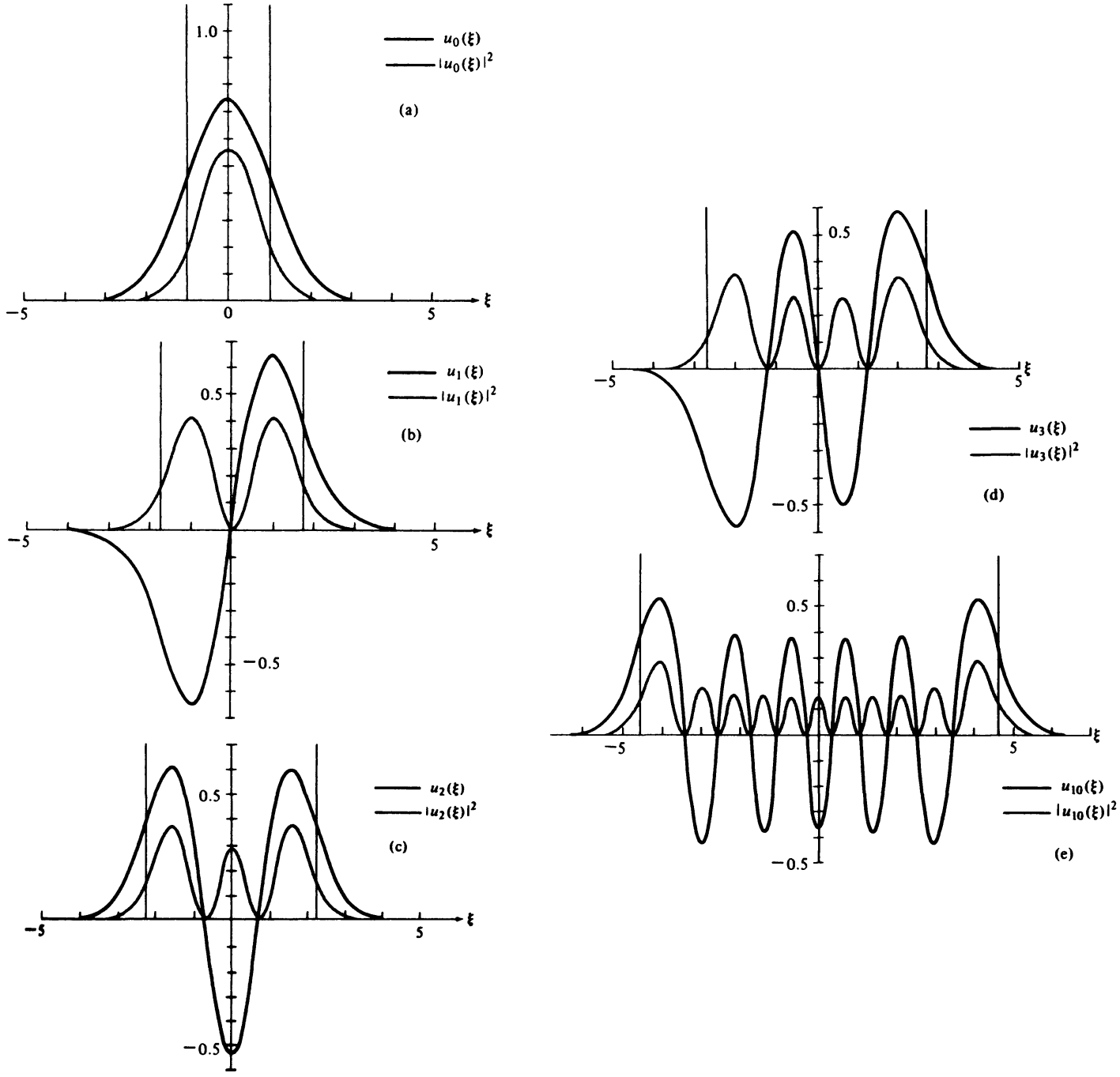


Figure 2.7 Hermite–Gaussian functions $u_l(\xi) = (\pi^{1/2} l! 2^l)^{-1/2} H_l(\xi) e^{-\xi^2/2}$ corresponding to higher-order beam solutions (Equation (2.8-1)). The curves are normalized so as to represent a fixed amount of total beam power in all the modes $\left(\int_{-\infty}^{\infty} u_l^2(\xi)\, d\xi = 1\right)$. The solid curves are the functions $u_l(\xi)$ for $l = 0$, 1, 2, 3, and 10. The curves limited to the upper half-space are $u_l^2(\xi)$.

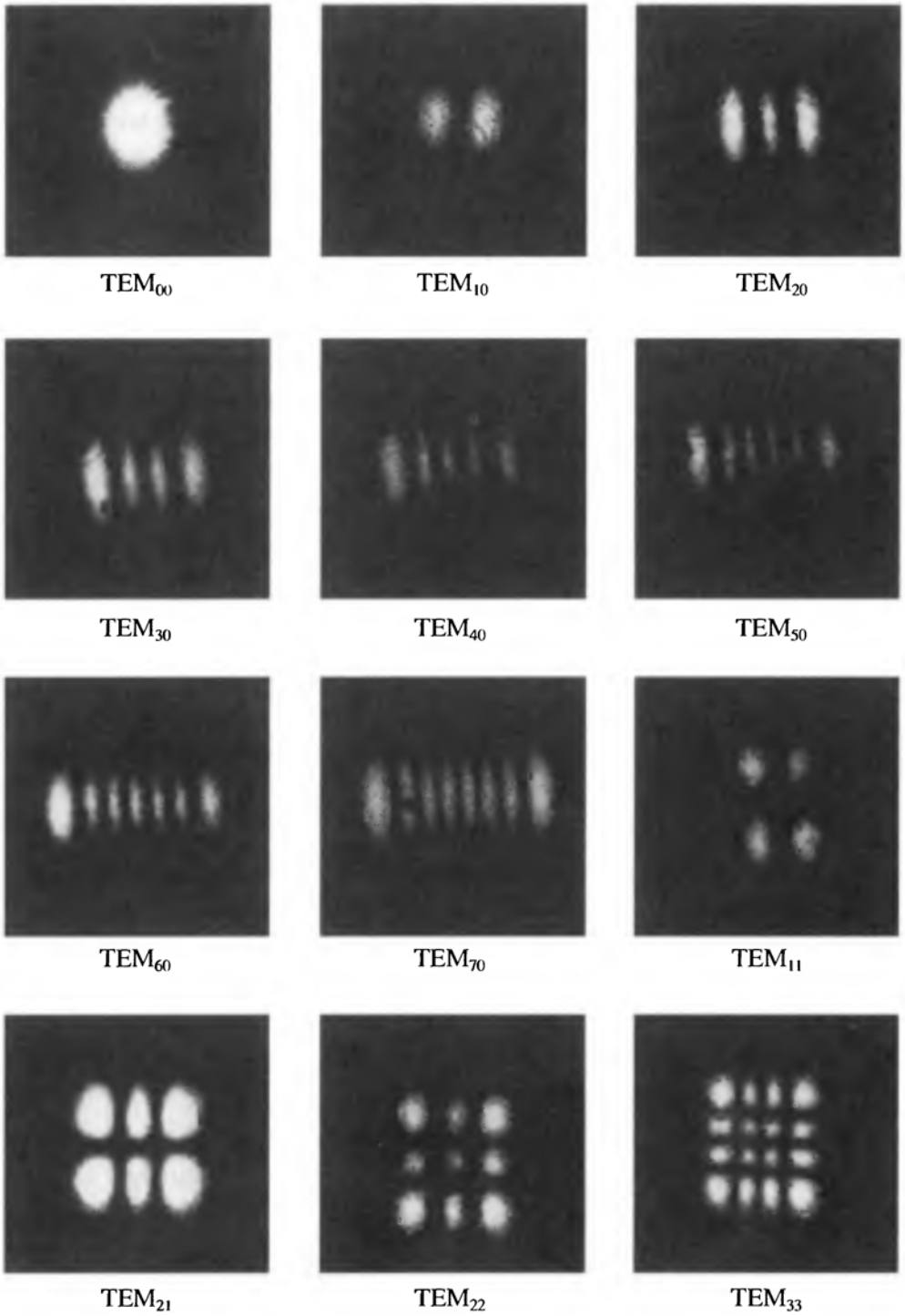


Figure 2.8 Intensity photographs of some low-order Gaussian beam modes. (After Reference [4].)

2.9 GAUSSIAN BEAM MODES IN QUADRATIC INDEX MEDIA

In Section 2.6 we treated the propagation of a circularly symmetric Gaussian beam in a lenslike medium by using the Gaussian beam solution of the homogeneous medium. Since the Gaussian beam solution of the homogeneous medium is not a solution in lenslike media, this treatment is possible provided we allow the complex Gaussian beam parameter q to

vary as a function of z according to Equation (2.6-1). Thus, in this approach, the complex beam parameter q of the Gaussian beam of the homogeneous medium will vary as a function of z according to Equation (2.6-4).

Here we describe a different approach. In this approach, we find directly the Gaussian beam solution of the lenslike medium. In other words, we are looking for the modes of propagation in a lenslike medium whose index of refraction can be described by

$$n^2(r) = n_0^2(1 - g^2r^2) \qquad r^2 = x^2 + y^2 \tag{2.9-1}$$

where n_0 is the refractive index at the axis of symmetry and g is a constant. For modes of propagation, the transverse field distribution remains unchanged as the modes propagate in the medium. For these modes of propagation the complex beam parameter q is a constant.

The vector wave equation (2.4-4) for lenslike media takes the form

$$\nabla^2\mathbf{E} + k^2(1 - g^2r^2)\mathbf{E} = 0 \tag{2.9-2}$$

where $k = n_0\omega/c$ is the wavenumber. We consider some (scalar) component E of the last equation and assume a solution in the form

$$E(x, y) = \psi(x, y)\exp(-i\beta z) \tag{2.9-3}$$

where β is a constant, known as the propagation constant. By examining the index profile and the wave equation, we note that the wave equation in this case is separable. Taking $\psi(x, y) = U(x)V(y)$, the wave equation (2.9-2) becomes

$$\frac{1}{U}\frac{\partial^2 U}{\partial x^2} + \frac{1}{V}\frac{\partial^2 V}{\partial y^2} + k^2 - k^2g^2(x^2 + y^2) - \beta^2 = 0 \tag{2.9-4}$$

where U and V are one-dimensional functions. By grouping x-dependent terms and y-dependent terms in Equation (2.9-4), we obtain

$$\frac{1}{U}\frac{\partial^2 U}{\partial x^2} + (k^2 - \beta^2 - k^2g^2x^2) = C \tag{2.9-5}$$

$$\frac{1}{V}\frac{\partial^2 V}{\partial y^2} - k^2g^2y^2 = -C \tag{2.9-6}$$

where C is some constant. We note that the differential equation (2.9-2) is now separated into two. Each of these equations can be solved separately. Defining a variable ξ by

$$\xi = \alpha y, \quad \alpha \equiv \sqrt{gk} \tag{2.9-7}$$

we can rewrite Equation (2.9-6) as

$$\frac{d^2V}{d\xi^2} + \left(\frac{C}{\alpha^2} - \xi^2\right)V = 0 \tag{2.9-8}$$

This is a well-known differential equation and is identical to the Schrödinger equation of the harmonic oscillator [2]. The solution and the corresponding eigenvalues can be written

$$\frac{C}{\alpha^2} = (2m + 1), \quad m = 1, 2, 3, \ldots \tag{2.9-9}$$

$$V_m(\xi) = H_m(\xi)e^{-\xi^2/2} \tag{2.9-10}$$

where H_m is the Hermite polynomial of order m.

We now repeat the procedure with Equation (2.9-5). Substituting

$$\zeta = \alpha x$$

we obtain

$$\frac{\partial^2 U}{\partial \zeta^2} + \left(\frac{k^2 - \beta^2 - C}{\alpha^2} - \zeta^2 \right) U = 0$$

The eigenvalues and eigenfunctions are given by

$$\frac{k^2 - \beta^2 - C}{\alpha^2} = (2l + 1), \quad l = 1, 2, 3, \ldots \tag{2.9-11}$$

and

$$U_l(\zeta) = H_l(\zeta) e^{-\zeta^2/2} \tag{2.9-12}$$

The total solution for the wavefunction ψ is thus

$$\psi(x, y) = H_l\left(\frac{\sqrt{2}x}{\omega_0} \right) H_m\left(\frac{\sqrt{2}y}{\omega_0} \right) e^{-(x^2+y^2)/\omega_0^2}$$

where the "spot size" ω_0 is, according to Equation (2.9-7),

$$\omega_0 = \frac{\sqrt{2}}{\alpha} = \sqrt{\frac{2}{gk}} = \sqrt{\frac{\lambda}{gn_0\pi}} \tag{2.9-13}$$

This beam spot size can also be obtained directly from Equation (2.4-13) by assuming $q' = 0$. In other words, the complex beam parameter of the modes of propagation is a constant given by $q = i/g$.

The complex field of the mode of propagation is

$$\begin{aligned} E_{l,m}(x, y, z) &= \psi_{l,m}(x, y) e^{-i\beta_{l,m} z} \\ &= E_0 H_l\left(\sqrt{2}\frac{x}{\omega_0} \right) H_m\left(\sqrt{2}\frac{y}{\omega_0} \right) \exp\left(-\frac{x^2 + y^2}{\omega_0^2} \right) \exp(-i\beta_{l,m} z) \end{aligned} \tag{2.9-14}$$

where E_0 is a constant. The propagation constant $\beta_{l,m}$ of the l, m mode is obtained from Equations (2.9-9) and (2.9-11):

$$\beta_{l,m} = k\left(1 - \frac{2}{k} g(l + m + 1) \right)^{1/2} \tag{2.9-15}$$

Two features of the mode solutions are noteworthy. (1) Unlike the homogeneous medium solution ($g = 0$), the mode spot size ω is independent of z. This can be explained by the focusing action of the index variation ($g > 0$), which counteracts the natural tendency of a confined beam to diffract (spread). In the case of an index of refraction that increases with r ($g < 0$), it follows from Equations (2.9-13) and (2.9-14) that $\omega^2 < 0$ and no confined solutions exist. The index profile in this case leads to defocusing, thus reinforcing the diffraction of the beam. (2) The dependence of β on the mode indices l, m causes the different modes to have phase velocities $v_{l,m} = \omega/\beta_{l,m}$ as well as group velocities $(v_g)_{l,m} = d\omega/d\beta_{l,m}$ that depend on l and m.

Let us consider modal dispersion (i.e., the dependence on l and m) of the group velocity of mode

$$(v_g)_{l,m} = \frac{d\omega}{d\beta_{l,m}} \tag{2.9-16}$$

If the index variation is small so that

$$\frac{1}{k} g(l + m + 1) \ll 1 \tag{2.9-17}$$

we can approximate Equation (2.9-15) as

$$\beta_{l,m} \cong k - g(l + m + 1) - \frac{g^2}{2k}(l + m + 1)^2 \tag{2.9-18}$$

so that, according to Equation (2.9-16),

$$(v_g)_{l,m} = \frac{c/n_0}{\left(1 + \frac{g^2}{2k^2}(l + m + 1)^2\right)} \tag{2.9-19}$$

We have shown in Chapter 1 that optical pulses propagate at the group velocity v_g. The effect of the group velocity dispersion on pulse propagation in lenslike media is considered next. We will also treat the effect of group velocity dispersion ($dv_g/d\omega \neq 0$) on pulse broadening.

Pulse Spreading in Quadratic Index Glass Fibers

Glass fibers with quadratic index profiles (2.9-1) can be employed for optical communications systems [5, 6]. The signals are coded into trains of optical pulses and the information capacity is thus fundamentally limited by the number of pulses that can be transmitted per unit time [7, 8]. There are two ways in which the dispersion limits the pulse repetition rate: modal dispersion and group velocity dispersion.

Modal Dispersion

If the optical pulses fed into the input end of the fiber excite a large number of modes (this will be the case if the input light is strongly focused so that the "rays" subtend a large angle), then each mode will travel with a group velocity $(v_g)_{l,m}$, as given by Equation (2.9-19). If all the modes from (0, 0) to ($l_{\max}$, $m_{\max}$) are excited, the output pulse at $z = L$ will broaden to

$$\Delta\tau \cong L\left(\frac{1}{(v_g)_{l_{\max},m_{\max}}} - \frac{1}{(v_g)_{0,0}}\right) \tag{2.9-20}$$

We can use Equation (2.9-19) to obtain

$$\Delta\tau \cong \frac{n_0 g^2 L}{2ck^2}[(l_{\max} + m_{\max} + 1)^2 - 1] \tag{2.9-21}$$

The maximum number of pulses per second that can be transmitted without serious overlap of adjacent output pulses is thus $f_{\max} \sim 1/\Delta\tau$. High data rate transmission will thus require the use of single-mode excitation, which can be achieved by the use of coherent single-mode laser excitation [6–8, 9].

EXAMPLE: MAXIMUM PULSE RATE

Consider a 1 km long quadratic index fiber with $n_0 = 1.5$ and $g^2 = 3.4 \times 10^3$ cm^{-2}. Let the input optical pulses at $\lambda = 1.0$ μm excite the modes up to $l_{max} = m_{max} = 30$. Substitution in Equation (2.9-21) gives

$$\Delta\tau = 3.6 \times 10^{-9} \text{ s}$$

and $f_{max} \sim (\Delta\tau)^{-1} = 2.8 \times 10^8$ pulses per second for the maximum pulse rate.

Group Velocity Dispersion

The pulse spreading (2.9-21) due to multimode excitation can be eliminated if one were to excite a single mode, say, l, m only. In this case, pulse spreading would still result from the dependence of $(v_g)_{l,m}$ on frequency. This spreading can be explained by the fact that a pulse with a spectral width $\Delta\omega$ will spread in a distance L by

$$\Delta\tau = \Delta\left(\frac{L}{v_g}\right) \approx L\left|\frac{d}{d\omega}\left(\frac{1}{v_g}\right)\right|\Delta\omega = \frac{L}{v_g^2}\left|\frac{dv_g}{d\omega}\right|\Delta\omega \tag{2.9-22}$$

If the pulse is derived, say, by gating, from a coherent continuous source with a negligible spectral width, the pulse spectral width is related to the pulse duration τ by $\Delta\omega \sim 2\pi/\tau$ and Equation (2.9-22) becomes

$$\Delta\tau \approx \frac{2\pi L}{v_g^2\tau}\left|\frac{dv_g}{d\omega}\right| \tag{2.9-23}$$

If the source bandwidth $\Delta\omega$ exceeds $2\pi/\tau$, then we need to replace $\Delta\omega$ in Equation (2.9-22) by the actual spectral width $\Delta\omega_s$. A rigorous treatment of this subject is reserved for Chapter 7.

In the situation when $g(l + m + 1) \ll k$, so that we can neglect the second-order term in Equation (2.9-18), we have

$$\beta_{l,m} \cong k - g(l + m + 1) \tag{2.9-24}$$

We note that the propagation constants of the modes are equally spaced with a spacing of g. It can easily be shown that the field distribution in the multimode waveguide with equally spaced propagation constants will reproduce itself periodically. In other words,

$$E(x, y, z + 2\pi/g) = E(x, y, z) \tag{2.9-25}$$

This result is consistent with the periodic solutions of rays in lenslike media.

2.10 PROPAGATION IN MEDIA WITH A QUADRATIC GAIN PROFILE

In many laser media the gain is a strong function of position. This variation can be due to a variety of causes, among them: (1) the radial distribution of energetic electrons in the plasma region of gas lasers [10], (2) the variation of pumping intensity in solid state lasers, and (3) the dependence of the degree of gain saturation on the radial position in the beam. We can account for an optical medium with quadratic gain (or loss) variation by taking the complex propagation constant $k(r)$ in Equation (2.4-5) as

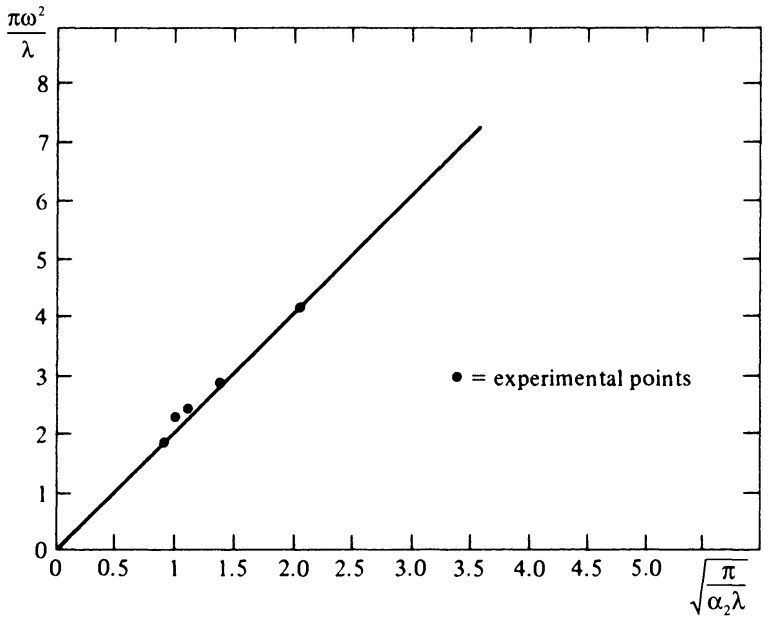


Figure 2.9 Theoretical curve showing the dependence of beam radius on quadratic gain constant α_2. Experimental points were obtained in a xenon 3.39 μm laser in which α_2 was varied by controlling the unsaturated laser gain. (After Reference [11].)

$$k(r) = k \pm i(\alpha_0 - \tfrac{1}{2}\alpha_2 r^2) \tag{2.10-1}$$

where α_0 and α_2 are constants. The plus (minus) sign applies to the case of gain (loss). The expression used previously for the quadratic index variation can be employed here, provided we treat the constant g in Equation (2.4-6) as a complex number and we assume $g^2 r^2 \ll 1$. In other words, Equation (2.10-1) can be obtained from Equation (2.4-6) if we take $g^2 = i\alpha_2/k$. Using this value in Equation (2.4-13) to obtain the steady-state ($(1/q)' = 0$) solution of the complex beam radius q yields

$$\frac{1}{q} = -ig = -i\sqrt{\frac{i\alpha_2}{k}} \tag{2.10-2}$$

The steady-state beam radius and spot size are obtained from Equations (2.6-7) and (2.10-2):

$$\begin{aligned} \omega_0^2 &= 2\sqrt{\frac{\lambda}{\pi n \alpha_2}} \\ R &= 2\sqrt{\frac{\pi n}{\lambda \alpha_2}} \end{aligned} \tag{2.10-3}$$

where n is the real part of the index of refraction. We thus find that the steady-state solution corresponds to a beam with a constant spot size but with a finite radius of curvature. The general (non-steady-state) behavior of the Gaussian beam in a quadratic gain medium is described in terms of its complex beam parameter $q(z)$ given by Equation (2.6-5), where $g^2 = i\alpha_2/k$.

Experimental data showing a decrease of the beam spot size with increasing gain parameter α_2 in agreement with Equation (2.10-3) are shown in Figure 2.9. "Steady state" here refers not to the intensity, which according to Equation (2.10-1) is growing or decaying with z, but to the beam radius of curvature and spot size.

2.11 ELLIPTIC GAUSSIAN BEAMS

All the beam solutions considered up to this point have one feature in common. The field drops off as in Equation (2.5-18), according to

$$E_{l,m} \propto \exp\left(-\frac{x^2 + y^2}{\omega^2(z)}\right) \tag{2.11-1}$$

so that the locus in the xy plane of the points where the field is down by a factor of $1/e$ from its value on the axis is a circle of radius $\omega(z)$. We will refer to such beams as *circular Gaussian beams*.

The wave equation (2.4-9) also admits solutions in which the variation in the x and y directions is characterized by

$$E_{l,m} \propto \exp\left(-\frac{x^2}{\omega_x^2(z)} - \frac{y^2}{\omega_y^2(z)}\right) \tag{2.11-2}$$

with $\omega_x \neq \omega_y$. Such beams, which we name *elliptic Gaussian*, result, for example, when a circular Gaussian beam passes through a cylindrical lens or when a laser beam emerges from an astigmatic resonator—that is, one whose mirrors possess different radii of curvature in the zy and zx planes.

We will not repeat the whole derivation for this case, but will indicate the main steps. Instead of Equation (2.4-11), we assume a solution

$$\psi = \exp\left[-i\left(P(z) + \frac{k}{2q_x(z)}x^2 + \frac{k}{2q_y(z)}y^2\right)\right] \tag{2.11-3}$$

where q_x and q_y are the complex Gaussian beam parameters. We also generalize Equation (2.4-6) and assume an astigmatic quadratic index medium with the following index profile:

$$n^2(x, y) = n_0^2(1 - g_x^2x^2 - g_y^2y^2)$$

where g_x and g_y are constants.

Using Equation (2.11-3) and following the approach in Section 2.4, we obtain

$$\left(\frac{1}{q_x}\right)^2 + \left(\frac{1}{q_x}\right)' + g_x^2 = 0, \qquad \left(\frac{1}{q_y}\right)^2 + \left(\frac{1}{q_y}\right)' + g_y^2 = 0 \tag{2.11-4}$$

and

$$\frac{dP}{dz} = -\frac{i}{2}\left(\frac{1}{q_x} + \frac{1}{q_y}\right) \tag{2.11-5}$$

where the prime indicates a differentiation with respect to z. In the case of a homogeneous ($g_x = g_y = 0$) beam we obtain, as in Equation (2.5-6),

$$q_x(z) = z + C_x \tag{2.11-6}$$

where C_x is an arbitrary constant of integration. We find it useful to write C_x as

$$C_x = -z_x + q_{0x} \tag{2.11-7}$$

where z_x is real and q_{0x} is imaginary. The physical significance of these two constants will become clear in what follows. A similar result with $x \rightarrow y$ is obtained for $q_y(z)$. Using the solutions of $q_x(z)$ and $q_y(z)$ in Equation (2.11-5) gives

$$P = -\frac{i}{2}\left[\ln\left(1 + \frac{z - z_x}{q_{0x}}\right) + \ln\left(1 + \frac{z - z_y}{q_{0y}}\right)\right]$$

Proceeding straightforwardly, as in the derivation connecting Equations (2.5-6)–(2.5-18), we find

$$E(x, y, z) = E_0 \frac{\sqrt{\omega_{0x}\omega_{0y}}}{\sqrt{\omega_x(z)\omega_y(z)}} \exp\left(-i[kz - \eta(z)] - \frac{ikx^2}{2q_x(z)} - \frac{iky^2}{2q_y(z)}\right)$$

$$= E_0 \frac{\sqrt{\omega_{0x}\omega_{0y}}}{\sqrt{\omega_x(z)\omega_y(z)}} \exp\left[-i[kz - \eta(z)] - x^2\left(\frac{1}{\omega_x^2(z)} + \frac{ik}{2R_x(z)}\right) - y^2\left(\frac{1}{\omega_y^2(z)} + \frac{ik}{2R_y(z)}\right)\right] \tag{2.11-8}$$

where

$$q_{0x} = i\frac{\pi\omega_{0x}^2 n}{\lambda}$$

$$\omega_x^2(z) = \omega_{0x}^2\left[1 + \left(\frac{\lambda(z - z_x)}{\pi\omega_{0x}^2 n}\right)^2\right] \tag{2.11-9}$$

$$R_x(z) = (z - z_x)\left[1 + \left(\frac{\pi\omega_{0x}^2 n}{\lambda(z - z_x)}\right)^2\right]$$

where $n = n_0$ is the index of refraction of the medium. Similar expressions for q_{0y}, ω_y, and R_y can be obtained by $x \rightarrow y$.

The phase delay $\eta(z)$ in Equation (2.11-8) is now given by

$$\eta(z) = \tfrac{1}{2}\tan^{-1}\left(\frac{\lambda(z - z_x)}{\pi\omega_{0x}^2 n}\right) + \tfrac{1}{2}\tan^{-1}\left(\frac{\lambda(z - z_y)}{\pi\omega_{0y}^2 n}\right) \tag{2.11-10}$$

It follows that *all* the results derived for the case of circular Gaussian beams apply, separately, to the xz and to the yz behavior of the elliptic Gaussian beam. For the purpose of analysis, the elliptic beam can be considered as two independent "beams." The position of the waist is not necessarily the same for these two beams. It occurs at $z = z_x$ for the xz beam and at $z = z_y$ for the yz beam in the example of Figure 2.10, where z_x and z_y are arbitrary. It also follows from the similarity between Equations (2.11-4) and (2.4-11) that the *ABCD* transformation law (2.6-9) can be applied separately to $q_x(z)$ and $q_y(z)$, which, according to Equation (2.11-8), are given by

$$\frac{1}{q_x(z)} = \frac{1}{R_x(z)} - i\frac{\lambda}{\pi n\omega_x^2(z)}$$

$$\frac{1}{q_y(z)} = \frac{1}{R_y(z)} - i\frac{\lambda}{\pi n\omega_y^2(z)} \tag{2.11-11}$$

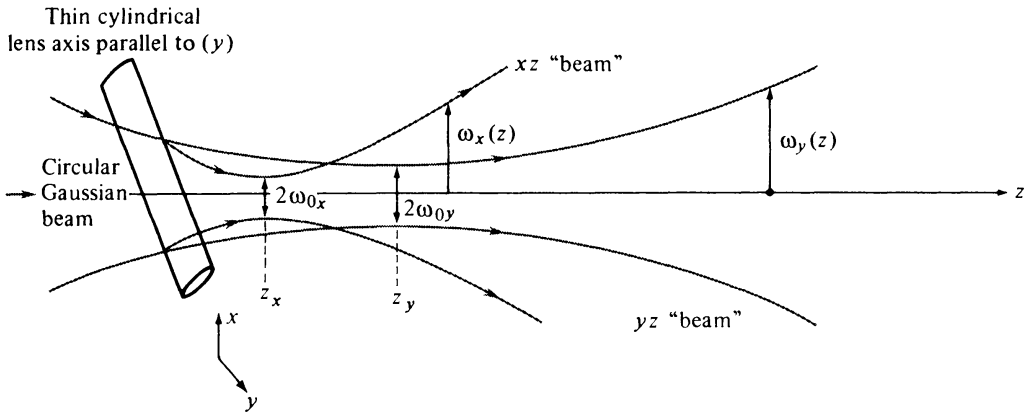


Figure 2.10 Illustration of an elliptic beam produced by cylindrical focusing of a circular Gaussian beam.

Elliptic Gaussian Beams in a Quadratic Lenslike Medium

Here we consider the mode of propagation in an astigmatic medium whose index of refraction is given by

$$n^2(x, y) = n_0^2(1 - g_x^2x^2 - g_y^2y^2) \tag{2.11-12}$$

where n_0, g_x, and g_y are constants. These modes of propagation have well-defined transverse wavefunctions that remain the same as they propagate. The derivation is identical to that presented in Section 2.9, resulting in the following wavefunctions for elliptic Gaussian beams:

$$E_{l,m}(r) = E_0 e^{-i\beta_{l,m}z} H_l\left(\sqrt{2}\frac{x}{\omega_x}\right) H_m\left(\sqrt{2}\frac{y}{\omega_y}\right) \exp\left(-\frac{x^2}{\omega_x^2} - \frac{y^2}{\omega_y^2}\right) \tag{2.11-13}$$

where E_0 is a constant and

$$\omega_x = \sqrt{\frac{2}{g_x k}} = \sqrt{\frac{\lambda}{g_x n_0 \pi}}, \quad \omega_y = \sqrt{\frac{2}{g_y k}} = \sqrt{\frac{\lambda}{g_y n_0 \pi}} \tag{2.11-14}$$

$$\beta_{l,m} = k\left\{1 - \frac{2}{k}\left[g_x\left(l + \frac{1}{2}\right) + g_y\left(m + \frac{1}{2}\right)\right]\right\}^{1/2}, \quad m, l = 0, 1, 2, 3, \ldots \tag{2.11-15}$$

Stripe-geometry GaAs-AlGaAs semiconductor lasers have an astigmatic quadratic-index profile approximately described by Equation (2.11-12) with $g_x \neq g_y$. Figure 2.11 shows the intensity patterns of the elliptic Gaussian beam modes of these lasers with various stripe widths (ω = 10 μm–50 μm). The principles of operation of semiconductor lasers will be discussed in Chapters 15 and 16.

2.12 BEAM PROPAGATION AND DIFFRACTION INTEGRAL

In previous sections, we considered the propagation of rays and Gaussian beams in homogeneous media. Their propagation can easily be described by using *ABCD* matrices. Here, we consider the propagation of a general beam of light with a finite transverse dimension. We will first describe the beam propagation by decomposing the beam into a superposition of plane waves. By virtue of the linearity of the wave equation, all optical wave of finite

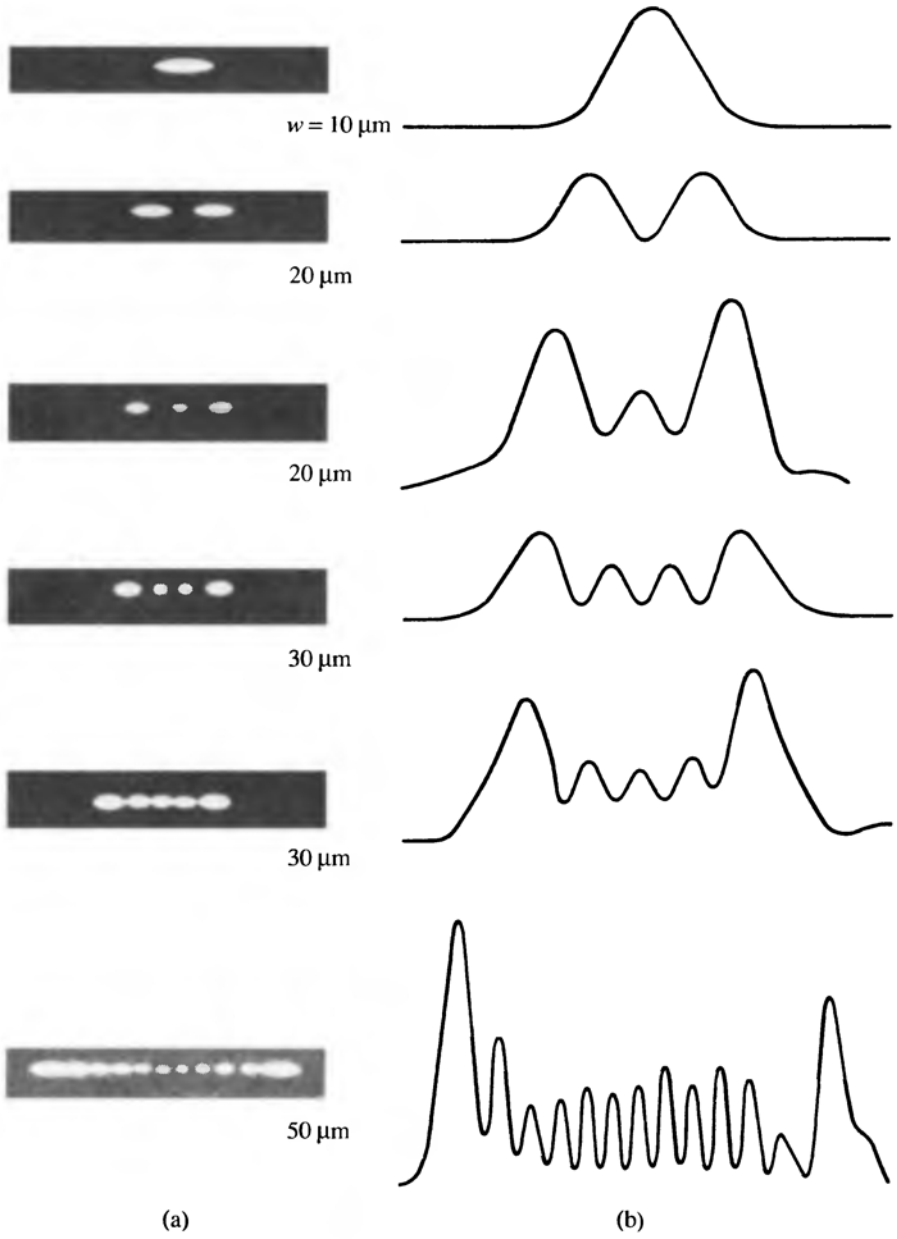


Figure 2.11 (a) Near-field and (b) far-field intensity distributions of the output of stripe contact GaAs–GaAlAs lasers. (After Reference [12].)

extent can be expressed in terms of a linear combination of plane waves. The propagation can then be described by the propagation of all these plane wave components. This approach leads to an integral representation of the propagation, known as the *Fresnel–Kirchhoff diffraction integral.* Near the end of this section, we will show that *ABCD* matrices can also be applied to the Fresnel–Kirchhoff diffraction integral for the propagation of a general optical beam in lenslike media.

Referring to Figure 2.12, let us consider the propagation of a general beam of light along the z axis. Let $E(x, y)$ be the amplitude of an optical beam at $z = 0$. The amplitude can be written as a Fourier integral of plane waves,

$$E(x, y) = \iint A(k_x, k_y) \exp(-ik_x x - ik_y y)\, dk_x\, dk_y \quad (z = 0) \tag{2.12-1}$$

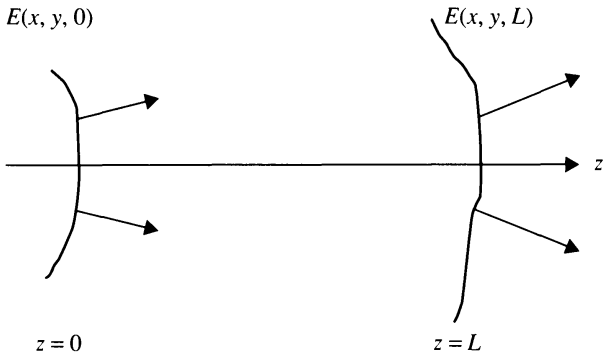


Figure 2.12 Propagation of a general beam from $z = 0$ to $z = L$.

where $A(k_x, k_y)$ is the amplitude of the (k_x, k_y) plane wave component. We assume that the beam is propagating along the z axis so that the amplitude $A(k_x, k_y)$ is centered around (0, 0) in the (k_x, k_y) plane. The amplitude distribution $A(k_x, k_y)$ is often called the Fourier transform of $E(x, y)$. Given the Fourier transform $A(k_x, k_y)$, the field amplitude at $z > 0$ can be written

$$E(x, y, z) = \iint A(k_x, k_y) \exp(-ik_x x - ik_y y) \exp(-ik_z z)\, dk_x\, dk_y \tag{2.12-2}$$

where k_z is the z component of the wavevector. This equation can be viewed as a linear summation of all the plane wave components. The components of the wavevector must satisfy

$$k_x^2 + k_y^2 + k_z^2 = \left(\frac{\omega}{c} n\right)^2 \equiv k^2 \tag{2.12-3}$$

where n is the index of refraction, ω is the angular frequency, and k is the wavenumber of light in the medium. For a well-confined beam or a paraxial wave, the Fourier transform $A(k_x, k_y)$ is significant only in the regime where

$$k_x, k_y \ll k \tag{2.12-4}$$

In this case, the z component k_z can be written

$$k_z = k - \frac{k_x^2 + k_y^2}{2k} \tag{2.12-5}$$

Substituting Equation (2.12-5) into Equation (2.12-2), we obtain

$$E(x, y, z) = \iint A(k_x, k_y)\, e^{-ikz + i[(k_x^2 + k_y^2)/2k]z} \exp(-ik_x x - ik_y y)\, dk_x\, dk_y \tag{2.12-6}$$

By comparing with Equation (2.12-1), we note that the Fourier transform of $E(x, y, z)$ can be written

$$A'(k_x, k_y) = H(k_x, k_y) A(k_x, k_y) = e^{-ikz + i[(k_x^2 + k_y^2)/2k]z} A(k_x, k_y) \tag{2.12-7}$$

where H is known as the *transfer function*. The transfer function for propagation over a distance z in space is thus

$$H(k_x, k_y) = e^{-ikz + i[(k_x^2 + k_y^2)/2k]z} \tag{2.12-8}$$

By inverting Equation (2.12-1), the Fourier transform of $E(x, y)$ can be written

$$A(k_x, k_y) = \frac{1}{(2\pi)^2} \iint E(x, y) \exp(ik_x x + ik_y y)\, dx\, dy \tag{2.12-9}$$

To find the field amplitude at $z > 0$, we can substitute the Fourier transform of Equation (2.12-9) into Equation (2.12-6). We obtain

$$E(x, y, z) = \frac{1}{(2\pi)^2} \iint dx'\, dy' \iint E(x', y') e^{-ik_x(x-x')-ik_y(y-y')} e^{-ikz+i[(k_x^2+k_y^2)/2k]z}\, dk_x\, dk_y \quad (2.12\text{-}10)$$

We now carry out the integration over k_x and k_y. This leads to

$$E(x, y, z) = \frac{i}{\lambda z} e^{-ikz} \iint E(x', y') e^{-ik[(x-x')^2+(y-y')^2]/2z}\, dx'\, dy' \quad (2.12\text{-}11)$$

This is known as the Fresnel–Kirchhoff diffraction integral and is often written

$$E(x, y, L) = \frac{i}{\lambda L} e^{-ikL} \iint E(x', y') e^{-ik[(x-x')^2+(y-y')^2]/2L}\, dx'\, dy' \quad (2.12\text{-}12)$$

where L is the distance of propagation from $z = 0$. Notice that this equation is consistent with the *Huygens principle* in which the field at point (x, y) is a sum of all spherical waves emitted from source point (x', y'). In arriving at the above equations, we have used the following integral:

$$\int_{-\infty}^{\infty} \exp(-\alpha x^2 - \beta x)\, dx = \sqrt{\frac{\pi}{\alpha}} \exp\left(\frac{\beta^2}{4\alpha}\right) \quad (2.12\text{-}13)$$

Thus, given a field of $E(x', y')$ at the input plane $z = 0$, the field $E(x, y, L)$ at the output plane $z = L$ can be obtained by carrying out the integral (2.12-12) over the source plane (x', y').

EXAMPLE: FRESNEL–KIRCHHOFF DIFFRACTION

Consider an optical beam with a Gaussian beam profile at $z = 0$. In other words, the field amplitude can be written

$$E(x, y, 0) = E_0 \exp[-(x^2 + y^2)/a^2]$$

where a is a constant. Physically, a is a measure of the beam size.

Substituting the above expression into Equation (2.12-11), and carrying out the integration with the help of Equation (2.12-13), we obtain

$$E(x, y, z) = \frac{E_0 e^{-ikz}}{1 - i\dfrac{\lambda z}{\pi a^2 n}} \exp\left(-i\frac{k}{2R}(x^2 + y^2)\right) \exp\left(-\frac{x^2 + y^2}{\omega^2}\right)$$

with

$$R = z + \frac{1}{z}\left(\frac{\pi a^2 n}{\lambda}\right)^2, \quad \omega^2 = a^2\left[1 + \left(\frac{z\lambda}{\pi a^2 n}\right)^2\right]$$

We note that these results are in agreement with the propagation of a Gaussian beam.

The Fresnel–Kirchhoff diffraction integral (2.12-12) is now reproduced here with a slight change of notation:

$$f_1(x_1, y_1) = \frac{ik}{2\pi L} \exp(-ikL) \iint f_0(x_0, y_0) \exp\left(-\frac{ik}{2L}[(x_1 - x_0)^2 + (y_1 - y_0)^2]\right) dx_0\, dy_0 \quad (2.12\text{-}14)$$

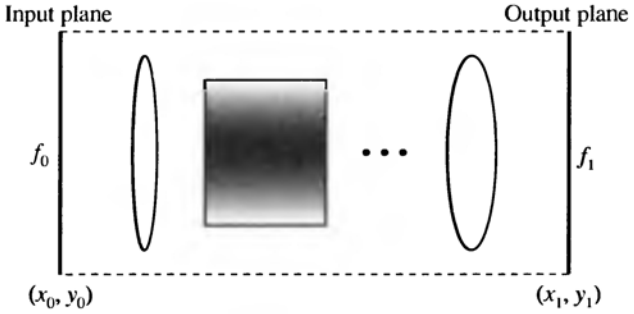


Figure 2.13 A cascade of lenslike (paraxial) elements with an overall *ABCD* matrix.

which relates the field $f_1(x_1, y_1)$ at $z = L$ to the source field $f_0(x_0, y_0)$ at $z = 0$ in the case of a homogeneous and isotropic medium. Equation (2.12-14) is often written in the following form:

$$f_1(x_1, y_1) = \frac{ik}{2\pi L}\iint f_0(x_0, y_0) \exp[-ik\rho(x_0, y_0; x_1, y_1)]\, dx_0\, dy_0 \tag{2.12-15}$$

where

$$\rho(x_0, y_0; x_1, y_1) = L + \frac{1}{2L}(x_1 - x_0)^2 + \frac{1}{2L}(y_1 - y_0)^2 \tag{2.12-16}$$

is the distance between the source point (x_0, y_0) and the observation point (x_1, y_1) in free space. We recall that the above is valid provided $\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} \ll L$. This result can be extended to the case where the intervening medium consists of a cascade of paraxial and axially symmetric components, which can be described by an overall *ABCD* matrix. An example of such a system is sketched in Figure 2.13.

The result is

$$f_1(x_1, y_1) = \frac{ik_0}{2\pi B}\exp(-ik_0 L)\iint f_0(x_0, y_0) \exp\left(-\frac{ik_0}{2B}[A(x_0^2 + y_0^2) - 2x_0x_1 - 2y_0y_1 + D(x_1^2 + y_1^2)]\right) dx_0\, dy_0 \tag{2.12-17}$$

$$k_0 = \frac{2\pi}{\lambda} = \frac{\omega}{c} = \text{wavenumber in vacuum}$$

$$L = \sum_i n_i L_i = \text{opticla path length along the axis}$$

This important result can be proved as follows by using Fermat's principle of least time [3, 13–14]. Referring to Figure 2.14, we consider propagation through an optical system that can be represented by an overall *ABCD* matrix. Let X_0 and X_1 be general points at the input plane and output plane, respectively. Without loss of generality, we further assume that the medium before the input and the medium after the output are vacuum with a unity refractive index. Let the optical path between X_0 and X_1 be written $\rho(x_0, x_1)$. We extrapolate the rays at points X_0 and X_1. This leads to intersections with the optical axis at P_0 and P_1.

According to Fermat's principle, the optical path $P_0Q_0Q_1P_1$ along the axis should be the same as that of $P_0X_0X_1P_1$. Let the radii of curvature be

$$\begin{aligned} R_0 &= P_0Q_0 \\ R_1 &= Q_1P_1 \end{aligned} \tag{2.12-18}$$

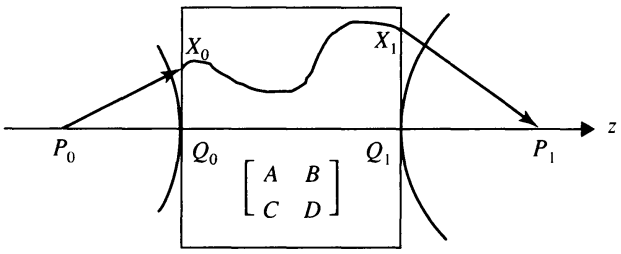


Figure 2.14 Schematic drawing showing the propagation of a beam through a general optical system described by a ray matrix with elements A, B, C, and D. This figure is used in the proof of the Fresnel–Kirchhoff integral (2.12-17) using Fermat's principle.

then the optical paths can be written

$$P_0Q_0Q_1P_1 = R_0 + \sum_i n_iL_i + R_1 \tag{2.12-19}$$

$$P_0X_0X_1P_1 = R_0 + \frac{x_0^2}{2R_0} + \rho(x_0, x_1) + R_1 + \frac{x_1^2}{2R_1} \tag{2.12-20}$$

By using the above equations and $P_0Q_0Q_1P_1 = P_0X_0X_1P_1$, we obtain

$$\rho(x_0, x_1) = \sum_i n_iL_i - \frac{x_0^2}{2R_0} - \frac{x_1^2}{2R_1} \tag{2.12-21}$$

where R_0 and R_1 are related to the ray slopes as

$$x_0' = \frac{x_0}{R_0} \quad \text{and} \quad x_1' = \frac{x_1}{R_1} \tag{2.12-22}$$

Furthermore, the ray parameters (including the slopes) can be written

$$\begin{aligned} x_1 &= Ax_0 + Bx_0' \\ x_1' &= Cx_0 + Dx_0' \end{aligned} \tag{2.12-23}$$

By eliminating R_0, R_1, and the ray slopes, we obtain

$$\rho(x_0, x_1) = \sum_i n_iL_i + \frac{1}{2B}(Ax_0^2 + Dx_1^2 - 2x_0x_1) \tag{2.12-24}$$

The extension to the general case of from point (x_0, y_0) to point (x_1, y_1) leads to

$$\rho(x_0, y_0; x_1, y_1) = \sum_i n_iL_i + \frac{1}{2B}[A(x_0^2 + y_0^2) + D(x_1^2 + y_1^2) - 2x_0x_1 - 2y_0y_1] \tag{2.12-25}$$

Substituting Equation (2.12-25) into Equation (2.12-15), we obtain the general expression (2.12-17). Note that the factor $(ik/2\pi L)$ is replaced by $(ik/2\pi B)$. This is essential to conserve power and to make the general result agree with the free-space result when the *ABCD* system consists only of free space. The matrix element B plays the same role as the distance of propagation $L = z_1 - z_0$. Thus an arbitrary optical wave can propagate through a sequence of paraxial optical systems, including all diffraction effects, by using knowledge only of the overall *ABCD* coefficients of the optical system between the input and output plane.

To demonstrate the power of this general result (2.12-17), we will use it to derive the imaging condition of a generalized cascade of optical elements described by an overall *ABCD* matrix. By rearranging and manipulating the exponent of Equation (2.12-17), we can rewrite it as

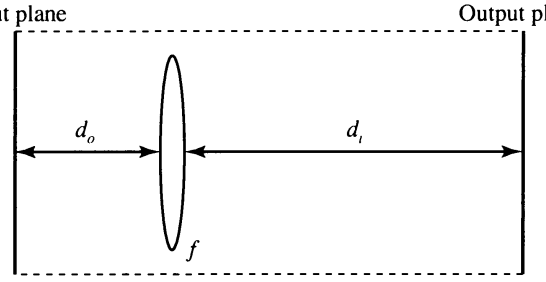


Figure 2.15 A single lens imaging system.

$$f_1(x_1, y_1) = \frac{ik_0}{2\pi B}\exp\left[-ik_0 L - \frac{ik_0}{2B}\left(D - \frac{1}{A}\right)(x_1^2 + y_1^2)\right]$$

$$\times \iint f_0(x_0, y_0)\exp\left\{-\frac{ik_0}{2B}\left[A\left(x_0 - \frac{x_1}{A}\right)^2 + A\left(y_0 - \frac{y_1}{A}\right)^2\right]\right\} dx_0\, dy_0 \quad (2.12\text{-}26)$$

It is a straightforward matter to show [15] that

$$\lim_{B\to 0}\sqrt{\frac{i}{2\pi B}}\exp\left(-i\frac{x^2}{2B}\right) = \delta(x) \quad (2.12\text{-}27)$$

where $\delta(x)$ is the Dirac delta function. Using Equation (2.12-27) in (2.12-26), we obtain

$$\lim_{B\to 0} f_1(x_1, y_1) = \frac{\exp(-ik_0 L)}{A} f_0\left(\frac{x_1}{A}, \frac{y_1}{A}\right)\exp\left(-i\frac{k_0(DA-1)}{2AB}(x_1^2 + y_1^2)\right) \quad (2.12\text{-}28)$$

Using the property $AD - BC = 1$, which holds when the individual A, B, C, D matrices of the cascade possess unity determinants, we rewrite Equation (2.12-28) as

$$\lim_{B\to 0} f_1(x_1, y_1) = \frac{\exp(-ik_0 L)}{A}\exp\left(-i\frac{k_0 C}{2A}(x_1^2 + y_1^2)\right) \times f_0\left(\frac{x_1}{A}, \frac{y_1}{A}\right) \quad (2.12\text{-}29)$$

This is the main result. It shows that in the limit of $B \to 0$, $f_1(x_1, y_1)$ is magnified by a factor A to $f_0(x_1/A, y_1/A)$ as well as multiplied by a quadratic phase factor (in x_1 and y_1). We thus associate $B = 0$ with the generalized imaging condition and A with the magnification. We leave it as a (simple) exercise to prove that in the case of imaging by a single thin lens, as shown in Figure 2.15, the imaging condition $B \to 0$ leads to the familiar lens formula

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i} \quad (2.12\text{-}30)$$

while the magnification is $A = -d_i/d_o$, where d_o and d_i are positions of the object and image, respectively.

The Fraunhofer Approximation

The diffraction integral (2.12-14) can be further simplified in the so-called Fraunhofer regime, where

$$k(x_0^2 + y_0^2) \ll 2L \quad (2.12\text{-}31)$$

for all points (x_0, y_0) of the source field. In this case, the quadratic phase function $\exp[-ik(x_0^2 + y_0^2)/2L]$ is approximately unity over the entire source field (aperture). The Fresnel–Kirchoff integral becomes, according to Equation (2.12-14),

$$f_1(x_1, y_1) = \frac{ik}{2\pi L}\exp[-ikL - ik(x_1^2 + y_1^2)/2L]\iint f_0(x_0, y_0)\exp\left(\frac{ik}{L}(x_1x_0 + y_1y_0)\right)dx_0\, dy_0 \quad (2.12\text{-}32)$$

We note that the field distribution $f_1(x_1, y_1)$ at $z = L$ is proportional to the Fourier transform of the source field distribution $f_0(x_0, y_0)$, provided condition (2.12-31) is satisfied. This is the regime of the Fraunhofer diffraction. The spatial region defined by Equation (2.12-31) is often called the far field. Thus, aside from a multiplicative phase factor, the field distribution in the far field is simply the Fourier transform of the source field.

EXAMPLE: FAR FIELD OF A CIRCLE APERTURE

Consider the passage of light with $\lambda = 1$ μm through a circular aperture of radius $a = 1$ cm. The far field according to Equation (2.12-31) is

$$L \gg k(x_0^2 + y_0^2)_{\max}/2 = \pi a^2/\lambda = 314 \text{ m}$$

In the situation when the far field is indeed very far from the aperture, the far-field distribution can often be observed with the help of an imaging lens.

PROBLEMS

2.1 The unit cell matrix of a periodic lens waveguide is a unimodular matrix ($AD - BC = 1$). Solve the following eigenvalue problem:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} r \\ r' \end{bmatrix} = \lambda\begin{bmatrix} r \\ r' \end{bmatrix}$$

(a) Show that the eigenvalues λ of the equation are $\lambda = e^{\pm i\theta}$ with θ given by Equation (2.1-13).

(b) Show that the eigenvectors are

$$\begin{bmatrix} r \\ r' \end{bmatrix}_1 = \begin{bmatrix} B \\ e^{i\theta} - A \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} r \\ r' \end{bmatrix}_2 = \begin{bmatrix} B \\ e^{-i\theta} - A \end{bmatrix}$$

These two eigenvectors are independent provided B is finite and θ is not an integral multiple of π (i.e., $\theta \neq m\pi$). Although they are independent, show that these two eigenvectors are, in general, not orthogonal.

(c) A matrix M can be found such that the ray matrix is diagonalized,

$$M^{-1}\begin{bmatrix} A & B \\ C & D \end{bmatrix}M = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}$$

Show that M can be written

$$M = \begin{bmatrix} B & B \\ e^{i\theta} - A & e^{-i\theta} - A \end{bmatrix}$$

The determinant of matrix M is $-2iB \sin\theta$. The inverse matrix exists provided B is finite and θ is not an integral multiple of π (i.e., $\theta \neq m\pi$). Find the inverse matrix M^{-1}.

2.2 Given two independent eigenvectors, it is possible to express an arbitrary ray vector in terms of a linear combination of the eigenvectors. In other words, we can write

$$\begin{bmatrix} r_0 \\ r_0' \end{bmatrix} = \alpha_0\begin{bmatrix} B \\ e^{i\theta} - A \end{bmatrix} + \beta_0\begin{bmatrix} B \\ e^{-i\theta} - A \end{bmatrix}$$

where α_0 and β_0 are constants.

(a) Find α_0 and β_0 in terms of r_0 and r_0'. Since both r_0 and r_0' are real, α_0 and β_0 form a complex conjugate pair.

(b) Show that the ray vector at the mth period can be written

$$\begin{bmatrix} r_m \\ r'_m \end{bmatrix} = \alpha_0 e^{im\theta} \begin{bmatrix} B \\ e^{i\theta} - A \end{bmatrix} + \beta_0 e^{-im\theta} \begin{bmatrix} B \\ e^{-i\theta} - A \end{bmatrix}$$

and show that a general expression for the ray location can be written in the form (2.1-23):

$$r_m = r_{\max} \sin(m\theta + \alpha)$$

(c) Derive Equation (2.1-23) directly from Equations (2.1-10) and (2.1-11).
(d) Derive Equations (2.1-24) through (2.1-26).

2.3 Derive Equation (2.3-19).

2.4 Consider a thin convex lens made of a homogeneous medium with spherical surfaces. Let the radii of curvatures of the surfaces be R_1 and R_2. Show that for a plane wave incident on a thin lens, the input–output relationship can be written

$$E_{\text{out}}(x, y) = E_{\text{in}}(x, y) \exp\left(ik \frac{x^2 + y^2}{2f}\right) \exp(-i\phi)$$

where f is the focal length of the thin lens. Show that f is given by

$$\frac{1}{f} = (n - 1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right)$$

What is the physical meaning of ϕ?

2.5 Show that a lenslike medium occupying the region $0 \le z \le l$ will image a point on the axis at $z < 0$ onto a single point. (If the image point occurs at $z < l$, the image is virtual.)

2.6 Derive the ray matrices of Table 2.1.

2.7 For confined Gaussian beams, the complex beam parameters q must be in the upper complex plane.

(a) Using the *ABCD* law of Gaussian beam transformation (2.6-9),

$$q_2 = \frac{Aq_1 + B}{Cq_1 + D}$$

Show that if q_1 is in the upper complex plane, so is q_2, provided A, B, C, and D are all real and $AD - BC = 1$. [Hint: Rationalize the denominator.]
(b) Using (a), show that q_N is always in the lower plane, if q_0 is in the lower complex plane.

$$q_N = \frac{\{A \sin(N\theta) - \sin[(N-1)\theta]\}q_0 + B \sin(N\theta)}{C \sin(N\theta)q_0 + D \sin(N\theta) - \sin[(N-1)\theta]}$$

provided θ is real.

(c) Show that for $|\cos\theta| > 1$, q_N approaches a real number when N becomes infinity. [Hint: First show the existence of a limit. Then show that the limit must satisfy $q = (Aq + B)/(Cq + D)$.]

2.8 Consider the incidence of a Gaussian beam normally on a solid prism with an index of refraction n (as shown in Figure P2.8).

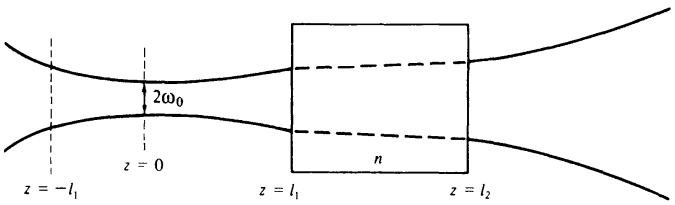


Figure P2.8

(a) What is the far-field diffraction angle of the output beam?
(b) Assume that the prism is moved to the left until its input face is at $z = -l_1$. What is the new beam waist and what is its location? (Assume that the crystal is long enough that the beam waist is inside the crystal.)

2.9 A Gaussian beam with a waist located at $z = 0$, and with a wavelength λ, is incident on a lens placed at $z = l$ as shown in Figure P2.9. Find the lens focal length, f, so that the output beam has a waist located at a given location $z = L$. Show that (given l and L) two solutions for the focal length exist. Sketch the beam behavior for each of these solutions.

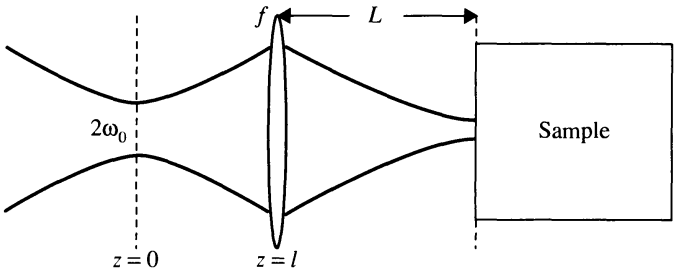


Figure P2.9

2.10 Complete all the missing steps in the derivation of Section 2.11.

2.11 Find the beam spot size and the maximum number of pulses per second that can be carried by an optical beam ($\lambda = 1$ μm) propagating in a quadratic index glass fiber with $n = 1.5$, $g^2 = 3.3 \times 10^2$ cm^{-2}: (a) in the case of a single-mode excitation $l = m = 0$ and (b) in the case

where all the modes with $l, m < 5$ are excited. Using dispersion data of any typical commercial glass and taking $g^2 = 3.3 \times 10^2$ cm^{-2}, $l_{max} = m_{max} = 30$, compare the relative contributions of modal and glass dispersion to pulse broadening.

2.12 Consider a thick lens with radii of curvature R_1 and R_2 on its entrance and exit surfaces, an index of refraction n, and a thickness d.

(a) Obtain the *ABCD* matrix of the lens.
(b) What is its focal distance for light incident from the left?

2.13 Show that

$$\lim_{B\to 0} \sqrt{\frac{i}{2\pi B}} \exp\left(-i\frac{x^2}{2B}\right) = \delta(x)$$

where $\delta(x)$ is the Dirac δ function.

2.14 Show that, in the case of imaging by a thin lens, the generalized imaging condition $B = 0$ leads to Equation (2.12-30).

2.15 Show that Snell's law is consistent with Fermat's principle of least time.

2.16 Derive Snell's law directly from the ray equation.

2.17 The index of refraction of the atmosphere is given approximately as

$$n = 1 + \Delta n \exp[-\alpha(r - R)], \quad R < r$$

where Δn is approximately 0.0003, r is the distance measured from the center of the Earth, R is the radius of Earth, and α is approximately 1/7.8 km. For sunset and sunrise, the rays are nearly horizontal.

(a) Define β as the angle between the tangent vector of the ray and the horizon (x axis). Show that the ray equation can be written approximately

$$\frac{d}{dx}(n\beta) = \hat{\mathbf{y}} \cdot \nabla n = \frac{dn}{dr}\cos\theta$$

where r is the radial coordinate of the ray and y is the vertical axis with $r^2 = R^2 + x^2$.

(b) Define $r = R + z$, with R = Earth's radius. The ray bending can be obtained by integrating the above equation from $z = 0$ to infinity. Show that

$$\beta_\infty = \int_0^\infty \left(\frac{dn}{dr}\right) \frac{R}{\sqrt{(R+z)^2 - R^2}}\, dz$$

(c) Show that

$$\beta_\infty = -\Delta n \sqrt{\frac{\alpha R \pi}{2}}$$

[Hint: Neglect the z^2 term in the denominator ($z^2 \ll 2Rz$). This is legitimate as $R = 6400$ km and (dn/dr) is almost zero for altitudes above 100 km.] Using $\alpha = 1/7.8$ km, $R = 6400$ km, and $\Delta n = 0.0003$, $\beta_\infty = 0.011 = 0.6°$.

2.18

(a) Show that the inverse matrix M^{-1} of Problem 2.1 can be written

$$M^{-1} = \frac{i}{2B\sin\theta}\begin{bmatrix} e^{-i\theta} - A & -B \\ A - e^{i\theta} & B \end{bmatrix}$$

(b) Show that

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^N = M\begin{bmatrix} e^{iN\theta} & 0 \\ 0 & e^{-iN\theta} \end{bmatrix} M^{-1}$$

(c) Carry out the matrix multiplications and derive the matrix identity (2.1-11).

2.19 Rays that are not confined to meridional planes exist in lenslike media, described by Equation (2.3-10). These rays are the so-called skew rays.

(a) Starting from the ray equation (2.3-8) or (2.3-11), show that rays with helical paths exist. In other words, show that the solution

$$r = a, \quad \theta = bz$$

satisfies the ray equation, where a and b are constants. Note that the helical pitch is $2\pi/b$.

(b) Show that b is related to the helical radius a by

$$b^2 = \frac{2n_0^2 g^2}{n^2 - n_0^2 g^2 a^2} = \frac{2n_0^2 g^2}{n_0^2 - 2n_0^2 g^2 a^2}$$

where n is the index of refraction at $r = a$.

(c) Show that circular paths also exist. Show that the radius of such a circular path satisfies the equation

$$n^2 = n_0^2 g^2 r^2$$

Note that the circular path may be viewed as a degenerate helix with zero pitch or an infinite b.

2.20 The lenslike medium described by Equation (2.3-9) can be shown to be a limiting case of the periodic lens waveguide consisting of identical lenses separated by an equal distance of d. Let $z = Nd$. Consider the limit when $d \to 0$ and $f \to \infty$ such that $df = 1/g^2$.

(a) Show that $\theta \to \sqrt{d/f}$ and $N\theta \to gz$.

(b) Using $A = 1$, $B = d$, $C = -1/f$, and $D = 1 - d/f$, show that in the limit when $N \to \infty$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^N = \begin{bmatrix} \cos(gz) & \sin(gz)/g \\ -g\sin(gz) & \cos(gz) \end{bmatrix}$$

So the lenslike medium can be viewed as a distributed case of a lens waveguide.

2.21 For the following problems, do the cases with $p = 0, 1$ and $m = 0, 1, 2$.

(a) Plot the intensity pattern of the first few low-order modes of Laguerre–Gaussian beams.

(b) Plot the sum of (p, m) and $(p, -m)$ modes.

(c) Express the first few low-order modes of Laguerre–Gaussian beams in terms of the linear combination of Hermite–Gaussian beams.

2.22

(a) Using the fundamental Gaussian beam in homogeneous media as an example, show that

$$k(\partial\psi/\partial z) = \left[-\frac{k}{q} + \frac{ik(kr^2/2)}{q^2}\right]\psi,$$

$$\partial^2\psi/\partial z^2 = \left[\frac{2}{q^2} - \frac{4i(kr^2/2)}{q^3} - \frac{(kr^2/2)^2}{q^4}\right]\psi$$

(b) Show that within the Gaussian beam, $(kr^2/2q) \leq 1$, the SVA is justified provided

$$\omega_0 \gg \lambda$$

that is, the beam spot size is much bigger than the wavelength.

(c) Show that the SVA approximation is equivalent to paraxial ray approximation.

REFERENCES

1. Marcuse, D., *Light Transmission Optics*. Van Nostrand, Princeton, NJ, 1972.
2. Yariv, A., *Quantum Electronics*, 2nd ed. Wiley, New York, 1975, Section 2.2.
3. Siegman, A. E., *Lasers*, University Science Books, Mill Valley, CA, 1986, pp. 779–782.
4. Kogelnik, H., and W. Rigrod, *Proc. IRE* **50**:230 (1962).
5. Kawakami, S., and J. Nishizawa, An optical waveguide with the optimum distribution of the refractive index with reference to waveform distortion. *IEEE Trans. Microwave Theory Technique* **10**:814 (1968).
6. Miller, S. E., E. A. J. Marcatili, and T. Li, Research toward optical fiber transmission systems. *Proc. IEEE* **61**:1703 (1973).
7. Cohen, L. G., and H. M. Presby, Shuttle pulse measurement of pulse spreading in a low loss graded index fiber. *Appl. Opt.* **14**:1361 (1975).
8. Bloom, D. M., L. F. Mollenauer, Chinlon Lin, D. W. Taylor, and A. M. DelGaudio, Direct demonstration of distortionless picosecond-pulse propagation in kilometer-length optical fibers. *Opt. Lett.* **4**:297 (1979).
9. Suematsu, Y., Long wavelength optical fiber communication. *Proc. IEEE* **71**:692 (1983).
10. Bennett, W. R., Inversion mechanisms in gas lasers. *Appl. Opt. Suppl. Chem. Lasers* **2**:3 (1965).
11. Casperson, L., and A. Yariv, The Gaussian mode in optical resonators with a radial gain profile. *Appl. Phys. Lett.* **12**:355 (1968).
12. Zachos, T. H., Gaussian beams from GaAs junction lasers. *Appl. Phys. Lett.* **12**:318 (1969).
13. Baues, P., Huygens' principle in inhomogeneous isotropic media. *Optoelectronics* **1**:37 (1969).
14. Collins, S. A., Lens-system diffraction integral written in terms of matrix optics. *J. Opt. Soc. Am.* **60**:1168 (1970).
15. Yariv, A., Imaging of coherent fields through lenslike systems. *Opt. Lett.* **19**:1607 (1994).

CHAPTER

GUIDED WAVES IN DIELECTRIC SLABS AND FIBERS

3.0 INTRODUCTION

So far, we have discussed wave propagation in free space, including the propagation of plane waves and beams. As a result of diffraction, a beam of light with a finite cross section will spread as it propagates in free space. Lenses or gradient index media may be employed at appropriate locations to focus the beams. Generally, dielectric media of high refractive index can be employed to confine the propagation of the beam. In this chapter, we will show that both dielectric slabs and circular fibers can support confined electromagnetic propagation. These modes of propagation are the so-called guided waves (or guided modes), and the structures that support guided waves are called waveguides. In this chapter, we first discuss the propagation of guided waves in dielectric slabs. As we know, any light beam with a finite transverse dimension will diverge as it propagates in a homogeneous medium. This divergence disappears in guiding dielectric structures under the appropriate conditions. In dielectric waveguides, the transverse dimension of these modes of propagation is determined by the dielectric waveguide.

We shall derive first the properties of guided modes in a dielectric slab structure. Optical modes are presented as the solution of the eigenvalue equation, which is derived from Maxwell's equations subject to the boundary conditions imposed by waveguide geometry. Both transverse electric (TE) and transverse magnetic (TM) modes of propagation are derived. The physics of confined propagation is explained in terms of the total internal reflection of plane waves from the dielectric interfaces. After discussion of the slab waveguides, we will cover the important subject of guided waves in circular fibers. We will then introduce the simple theory of effective index, which is useful in understanding waveguiding in two-dimensional structures. In the last part of the chapter, we will take up the subject of signal corruption due to chromatic dispersion in optical fibers, and the signal attenuation due to scattering and absorption. The subject of coupling between waveguides and the method of dispersion compensation in fiber transmission will be discussed later in this book.

3.1 TE AND TM CONFINED MODES IN SYMMETRIC SLAB WAVEGUIDES

Dielectric slabs are the simplest optical waveguides. Figure 3.1 shows a typical example of a slab waveguide. It consists of a thin dielectric layer (called the guiding layer, or simply the core)

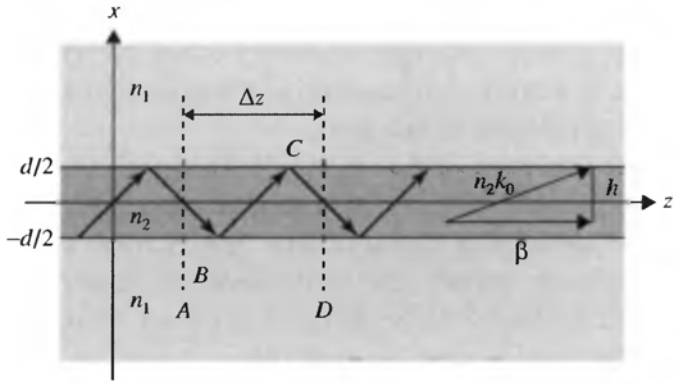


Figure 3.1 Schematic drawing of a symmetric slab waveguide. The waveguide consists of a guiding layer of thickness d with a refractive index n_2, surrounded by media of refractive index $n_1 \cdot k_0 = \omega/c$.

sandwiched between two semi-infinite bounding media (clad). Generally, the index of refraction of the guiding layer must be greater than those of the surrounding media. In addition, the thickness of the guiding layer is typically on the order of a wavelength. In symmetric slab waveguides, the two bounding media are identical. The simplest example will be a thin glass film (or layer) immersed in air or another fluid (or solid) of a lower index of refraction.

The following equation describes the index profile of a symmetric dielectric waveguide:

$$n(x) = \begin{cases} n_2, & |x| < d/2 \\ n_1, & \text{otherwise} \end{cases} \tag{3.1-1}$$

where d is the thickness of the guiding layer (core), n_2 is the index of refraction of the core, and n_1 is the index of refraction of the bounding media. To support guided modes, n_2 must be greater than n_1. The problem at hand is to find these guided modes.

The electromagnetic treatment of such a problem is relatively easy because the medium is homogeneous in each segment of the dielectric structure. In addition, the solutions of Maxwell's equations in homogeneous media are simply plane waves. Thus all we need to do is write the plane wave solutions for each segment and then match the boundary conditions at the interfaces to ensure the continuity of the wavefunction. We now consider the propagation of monochromatic radiation along the z axis. Maxwell's equation can be written in the form

$$\nabla \times \mathbf{H} = i\omega\varepsilon_0 n^2 \mathbf{E}, \qquad \nabla \times \mathbf{E} = -i\omega\mu\mathbf{H} \tag{3.1-2}$$

where n is the refractive index profile given in Equation (3.1-1). Since the whole structure is homogeneous along the z axis, solutions to the wave equations (3.1-2) can be taken as

$$\begin{aligned} \mathbf{E}(x, t) &= \mathbf{E}_m(x) \exp[i(\omega t - \beta z)] \\ \mathbf{H}(x, t) &= \mathbf{H}_m(x) \exp[i(\omega t - \beta z)] \end{aligned} \tag{3.1-3}$$

where β is the z component of the wavevectors and is known as the propagation constant to be determined from Maxwell's equations, and $\mathbf{E}_m(x)$ and $\mathbf{H}_m(x)$ are wavefunctions of the guided modes, with subscript m being an integer (called the mode number). For a layered dielectric structure that consists of homogeneous and isotropic materials, the wave equation can be obtained by eliminating $\mathbf{H}$ from Equartion (3.1-2):

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\mathbf{E}(x, y) + [k_0^2 n^2(r) - \beta^2]\mathbf{E}(x, y) = 0 \tag{3.1-4}$$

where $k_0 = \omega/c$ is the wavenumber in vacuum. Note that the preceding equation is not valid at the dielectric interfaces for TM waves (the $\mathbf{H}$ vector is transverse to the xz plane). We will solve for this equation separately in each segment of the structure and then match the

tangential components of the field vectors at each interface. In addition to the continuity conditions at the interfaces, another important boundary condition for guided modes is that the field amplitudes be zero at infinity. A discussion of the general properties of the solutions of Equation (3.1-4) will be given later in this book.

We now proceed with the solution of Equation (3.1-4) in each segment of the dielectric structure. The propagation constant β is a very important parameter because it determines whether the field varies sinusoidally or exponentially. For confined modes, the field amplitude must fall off exponentially outside the guide structure. Consequently, the quantity $(n\omega/c)^2 - \beta^2$ in Equation (3.1-4) must be negative for $|x| > \frac{1}{2}d$. In other words, the propagation constant β of a confined mode must be such that

$$\beta > \frac{n_1\omega}{c} \tag{3.1-5}$$

where we recall that n_1 is the index of refraction of the bounding media. On the other hand, the continuity of the field requires that the magnitude of the field $\mathbf{E}_m(x)$ attain a maximum value. The existence of a maximum requires that the Laplacian of the field be negative. In other words, the propagation constant of a confined mode must be such that

$$\beta < \frac{n_2\omega}{c} \tag{3.1-6}$$

Thus we will find confined modes whose propagation constant satisfies these conditions, Equations (3.1-5) and (3.1-6). The modes can also be classified as either TE or TM modes. The TE modes have their electric field perpendicular to the xz plane (plane of incidence, or plane of propagation) and thus have only the field components E_y, H_x, and H_z. The TM modes have the field components H_y, E_x, and E_z.

Guided TE Modes

The electric field amplitude of the guided TE modes can be written in the form

$$E_y(x, z, t) = E_m(x) \exp[i(\omega t - \beta z)] \tag{3.1-7}$$

In a manner very similar to the wavefunction of a particle in a square-well potential, the mode function $E_m(x)$ is taken as

$$E_m(x) = \begin{cases} A \sin hx + B \cos hx, & |x| < \frac{1}{2}d \\ C \exp(-qx), & x > \frac{1}{2}d \\ D \exp(qx), & x < -\frac{1}{2}d \end{cases} \tag{3.1-8}$$

where A, B, C, and D are constants, and the parameters h and q are related to the propagation constant by

$$h = \left[\left(\frac{n_2\omega}{c}\right)^2 - \beta^2\right]^{1/2}$$
$$q = \left[\beta^2 - \left(\frac{n_1\omega}{c}\right)^2\right]^{1/2} \tag{3.1-9}$$

The parameter h may be considered as the transverse component of the wavevector in the guiding layer. To be acceptable solutions, the tangential component of the electric and magnetic

fields, E_y and H_z, must be continuous at the interfaces. Since $H_z = (i/\omega\mu)(\partial E_y/\partial x)$, we must match the magnitude as well as the slope of the TE mode functions $E_m(x)$ at the interfaces. This leads to

$$A \sin(\tfrac{1}{2}hd) + B \cos(\tfrac{1}{2}hd) = C \exp(-\tfrac{1}{2}qd)$$

$$hA \cos(\tfrac{1}{2}hd) - hB \sin(\tfrac{1}{2}hd) = -qC \exp(-\tfrac{1}{2}qd)$$

$$-A \sin(\tfrac{1}{2}hd) + B \cos(\tfrac{1}{2}hd) = D \exp(-\tfrac{1}{2}qd)$$

$$hA \cos(\tfrac{1}{2}hd) + hB \sin(\tfrac{1}{2}hd) = qD \exp(-\tfrac{1}{2}qd)$$

from which we obtain

$$2A \sin(\tfrac{1}{2}hd) = (C - D) \exp(-\tfrac{1}{2}qd) \tag{3.1-10}$$

$$2hA \cos(\tfrac{1}{2}hd) = -q(C - D) \exp(-\tfrac{1}{2}qd) \tag{3.1-11}$$

$$2B \cos(\tfrac{1}{2}hd) = (C + D) \exp(-\tfrac{1}{2}qd) \tag{3.1-12}$$

$$2hB \sin(\tfrac{1}{2}hd) = q(C + D) \exp(-\tfrac{1}{2}qd) \tag{3.1-13}$$

By examining the above equations, we find there are two sets of solutions.

(a) Symmetric modes ($A = 0$ and $C = D$): Equations (3.1-12) and (3.1-13) yield

$$h \tan(\tfrac{1}{2}hd) = q \quad \text{(for symmetric TE modes)} \tag{3.1-14}$$

(b) Antisymmetric modes ($B = 0$ and $C = -D$): Equations (3.1-10) and (3.1-11) give

$$h \cot(\tfrac{1}{2}hd) = -q \quad \text{(for antisymmetric TE modes)} \tag{3.1-15}$$

Note that both Equations (3.1-14) and (3.1-15) cannot be satisfied simultaneously since the elimination of q would lead to a pure imaginary h and a negative q. However, these two equations can be combined into a single equation (see Problem 3.14):

$$\tan(hd) = \frac{2hq}{h^2 - q^2} \tag{3.1-16}$$

The solutions of TE modes may thus be divided into two classes. For the first class,

$$A = 0, \quad C = D, \quad h \tan(\tfrac{1}{2}hd) = q \tag{3.1-17}$$

and for the second class,

$$B = 0, \quad C = -D, \quad h \cot(\tfrac{1}{2}hd) = -q \tag{3.1-18}$$

Note that the solutions in the first class have symmetric wavefunctions, whereas those of the second class have antisymmetric wavefunctions.

The propagation constants of the TE modes are found from a numerical or graphical solution of Equations (3.1-17) and (3.1-18), with the definition of h and q given by Equation (3.1-9). A very simple and well-known graphic solution is described here, since it clearly shows the way in which the number of TE modes depends on both the thickness d and the difference of indices of refraction. By putting $u = \frac{1}{2}hd$ and $v = \frac{1}{2}qd$, Equation (3.1-17) becomes $u \tan u = v$, with

$$u^2 + v^2 = (n_2^2 - n_1^2)\left(\frac{\omega d}{2c}\right)^2 = (n_2^2 - n_1^2)\left(\frac{\pi d}{\lambda}\right)^2 \equiv V^2 \tag{3.1-19}$$

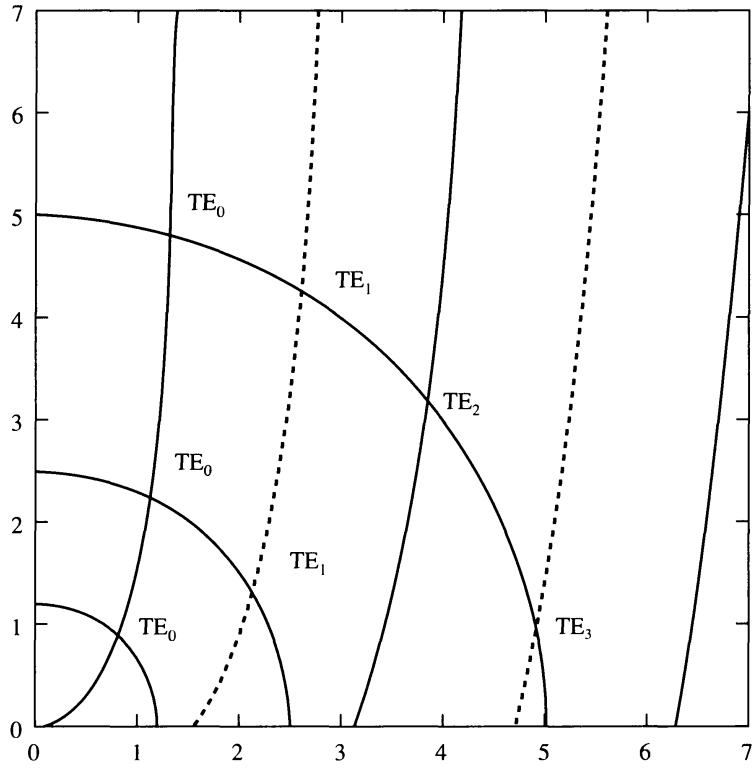


Figure 3.2 Graphic solution of Equations (3.1-17) and (3.1-18) for three values of V. Solid curves are $v = u \tan u$, and the dotted curves are $v = -u \cot u$.

Since u and v are restricted to positive values, the propagation constants may be found in this case from the intersection of both the curve $v = u \tan u$ and a circle of known radius $V = (n_2^2 - n_1^2)^{1/2}(\pi d/\lambda)$ in the first quadrant of the uv plane. A similar graphic construction for the solution of Equation (3.1-18) can be obtained by plotting $v = -u \cot u$ and the circle on the uv plane. Figure 3.2 shows such a graphic method for three values of V. For $V = 1.2$, there is only one solution—the TE_0 mode. There are two solutions (TE_0 and TE_1) when $V = 2.5$ and four solutions when $V = 5$. Note that the number of solutions depends on the value of V.

From Figure 3.2, it is clear that the number of confined TE modes depends on the magnitude of the parameter V. For V between zero and $\frac{1}{2}\pi$, there is just one TE mode of the first class. The first mode of the second class appears when the parameter V is greater than $\frac{1}{2}\pi$. As this parameter V increases, confined modes appear successively, first of one class and then of the other. Figure 3.3 plots the wavefunctions of a symmetrical slab waveguide with $n_2 = 1.6$, $n_1 = 1.5$, $d = 5$ μm, and $\lambda = 1.55$ μm. According to Equation (3.1-19), the parameter $V = 5.64$. This waveguide supports four TE modes. It is not difficult to see from Figure 3.3 that, when ordered according to the propagation constant β, the mth wavefunction has $m - 1$ nodes. We also notice that the wavefunctions are either symmetric or antisymmetric with respect to the origin $x = 0$. It follows from the discussion earlier that the wavefunctions are divided into two classes (see Equations (3.1-17) and (3.1-18)). This division is a direct consequence of the fact that the index profile $n(x)$ is symmetric about $x = 0$.

Knowledge that the solution possesses a definite symmetry sometimes simplifies the determination of the propagation constant, since we need only find the solution for positive x. Even solutions have zero slope and odd solutions have zero value at the origin $x = 0$. Thus the wavefunction of the even solutions can be written as $\cos(hx)$, whereas those of the odd solutions can be written as $\sin(hx)$. Both types of solutions decay exponentially in the region $|x| > \frac{1}{2}d$. The solutions are then obtained by matching the value and the slope at $|x| = \frac{1}{2}d$.

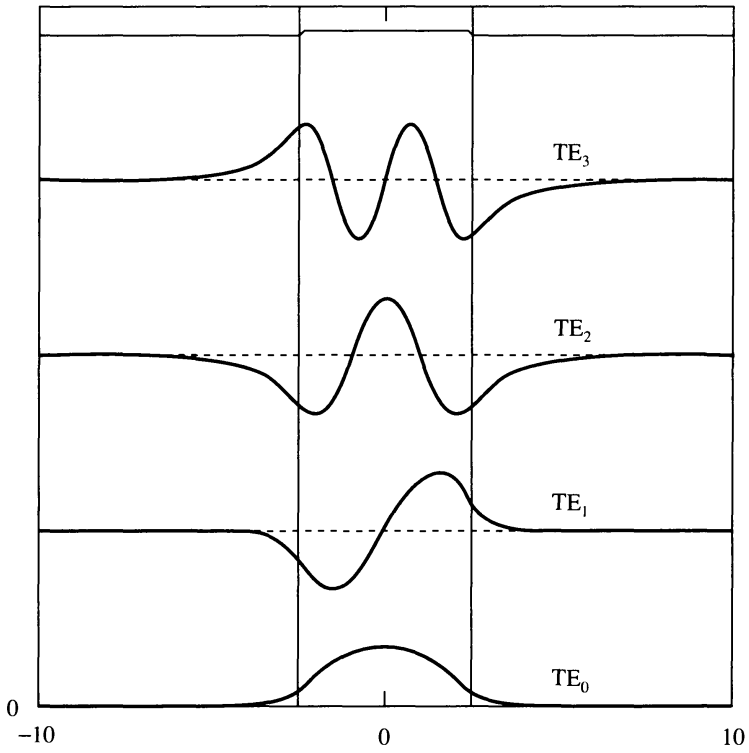


Figure 3.3 Wavefunctions of a symmetrical slab waveguide with $n_2 = 1.6$ and $n_1 = 1.5$. The thickness of the core is equal to $d = 5$ μm, and $\lambda = 1.55$ μm. The normalized propagation constants are 1.5946, 1.5785, 1.5521, and 1.5175. The fundamental mode has the largest propagation constant. The confinement factors (fraction of energy inside the core) for the modes are $\Gamma = 0.9914$, 0.9631, 0.9033 and 0.7511. The fundamental mode has the highest confinement.

For the purpose of describing and comparing the confined modes, it is convenient to define the normalized propagation constant as

$$\bar{\beta} = \frac{\beta}{\omega/c} \tag{3.1-20}$$

Such a normalized propagation is often called the effective index of refraction of the mode, n_{eff}, and is related to the phase velocity of the mode:

$$\bar{\beta} = n_{\text{eff}} = \frac{c}{v_p} \tag{3.1-21}$$

where v_p is the phase velocity of the mode, $v_p = \omega/\beta$. Thus, for confined modes, the normalized propagation constant $\bar{\beta}$ or the effective index n_{eff} is between n_2 and n_1.

Guided TM Modes

We now consider the TM modes whose magnetic field vector is perpendicular to the plane of propagation (xz plane). The derivation of the confined TM modes is similar in principle to that of the TE modes. The field amplitudes are written

$$\begin{aligned} H_y(x, z, t) &= H_m(x) \exp[i(\omega t - \beta z)] \\ E_x(x, z, t) &= \frac{i}{\omega\mu}\frac{\partial}{\partial z} H_y \\ E_z(x, z, t) &= -\frac{i}{\omega\mu}\frac{\partial}{\partial x} H_y \end{aligned} \tag{3.1-22}$$

The wavefunction $H_m(x)$ is

$$H_m(x) = \begin{cases} A \sin hx + B \cos hx, & |x| < \frac{1}{2}d \\ C \exp(-qx), & x > \frac{1}{2}d \\ D \exp(qx), & x < -\frac{1}{2}d \end{cases} \tag{3.1-23}$$

where A, B, C, and D are constants, and the parameters h and q are given by Equation (3.1-9).

The continuity of H_y and E_z at the two interfaces $x = \pm\frac{1}{2}d$ leads, in a manner similar to Equations (3.1-14) and (3.1-15), to the following eigenvalue equation:

$$\begin{aligned} h \tan(\tfrac{1}{2}hd) &= \frac{n_2^2}{n_1^2}q \quad \text{for even solutions} \\ h \cot(\tfrac{1}{2}hd) &= -\frac{n_2^2}{n_1^2}q \quad \text{for odd solutions} \end{aligned} \tag{3.1-24}$$

These two equations can also be combined into a single equation,

$$\tan(hd) = \frac{2h\bar{q}}{h^2 - \bar{q}^2} \tag{3.1-25}$$

where

$$\bar{q} = \frac{n_2^2}{n_1^2}q \tag{3.1-26}$$

Equation (3.1-24) can also be solved by using the graphic method described earlier. Figure 3.4 shows the dispersion relation (effective index n_{eff} versus normalized frequency V) of a typical symmetric waveguide.

The frequency at which $q = 0$ is a cutoff frequency. For a mode with $q = 0$, the field is no longer exponentially decaying in the cladding region and the propagation is no longer confined. Referring to Figure 3.4, we note that TE_0 and TM_0 modes have no cutoff frequency. In other words, these two modes are always confined in a symmetric waveguide. $V = \pi/2$ is

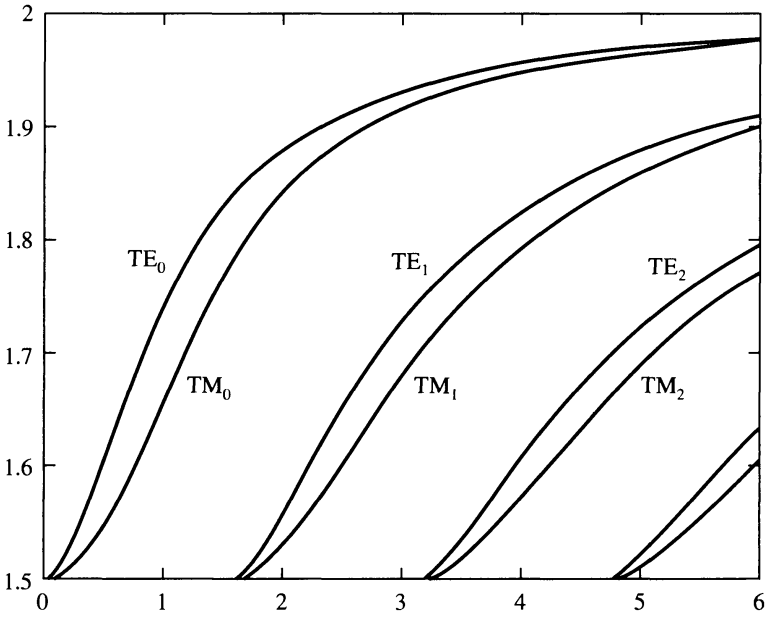


Figure 3.4 The effective index n_{eff} versus normalized frequency V of a typical symmetric waveguide with $n_1 = 1.5$ and $n_2 = 2.0$.

the cutoff frequency for TE_1 and TM_1 modes. For frequency in the range $V < \pi/2$, the waveguide can only support TE_0 and TM_0 modes. At a higher frequency in the range $\pi/2 < V < \pi$, additional modes TE_1 and TM_1 can be supported by the waveguide.

Geometric Optics Treatment

In the preceding, we solved the wave equation for the TE and TM modes of a symmetric slab waveguide. These modes can also be derived by using geometric optics. This is possible because the waveguide consists of layers of homogeneous dielectric materials. Wave propagation in each of the homogeneous regions can be represented by a superposition of two plane waves. One of the plane waves may be considered as the incident wave, whereas the other may be viewed as the reflected one. According to Equations (3.1-5) and (3.1-6), the plane waves that represent the confined modes experience total internal reflection at both interfaces $|x| = \frac{1}{2}d$. It seems that the condition of total internal reflection at both interfaces was enough to assure the confinement of energy in the guiding layer. However, it turns out that total internal reflection is only a necessary condition. In other words, not all rays trapped by internal reflection constitute a mode. A mode, by definition, must have a unique propagation constant and a well-defined field amplitude at each point in space and time. Consider the representation of a mode $E_m(x) \exp[i(\omega t - \beta z)]$ by a zigzagging plane wave of the form

$$E(x, z, t) = E_0(x) \exp[i(\omega t - \beta z - hx)] \tag{3.1-27}$$

where E_0 is a constant, and h is the transverse wavenumber. Let the plane wave travel a distance Δz in a time Δt in one zigzag $ABCD$ (see Figure 3.1). If we follow the ray path $ABCD$ and add the phase shifts due to propagation and total reflections at points B and C, we obtain a total phase shift of

$$\omega \Delta t - \beta \Delta z - 2hd + 2\phi \tag{3.1-28}$$

where ϕ is the phase shift upon total internal reflection at one of the interfaces. A mode of the form $E_m(x) \exp[i(\omega t - \beta z)]$ propagating from point A to point D will gain a phase shift of $\omega \Delta t - \beta \Delta z$. Therefore a total reflecting zigzag ray will become a mode only when the extra transverse phase shift is an integral multiple of 2π; that is,

$$-2hd + 2\phi = -2m\pi \tag{3.1-29}$$

where m is an integer. The minus sign preceding $2m\pi$ is chosen such that m corresponds to the TE_m and TM_m modes of the waveguide. This phase shift, 2ϕ, upon total internal reflection for TE or TM waves, can be expressed conveniently in terms of h and q as [1, 2]

$$\phi = \begin{cases} 2 \tan^{-1}\left(\dfrac{q}{h}\right), & \text{TE} \\ 2 \tan^{-1}\left(\dfrac{n_2^2 q}{n_1^2 h}\right), & \text{TM} \end{cases} \tag{3.1-30}$$

These phase shifts are limited between 0 and π. Therefore the fundamental mode corresponds to the situation when $2hd = 2\phi$ (i.e., $m = 0$). Higher-order modes involve larger hd. The integer m in Equation (3.1-27) therefore assumes only nonnegative values, that is, $m = 0, 1, 2, 3, \ldots$. Condition (3.1-29) is equivalent to the eigenvalue Equations (3.1-16) and (3.1-25).

Consider the example shown in Figure 3.3. This waveguide has $n_1 = 1.5$, $n_2 = 1.6$, $d = 5.0\ \mu m$, $\lambda = 1.550\ \mu m$, with a normalized frequency of $V = 5.6425$. There are four confined TE

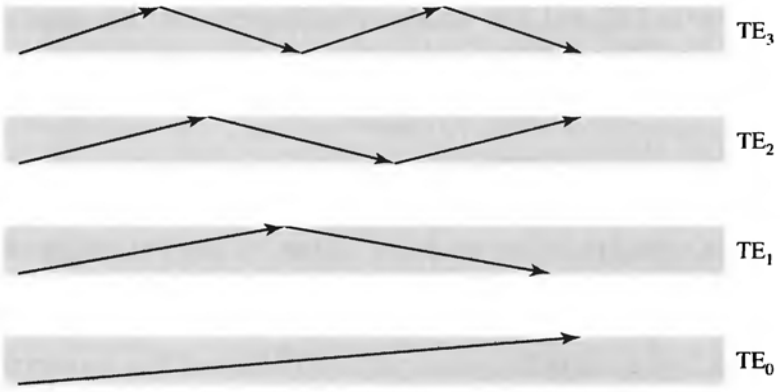


Figure 3.5 Zigzag ray representation of the four TE modes of a symmetrical slab waveguide.

modes in this slab waveguide. The normalized propagation constants $\beta/(\omega/c)$ for these modes are 1.5946, 1.5785, 1.5521, and 1.5175. These normalized propagation constants correspond to the ray incident angles of 85.29°, 80.60°, 75.94°, and 71.52°, in that order. Note that all these angles are greater than the critical angle 69.64°. The fundamental mode ($m = 0$) always has the largest propagation constant, and the angle of incidence with the normalized propagation constant approaches n_2. The highest order mode has the least normalized propagation constant approaching n_1 and the smallest angle of incidence approaching the critical angle. Figure 3.5 shows the zigzag rays corresponding to the modes previously discussed.

Ray optics offers a convenient way of obtaining the mode condition of a slab waveguide. In a more complicated waveguide structure that involves multilayers or inhomogeneous layers, it is very difficult to use the ray optics approach. In addition, the ray optics approach only yields the mode condition. Nothing about the field distribution and the mode orthogonality is obtained.

3.2 TE AND TM CONFINED MODES IN ASYMMETRIC SLAB WAVEGUIDES

The symmetric slab waveguides described in the previous section are useful for introducing the concept of confined propagation because of their simplicity in mathematics. In practice, however, most slab waveguides are not symmetric. In fact, most of the guiding layers are so thin that a substrate is necessary for support of the structure. In this section, we treat the confined modes in asymmetric slab waveguides. The symmetric slab waveguides described in the last section may be viewed as a special case of the asymmetric ones.

Referring to Figure 3.6, we now consider the propagation of confined modes in an asymmetric slab waveguide whose index profile is given by

$$n(x) = \begin{cases} n_1, & 0 < x \\ n_2, & -t < x < 0 \\ n_3, & x < -t \end{cases} \tag{3.2-1}$$

where t is the thickness of the guiding layer, and the index of refraction of the guiding media, n_2, is greater than those of the bounding media, n_1 and n_3. Without loss of generality, we assume that $n_1 < n_3 < n_2$. To be a satisfactory wavefunction, a solution to Maxwell's wave equation must be continuous, single valued, and finite throughout the space. We will first examine the physical nature of the solution as a function of the propagation constant β at a fixed frequency ω. For $\beta > n_2\omega/c$, it follows directly from Equation (3.1-4) that $(1/E)(\partial^2 E/\partial x^2) > 0$ everywhere,

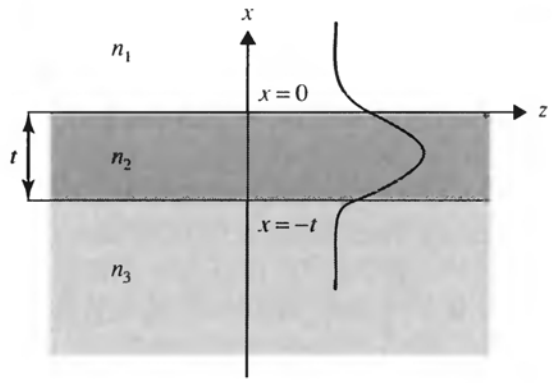


Figure 3.6 Schematic drawing of an asymmetric slab waveguide with a core index of n_2.

and $E(x)$ is exponential at all three regions of the waveguides. If we take $E(x) = \exp(-qx)$, which decays to zero at $x = +\infty$, because of the need to match both $E(x)$ and its derivatives at the two interfaces, the resulting field distribution is infinite at $x = -\infty$, as shown in Figure 3.7a. Such a solution would correspond to a field of infinite energy. It is not physically realizable and thus does not correspond to a real wave.

For $n_3(\omega/c) < \beta < n_2(\omega/c)$, as in Figures 3.7b and 3.7c, it follows from Equation (3.1-4) that the solution is sinusoidal in the core ($-t < x < 0$), since $(1/E)(\partial^2 E/\partial x^2) < 0$, but is exponential

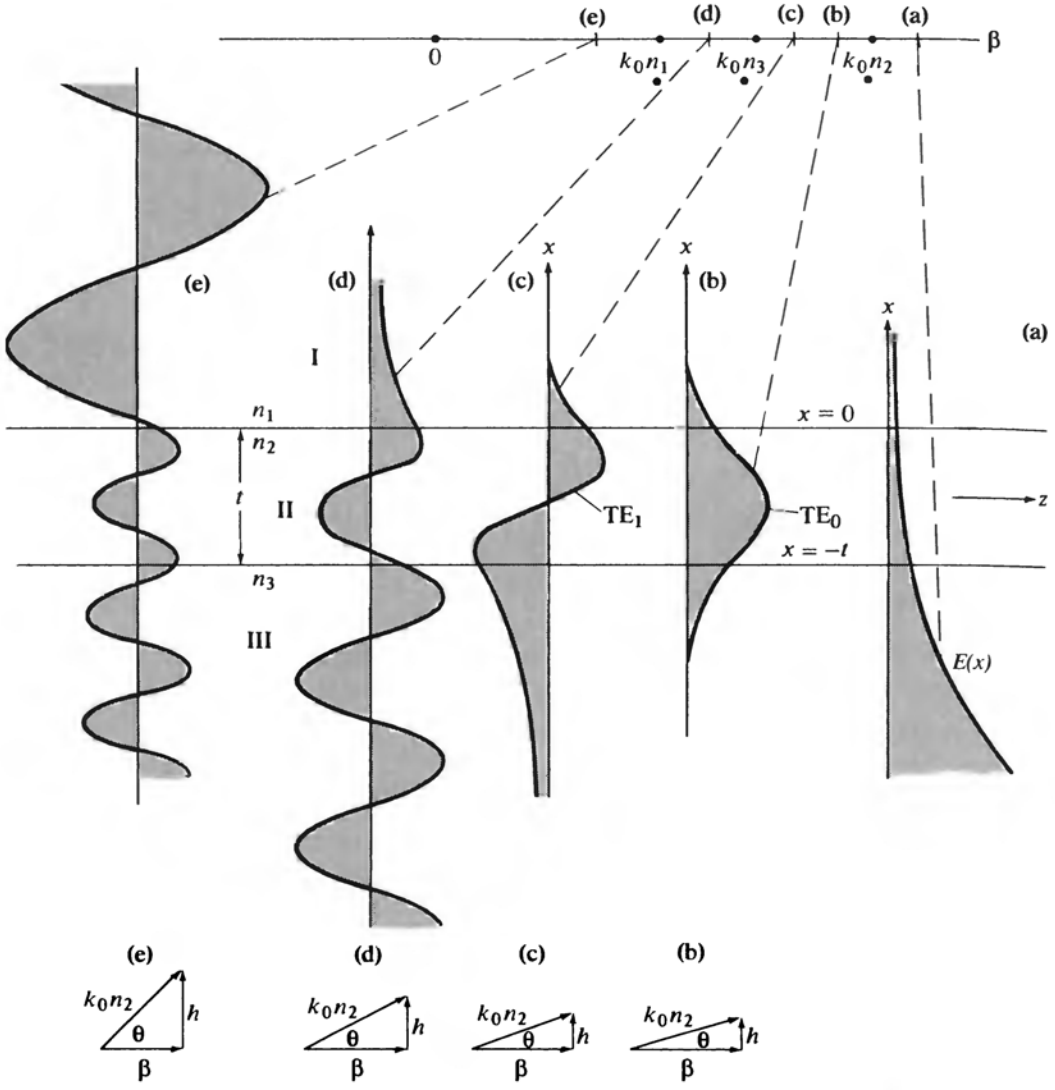


Figure 3.7 Typical field distributions corresponding to different values of β.

in the bounding media. This makes it possible to have a solution $E(x)$ that satisfies the boundary conditions while decaying exponentially in the regions $x < -t$ and $x > 0$. Two such solutions are shown in Figures 3.7b and 3.7c. The energy carried by these modes, as represented by the Poynting vector, is confined to the vicinity of the guiding layer, and, consequently, we will refer to them as confined or guided modes. In fact, only a small fraction of the energy is flowing outside the guiding layer. From the preceding discussion, it follows that a necessary condition for their existence is that $n_1(\omega/c)$, $n_3(\omega/c) < \beta < n_2(\omega/c)$, so that confined modes are possible only when $n_2 > n_1$, n_3; that is, the inner layer (core) possesses the highest index of refraction. In some basic sense, the confined modes in this regime are reminiscent of quantized states of an electron in a potential well, in which the electron is trapped by the potential well.

Mode solutions for $n_1(\omega/c) < \beta < n_3(\omega/c)$ (regime (d) in Figure 3.7) correspond, according to Equation (3.1-4), to exponential behavior in the region $x > 0$ and to sinusoidal behavior in the regions $x < 0$, as illustrated in Figure 3.7d. In this regime, almost all the energy is flowing in the substrate. We will refer to these modes as substrate radiation modes. For $0 < \beta < n_1(\omega/c)$, as in Figure 3.7e, the solution for $E(x)$ becomes sinusoidal in all three regions. These are the so-called radiation modes of the waveguides.

A solution of Equation (3.1-4), subject to the boundary conditions at the interfaces given in what follows, shows that while in regimes (d) and (e) of Figure 3.7, β is a continuous variable, the values of allowed β in the propagation regime $n_3(\omega/c) < \beta < n_2(\omega/c)$ are discrete. The number of confined modes depends on the thickness t, the frequency, and the indices of refraction n_1, n_2, n_3. At a given wavelength, the number of confined modes increases from zero with increasing t. At some cutoff thickness t, the mode TE_0 becomes confined. Further increases in the layer thickness will allow TE_1 to exist as well, and so on.

We now turn our attention to solution of the wave Equation (3.1-4) for the dielectric waveguide sketched in Figure 3.6. The derivation will be limited to the guided modes, which, according to Figure 3.7, have a propagation constant β such that

$$n_3\frac{\omega}{c} < \beta < n_2\frac{\omega}{c}$$

where $n_1 < n_3$. This guide can, in the general case, support a finite number of confined TE modes with field components E_y, H_x, and H_z and TM modes with components H_y, E_x, and E_z. We will derive the mode wavefunction and the corresponding propagation constants.

Guided TE Modes

For TE modes, the electric field vector is perpendicular to the plane of incidence (xz plane). Thus E_y is the only component of the electric field vector. The field component E_y of the TE mode can be taken in the form

$$E_y(x, y, z, t) = E_m(x)e^{i(\omega t - \beta z)} \tag{3.2-2}$$

where β is the propagation constant and $E_m(x)$ is the wavefunction of the mth mode.

The function $E_m(x)$ assumes the following forms in each of the three regions:

$$E_m = \begin{cases} C\exp(-qx), & 0 \le x \\ C\left(\cos(hx) - \dfrac{q}{h}\sin(hx)\right), & -t \le x \le 0 \\ C\left(\cos(ht) + \dfrac{q}{h}\sin(ht)\right)\exp[p(x+t)], & x \le -t \end{cases} \tag{3.2-3}$$

where C is a normalization constant and h, q, and p are given by

$$h = \left[\left(\frac{n_2\omega}{c}\right)^2 - \beta^2\right]^{1/2}, \quad q = \left[\beta^2 - \left(\frac{n_1\omega}{c}\right)^2\right]^{1/2}, \quad p = \left[\beta^2 - \left(\frac{n_3\omega}{c}\right)^2\right]^{1/2} \tag{3.2-4}$$

These relationships are obtained by substituting Equation (3.2-3) into Equation (3.1-4).

The boundary conditions require that E_y and $H_z = (i/\omega\mu)(\partial E_y/\partial x)$ be continuous at both $x = 0$ and $x = -t$. The wavefunction in Equation (3.2-3) has been chosen such that E_y is continuous at both interfaces as well as $\partial E_y/\partial x$ at $x = 0$. By imposing the continuity requirements on $\partial E_y/\partial x$ at $x = -t$, we get, from Equation (3.2-3),

$$h \sin(ht) - q \cos(ht) = p\left(\cos(ht) + \frac{q}{h}\sin(ht)\right)$$

or

$$\tan(ht) = \frac{p + q}{h(1 - pq/h^2)} \tag{3.2-5}$$

This is the so-called mode condition. The propagation constant β of a guided TE mode must satisfy this condition. Given a set of refractive indices n_1, n_2, and n_3 of a slab waveguide, Equation (3.2-5) in general yields a finite number of solutions for β provided the thickness t is large enough. These modes are mutually orthogonal.

The normalization constant C is chosen so that the field $E_m(x)$ in Equation (3.2-3) corresponds to a power flow of 1 W (per unit width in the y direction) along the z axis in the mode. A mode for which $E_y = AE_m(x)$ will thus correspond to a power flow of $|A|^2$ W/m. The normalization condition is given by

$$S_z = \tfrac{1}{2}\int \text{Re}[\mathbf{E} \times \mathbf{H}^*]_z \, dx = 1$$

or equivalently,

$$-\tfrac{1}{2}\int_{-\infty}^{\infty} E_y H_x^* \, dx = \frac{\beta_m}{2\omega\mu}\int_{-\infty}^{\infty} [E_m(x)]^2 dx = 1 \tag{3.2-6}$$

where m denotes the mth confined TE mode (corresponding to the mth eigenvalue of Equation (3.2-5)] and $H_x = -i(\omega\mu)^{-1}\partial E_y/\partial z$. Substitution of Equation (3.2-3) for the wavefunction in Equation (3.2-6) and carrying out the integration lead to, after a few steps of algebraic manipulation,

$$C_m = 2h_m\left(\frac{\omega\mu}{|\beta_m|[t + (1/q_m) + (1/p_m)](h_m^2 + q_m^2)}\right)^{1/2} \tag{3.2-7}$$

The orthonormalization of the modes can be written

$$\int_{-\infty}^{\infty} E_m E_l \, dx = \frac{2\omega\mu}{|\beta_m|}\delta_{l,m} \tag{3.2-8}$$

where $\delta_{l,m}$ is a Kronecker delta.

Guided TM Modes

For TM modes, the magnetic field vector is perpendicular to the plane of incidence (xz plane). The derivation of the confined TM modes is similar in principle to that of the TE modes. The field components are

$$H_y(x, z, t) = H_m(x)e^{i(\omega t-\beta z)}$$

$$E_x(x, z, t) = \frac{i}{\omega\varepsilon}\frac{\partial H_y}{\partial z} = \frac{\beta}{\omega\varepsilon}H_m(x)e^{i(\omega t-\beta z)} \tag{3.2-9}$$

$$E_z(x, z, t) = -\frac{i}{\omega\varepsilon}\frac{\partial H_y}{\partial x}$$

The function $H_m(x)$ assumes the following forms in each of the three regions:

$$H_m(x) = \begin{cases} -C\left(\dfrac{h}{\bar{q}}\cos(ht) + \sin(ht)\right)e^{p(x+t)}, & x \le -t \\ C\left(-\dfrac{h}{\bar{q}}\cos(hx) + \sin(hx)\right), & -t \le x \le 0 \\ -\dfrac{h}{\bar{q}}Ce^{-qx}, & 0 \le x \end{cases} \tag{3.2-10}$$

where C is a normalization constant, h, q, and p are given by Equation (3.2-4), and $\bar{q}$ is defined as follows.

The continuity of H_y and E_z at the two interfaces leads, in a manner similar to Equation (3.2-5), to the eigenvalue equation

$$\tan(ht) = \frac{h(\bar{p} + \bar{q})}{(h^2 - \bar{p}\bar{q})} \tag{3.2-11}$$

where

$$\bar{p} \equiv \frac{n_2^2}{n_3^2}p \quad \text{and} \quad \bar{q} \equiv \frac{n_2^2}{n_1^2}q$$

The normalization constant C is again chosen so that the field represented by Equations (3.2-9) and (3.2-10) carries 1 W of power flow along the z axis per unit width in the y direction. Thus we have

$$\frac{1}{2}\int_{-\infty}^{\infty} H_y E_x^* \, dx = \frac{\beta_m}{2\omega}\int_{-\infty}^{\infty}\frac{H_m^2(x)}{\varepsilon(x)}dx = 1$$

or, using $n^2(x) = \varepsilon(x)/\varepsilon_0$,

$$\int_{-\infty}^{\infty}\frac{|H_m(x)|^2}{n^2(x)}dx = \frac{2\omega\varepsilon_0}{\beta_m} \tag{3.2-12}$$

Carrying out the integration using Equation (3.2-10) gives

$$C_m = 2\sqrt{\frac{\omega\varepsilon_0}{|\beta_m| t_{\text{eff}}}}$$

$$t_{\text{eff}} = \frac{\bar{q}^2 + h^2}{\bar{q}^2}\left(\frac{t}{n_2^2} + \frac{q^2 + h^2}{\bar{q}^2 + h^2}\frac{1}{n_1^2 q} + \frac{p^2 + h^2}{\bar{p}^2 + h^2}\frac{1}{n_3^2 p}\right) \tag{3.2-13}$$

The TM modes are also mutually orthogonal. In fact, all TE and TM modes are mutually orthogonal. The general orthonormality property will be discussed later in this book.

Figure 3.8 illustrates the dependence of propagation constants β on the waveguide thickness (t/λ) for a waveguide with $n_2 = 2.00$, $n_3 = 1.70$, and $n_1 = 1.0$. According to this figure,

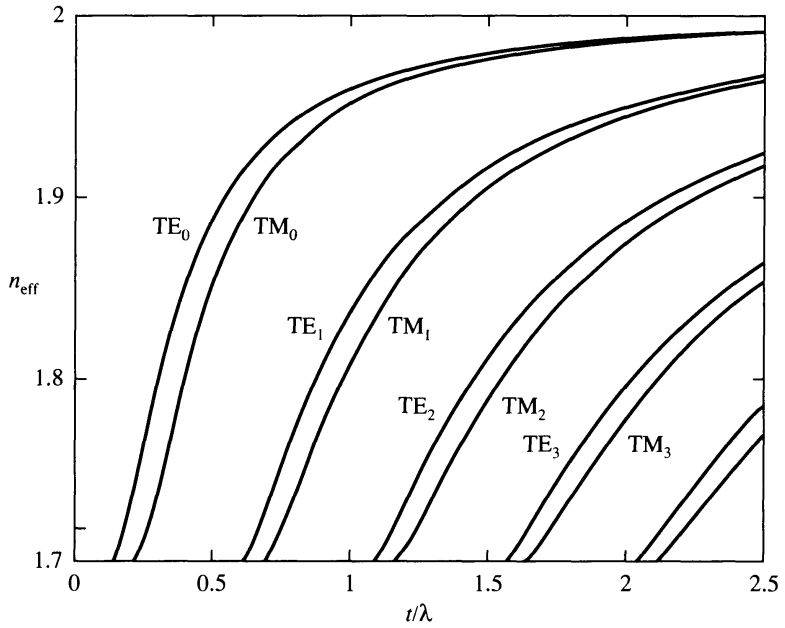


Figure 3.8 Effective index versus thickness/wavelength for the confined modes of an asymmetric waveguide with $n_1 = 1.0$, $n_2 = 2.0$, and $n_3 = 1.7$.

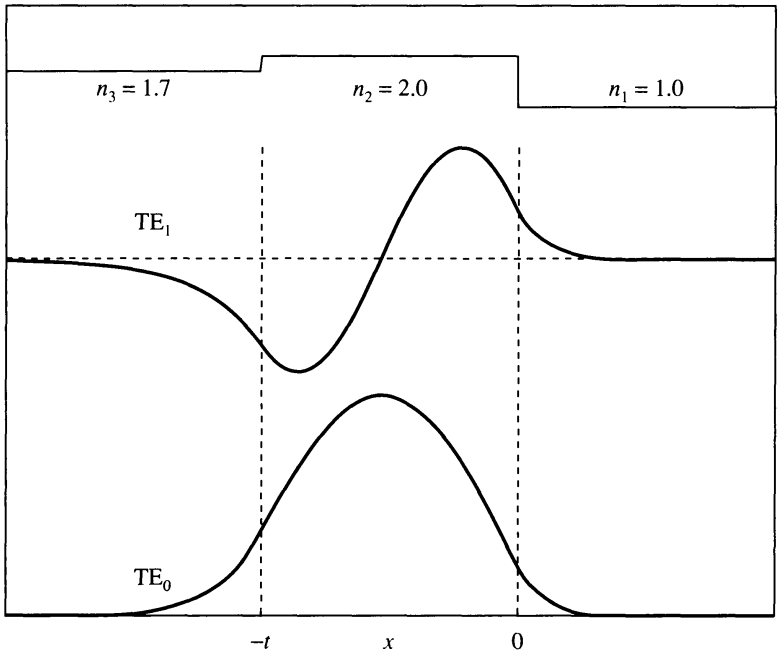


Figure 3.9 Field distribution of the two TE modes of a waveguide with $n_2 = 2.0$, $n_3 = 1.70$, $n_1 = 1.0$, and $t/\lambda = 1.0$. The top curve shows the index profile of the waveguide.

such a waveguide can support two TE and two TM modes when the core thickness is equal to the wavelength (i.e., $t = \lambda$). In fact, at $t/\lambda = 1$, the normalized propagation constants (or effective index) for the TE modes are $n_{eff} = 1.9594, 1.8375$. Wavefunctions of these two TE modes are plotted in Figure 3.9. We note that the wavefunctions are no longer symmetric. The wavefunctions are sinusoidal in the core and evanescent in the cladding regions.

Generally, a mode becomes confined above a certain (cutoff) value of t/λ. At the cutoff value, $p = 0$ and $\beta = n_3\omega/c$, and the mode extends to $x = -\infty$. According to the mode

conditions (3.2-5), (3.2-11), and (3.2-4), cutoff values of t/λ for TE and TM modes are given, respectively, by

$$\left(\frac{t}{\lambda}\right)_{\text{TE}} = \frac{1}{2\pi\sqrt{n_2^2 - n_3^2}}\left[m\pi + \tan^{-1}\left(\frac{n_3^2 - n_1^2}{n_2^2 - n_3^2}\right)^{1/2}\right]$$
$$\left(\frac{t}{\lambda}\right)_{\text{TM}} = \frac{1}{2\pi\sqrt{n_2^2 - n_3^2}}\left[m\pi + \tan^{-1}\frac{n_2^2}{n_1^2}\left(\frac{n_3^2 - n_1^2}{n_2^2 - n_3^2}\right)^{1/2}\right] \tag{3.2-14}$$

where m is an integer ($m = 0, 1, 2, 3, \ldots$) that denotes the mth confined TE (or TM) mode. Note that the cutoff thickness of the TM_m mode is always larger than that of the TE_m mode because $n_1 < n_2$. For values of t/λ slightly above the cutoff value, $p \geq 0$, and the mode is poorly confined. As the values of t/λ increase, so does the value of p, and the mode becomes increasingly confined to layer 2. This is reflected in the normalized propagation constant or the effective index, $\beta\lambda/2\pi$, which at cutoff is equal to n_3 and for large t/λ approaches n_2. In a symmetric waveguide ($n_1 = n_3$), the lowest order modes TE_0 and TM_0 have no cutoff and are confined for all values of t/λ. The confinement is, however, poor when t/λ becomes small.

The total number of confined modes that can be supported by a waveguide depends on the value of t/λ. To study the number of confined modes, we defined the parameter

$$V \equiv \frac{\pi t}{\lambda}\sqrt{n_2^2 - n_3^2} \tag{3.2-15}$$

Let us now consider what happens about the TE modes in a given waveguide (i.e., fixed n_1, n_2, n_3, and t) as the wavelength of the light decreases gradually, assuming that the medium remains transparent and the indices of refraction n_1, n_2, and n_3 do not vary significantly. Since $\omega/c = 2\pi/\lambda$, the effect of decreasing the wavelength is to increase the value of ω/c. At long wavelengths (low frequencies), such that

$$0 < V < \frac{1}{2}\tan^{-1}\left(\frac{n_3^2 - n_1^2}{n_2^2 - n_3^2}\right)^{1/2} \tag{3.2-16}$$

the value of t/λ is below the cutoff value, and no confined mode exists in the waveguide. As the wavelength decreases such that

$$\frac{1}{2}\tan^{-1}\left(\frac{n_3^2 - n_1^2}{n_2^2 - n_3^2}\right)^{1/2} < V < \frac{\pi}{2} + \frac{1}{2}\tan^{-1}\left(\frac{n_3^2 - n_1^2}{n_2^2 - n_3^2}\right)^{1/2} \tag{3.2-17}$$

one solution exists to the mode condition (3.2-5). The mode is designated as TE_0 and has a transverse h parameter falling within the range

$$0 < ht < \pi$$

so that it has no zero crossings in the interior of the guiding layer ($-t < x < 0$). When the wavelength decreases further such that the parameter falls within the range

$$\frac{\pi}{2} + \frac{1}{2}\tan^{-1}\left(\frac{n_3^2 - n_1^2}{n_2^2 - n_3^2}\right)^{1/2} < V < \pi + \frac{1}{2}\tan^{-1}\left(\frac{n_3^2 - n_1^2}{n_2^2 - n_3^2}\right)^{1/2} \tag{3.2-18}$$

the mode condition (3.2-5) yields two solutions. One corresponds to a value of $ht < \pi$ and is thus that of the lowest order TE_0 mode. In the second mode,

$$\pi < ht < 2\pi \tag{3.2-19}$$

and, consequently, it has one zero crossing (i.e., the point where $E_y = 0$) in the guiding region ($-t < x < 0$). This is the so-called TE_1 mode. Both of these modes correspond to the same frequency and can thus be excited simultaneously by the same input field. We notice, however, that the TE_0 mode has a larger value of p (i.e., $p_0 > p_1$) and is therefore more tightly confined to the guiding slab. It follows from Equation (3.2-4) that $\beta_0 > \beta_1$, so that the phase velocity $v_0 = \omega/\beta_0$ of the TE_0 mode is smaller than that of the TE_1 mode. We can now generalize and state that the mth mode (TE or TM) satisfies

$$(m-1)\pi < ht < m\pi \tag{3.2-20}$$

and has $m - 1$ zero crossings in the guiding layer ($-t < x < 0$). The general features of TM modes are similar to those of TE modes, except that the corresponding values of p are somewhat smaller, indicating a lesser degree of confinement. A larger fraction of the total TM mode power thus propagates in the outer medium compared to a TE mode of the same order. This point is taken up in Problem 3.2.

Geometric Optics Treatment

In a manner similar to that leading to Equation (3.1-29), geometric optics can also be used to obtain the mode condition. Following the procedure described in Section 3.1, we obtain

$$-2ht + \phi_{21} + \phi_{23} = -2m\pi \tag{3.2-21}$$

where m is an integer, ϕ_{21} is the phase shift upon total reflection from the interface $x = 0$, and ϕ_{23} is the phase shift upon total reflection from the interface $x = -t$. These phase shifts can be conveniently written in terms of p, q, and h as [1, 2]

$$\phi_{21} = \begin{cases} 2\tan^{-1}\left(\dfrac{q}{h}\right), & \text{TE} \\ 2\tan^{-1}\left(\dfrac{n_2^2 q}{n_1^2 h}\right), & \text{TM} \end{cases} \tag{3.2-22}$$

$$\phi_{23} = \begin{cases} 2\tan^{-1}\left(\dfrac{p}{h}\right), & \text{TE} \\ 2\tan^{-1}\left(\dfrac{n_2^2 p}{n_3^2 h}\right), & \text{TM} \end{cases} \tag{3.2-23}$$

The phase shifts in the preceding equations are again limited between 0 and π. Therefore the fundamental modes correspond to the situation when $m = 0$. Substitution of Equations (3.2-22) and (3.2-23) into Equation (3.2-21) will lead to the mode condition in Equations (3.2-5) and (3.2-11).

Mode Confinement Factor

As shown in Figures 3.3 and 3.9, the electric field and magnetic field of confined modes are finite in the cladding regions. Although the power associated with the mode is propagating

along the z direction, part of the electromagnetic field energy is actually propagating outside the guiding layer. Let Γ_i be the fraction of power flowing in medium i ($i = 1, 2, 3$) of the slab waveguide. By definition, we have

$$\Gamma_i = \frac{\mathrm{Re}\int_i (\mathbf{E} \times \mathbf{H}^*) \cdot \mathbf{z}\, dx}{\mathrm{Re}\int_{-\infty}^{\infty} (\mathbf{E} \times \mathbf{H}^*) \cdot \mathbf{z}\, dx}, \quad i = 1, 2, 3 \tag{3.2-24}$$

where the integral in the numerator is carried out over the ith layer.

Γ_2 is the fraction of mode power flowing in the guiding layer n_2 and is often called the mode confinement factor. The mode functions derived earlier in this section can be employed for calculation of the confinement factor. This is left for the student as an exercise (see Problem 3.2).

3.3 STEP-INDEX CIRCULAR DIELECTRIC WAVEGUIDES (LINEARLY POLARIZED MODES IN OPTICAL FIBERS)

Confined propagation can also occur in circular dielectric waveguides (e.g., silica glass fiber). By virtue of its transmission capacity, silica glass fiber has become the most important medium for long-distance, high-data-rate optical communications. It has caused what can be called, with very little exaggeration, a revolution in the art and practice of communication. This success is due mostly to the prediction [3] and realization [4] of low-loss propagation of confined optical modes in such fibers once the concentration of the absorbing impurities has been reduced to insignificance. The technology of optical communications in fibers can be stated, in simple terms, as that of feeding optical pulses at a maximal rate into one end of a fiber and retrieving them at the other end. The length of the fiber may vary from a few meters in the case of computer interconnect applications, to thousands of kilometers, in the case of transoceanic submarine cables. The main goal of a communication system is to receive the pulses at the output end with minimal loss of energy, minimal spread of pulses, and minimal contamination by noise. Nature, naturally, throws up many obstacles in the path of anyone trying to achieve these modest goals. Some of nature's tricks include diffraction; group velocity dispersion, which causes pulse spreading; and a variety of nonlinear scattering mechanisms. Much of this book is devoted to the understanding of these phenomena and, using this knowledge, to devising strategies for optimal transmission. A good understanding of the propagation of optical waves in fibers is very important. This section is devoted to the confined modes of propagation in fibers with a circular cross section. We will also investigate the problem of group velocity dispersion and propagation attenuation. The topics of detection and noise will be taken up in Chapters 10 and 11.

Referring to Figure 3.10, we consider a circular dielectric waveguide with a step-index profile,

$$n(r) = \begin{cases} n_1, & 0 < r < a \\ n_2, & a < r < b \end{cases} \tag{3.3-1a}$$

where a is the core radius and b is the clad radius.

It consists of a core of refractive index n_1 and radius a, and a cladding of refractive index n_2 and radius b. The radius b of the cladding is usually chosen to be large enough so that the

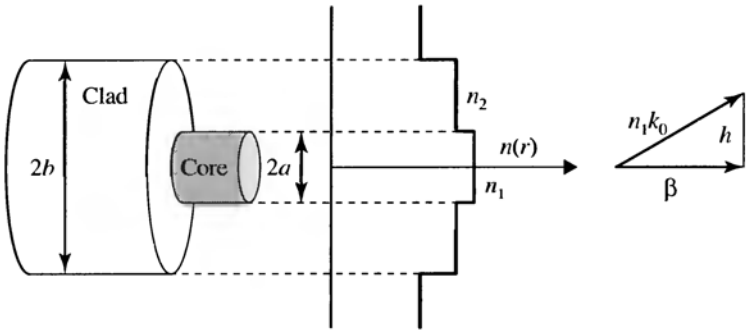


Figure 3.10 Schematic drawing of a circular dielectric waveguide with a clad radius b and a core radius $a \ll b$. $n(r)$ is the index profile of the fiber. $k_0 = \omega/c$.

field of confined modes is virtually zero at $r = b$. A typical optical fiber that supports a single mode of propagation at $\lambda = 1.55$ μm has a core radius of around 5 μm and a clad radius of greater than 100 μm. In the calculation below we will put $b = \infty$ for mathematical simplicity; this is a legitimate assumption in most waveguides, as far as confined modes are concerned.

Unlike the modes of propagation in dielectric slab waveguides, the modes of propagation in circular waveguides are in general a mixture of TE and TM waves. The exact solutions of these hybrid modes (*EH* and *HE* modes) require solving the wave equations in cylindrical coordinates and are very complicated. They are given in Appendixes A and B. In what follows, we will describe approximate solutions of the circular waveguides. A good approximation of the field components and mode condition can be obtained in most fibers whose core refractive index is only slightly higher than that of the cladding medium. As a matter of fact, the index difference $(n_1 - n_2)$ is of the order of 10^{-3} in most single-mode fibers. Assuming that

$$n_1 - n_2 \ll 1 \tag{3.3-1b}$$

it can be shown that the continuity condition on the tangential components of **H** at the interface $(r = a)$ between the core (n_1) and the clad (n_2) becomes identical to that of the tangential components of the field vector **E**. This leads to a tremendous simplification in matching the field components at the core–clad interface. Thus we may use Cartesian components of the field vectors without introducing much complexity in solving the wave equation. This simplified solution of the linearly polarized modes for the circular fibers using the assumption (3.3-1b) is due to Gloge [5]. In the limit (3.3-1b), all the transverse wavenumbers (h, q) of the confined modes $(n_2 k_0 < \beta < n_1 k_0)$ are much smaller compared to the propagation constant β: that is,

$$q, h \ll \beta \tag{3.3-2a}$$

where q and h are given by

$$h = \sqrt{n_1^2 k_0^2 - \beta^2} \quad \text{and} \quad q = \sqrt{\beta^2 - n_2^2 k_0^2} \tag{3.3-2b}$$

with k_0 being the wavenumber of light in vacuum $(k_0 = \omega/c)$.

We now start by solving the wave equation for the transverse Cartesian field components E_x, E_y, H_x, and H_y. These field components also satisfy the wave equations (A-1). For a step-index circular dielectric waveguide, the general solutions are Bessel functions as given in Appendix A. The tangential components (E_z, E_ϕ, H_z, and H_ϕ) must be continuous at the core–clad boundary. We now look for solutions where either the x or y component of the electric field vanishes. In other words, we look for either x-polarized or y-polarized modes of propagation. Since E_ϕ can be expressed in terms of E_x and E_y as

$$E_\phi = -E_x \sin\phi + E_y \cos\phi \tag{3.3-3}$$

it is apparent that the E_ϕ component is simply proportional to either E_x or E_y. Thus the continuity of E_ϕ becomes equivalent to the continuity of E_x or E_y in these new solutions. Take the E field of a y-polarized solution of the form

$$E_x = 0 \tag{3.3-4}$$

$$E_y = \begin{cases} AJ_l(hr)e^{il\phi}\exp[i(\omega t - \beta z)], & r < a \\ BK_l(qr)e^{il\phi}\exp[i(\omega t - \beta z)], & r > a \end{cases} \tag{3.3-5}$$

where A and B are constants, and J_l and K_l are Bessel functions. For confined modes ($n_2k_0 < \beta < n_1k_0$), the electric wavefunction is oscillatory inside the core and evanescent in the cladding region. The Bessel functions $K_l(qr)$ decay exponentially in the cladding region along the radial direction. We assume that $E_z \ll E_y$. The magnetic field components are then given, according to Appendix A, by

$$\begin{aligned} H_x &= \frac{-i}{\omega\mu}\frac{\partial}{\partial z}E_y = \frac{-\beta}{\omega\mu}E_y \\ H_y &\approx 0 \\ H_z &= \frac{i}{\omega\mu}\frac{\partial}{\partial x}E_y \end{aligned} \tag{3.3-6}$$

The longitudinal component of the electric field vector $\mathbf{E}$ is related to H_x according to the Maxwell equation $\nabla \times \mathbf{H} = \varepsilon\, \partial\mathbf{E}/\partial t$:

$$E_z = \frac{i}{\omega\varepsilon}\frac{\partial}{\partial y}H_x = \frac{-i\beta}{\omega^2\mu\varepsilon}\frac{\partial}{\partial y}E_y \tag{3.3-7}$$

where we used Equation (3.3-6) in arriving at the last equation. We note that the field components E_x and H_y are zero in this solution. The other four field components can be expressed in terms of E_y. In order to calculate H_z and E_z, we need to carry out the differentiation with respect to x and y, respectively, according to Equations (3.3-6) and (3.3-7). Since E_y is of the form (3.3-5), we need the relations

$$\frac{\partial}{\partial x} = \frac{\partial r}{\partial x}\frac{\partial}{\partial r} + \frac{\partial \phi}{\partial x}\frac{\partial}{\partial \phi} \tag{3.3-8}$$

and

$$\frac{\partial}{\partial y} = \frac{\partial r}{\partial y}\frac{\partial}{\partial r} + \frac{\partial \phi}{\partial y}\frac{\partial}{\partial \phi} \tag{3.3-9}$$

By using the definition of r and ϕ

$$r = (x^2 + y^2)^{1/2} \tag{3.3-10}$$

$$\phi = \tan^{-1}\left(\frac{y}{x}\right) \tag{3.3-11}$$

we obtain

$$\frac{\partial r}{\partial x} = \frac{x}{r} = \cos\phi \tag{3.3-12}$$

$$\frac{\partial r}{\partial y} = \frac{y}{r} = \sin\phi \tag{3.3-13}$$

$$\frac{\partial \phi}{\partial x} = -\frac{y}{r^2} = -\frac{1}{r}\sin\phi \tag{3.3-14}$$

and

$$\frac{\partial \phi}{\partial y} = \frac{x}{r^2} = \frac{1}{r}\cos\phi \tag{3.3-15}$$

We now substitute Equation (3.3-5) for E_y in Equations (3.3-6) and (3.3-7) and carry out the differentiation, using Equations (3.3-8)–(3.3-15). After some laborious algebra and using the following functional relations of the Bessel function,

$$\begin{aligned} J_l'(x) &= \tfrac{1}{2}[J_{l-1}(x) - J_{l+1}(x)] \\ K_l'(x) &= -\tfrac{1}{2}[K_{l-1}(x) + K_{l+1}(x)] \end{aligned} \tag{3.3-16}$$

$$\begin{aligned} \frac{l}{x}J_l(x) &= \tfrac{1}{2}[J_{l-1}(x) + J_{l+1}(x)] \\ \frac{l}{x}K_l(x) &= -\tfrac{1}{2}[K_{l-1}(x) - K_{l+1}(x)] \end{aligned} \tag{3.3-17}$$

we obtain the following expressions for the field components:

Core ($r < a$):

$$\begin{aligned} E_x &= 0 \\ E_y &= AJ_l(hr)e^{il\phi}\exp[i(\omega t - \beta z)] \\ E_z &= \frac{h}{\beta}\frac{A}{2}[J_{l+1}(hr)e^{i(l+1)\phi} + J_{l-1}(hr)e^{i(l-1)\phi}]\exp[i(\omega t - \beta z)] \\ H_x &= -\frac{\beta}{\omega\mu}AJ_l(hr)e^{il\phi}\exp[i(\omega t - \beta z)] \\ H_y &\cong 0 \\ H_z &= -\frac{ih}{\omega\mu}\frac{A}{2}[J_{l+1}(hr)e^{i(l+1)\phi} - J_{l-1}(hr)e^{i(l-1)\phi}]\exp[i(\omega t - \beta z)] \end{aligned} \tag{3.3-18}$$

Cladding ($r > a$):

$$\begin{aligned} E_x &= 0 \\ E_y &= BK_l(qr)e^{il\phi}\exp[i(\omega t - \beta z)] \\ E_z &= \frac{q}{\beta}\frac{B}{2}[K_{l+1}(qr)e^{i(l+1)\phi} - K_{l-1}(qr)e^{i(l-1)\phi}]\exp[i(\omega t - \beta z)] \\ H_x &= -\frac{\beta}{\omega\mu}BK_l(qr)e^{il\phi}\exp[i(\omega t - \beta z)] \\ H_y &\cong 0 \\ H_z &= -\frac{iq}{\omega\mu}\frac{B}{2}[K_{l+1}(qr)e^{i(l+1)\phi} + K_{l-1}(qr)e^{i(l-1)\phi}]\exp[i(\omega t - \beta z)] \end{aligned} \tag{3.3-19}$$

In arriving at Equations (3.3-18) and (3.3-19), we have also used $\beta = n_1 k_0 \sim n_2 k_0$, since $n_2 k_0 < \beta < n_1 k_0$ and $n_1 - n_2 \ll 1$. Note that E_y and H_x are the dominant field components because, in the limit (3.3-1b), $h, q \ll \beta$. In other words, the field is essentially transverse. The constant B is given by

$$B = \frac{AJ_l(ha)}{K_l(qa)} \tag{3.3-20}$$

to ensure the continuity of E_y ($E_\phi \propto E_y$) at the core boundary $r = a$. The constant A is then determined by the normalization condition.

The field solution (3.3-18) and (3.3-19) is a y-polarized wave ($E_x = 0$). For a complete field description, we also need the mode with the orthogonal polarization (i.e., an x-polarized wave). The field components E_x and E_y of this orthogonal mode are taken as

$$E_x = \begin{cases} AJ_l(hr)e^{il\phi} \exp[i(\omega t - \beta z)], & r < a \\ BK_l(qr)e^{il\phi} \exp[i(\omega t - \beta z)], & r > a \end{cases} \tag{3.3-21}$$

$$E_y = 0 \tag{3.3-22}$$

and the other field components are, according to Maxwell's equations,

$$\begin{aligned} E_z &= \frac{-i}{\omega\varepsilon}\frac{\partial}{\partial x}H_y = \frac{-i\beta}{\omega^2\mu\varepsilon}\frac{\partial}{\partial x}E_x \\ H_x &\approx 0 \\ H_y &= \frac{i}{\omega\mu}\frac{\partial}{\partial z}E_x = \frac{\beta}{\omega\mu}E_x \\ H_x &= \frac{-i}{\omega\mu}\frac{\partial}{\partial y}E_x \end{aligned} \tag{3.3-23}$$

where we have assumed that $E_z \ll E_x$. We note that $E_y = 0$ and $H_x \sim 0$ in this solution. By substituting Equation (3.3-21) for E_x in Equation (3.3-23) and carrying out the differentiation, using Equations (3.3-8)–(3.3-15), we obtain, again after some laborious algebra and using the relations (3.3-16) and (3.3-17), the following expressions for the field amplitudes:

Core ($r < a$):

$$\begin{aligned} E_x &= AJ_l(hr)e^{il\phi} \exp[i(\omega t - \beta z)] \\ E_y &= 0 \\ E_z &= i\frac{h}{\beta}\frac{A}{2}[J_{l+1}(hr)e^{i(l+1)\phi} - J_{l-1}(hr)e^{i(l-1)\phi}] \exp[i(\omega t - \beta z)] \\ H_x &\cong 0 \\ H_y &= \frac{\beta}{\omega\mu}AJ_l(hr)e^{il\phi} \exp[i(\omega t - \beta z)] \\ H_z &= \frac{h}{\omega\mu}\frac{A}{2}[J_{l+1}(hr)e^{i(l+1)\phi} + J_{l-1}(hr)e^{i(l-1)\phi}] \exp[i(\omega t - \beta z)] \end{aligned} \tag{3.3-24}$$

Cladding ($r > a$):

$$
\begin{aligned}
E_x &= BK_l(qr)e^{il\phi}\exp[i(\omega t - \beta z)] \\
E_y &= 0 \\
E_z &= i\frac{q}{\beta}\frac{B}{2}[K_{l+1}(qr)e^{i(l+1)\phi} + K_{l-1}(qr)e^{i(l-1)\phi}]\exp[i(\omega t - \beta z)] \\
H_x &\cong 0 \\
H_y &= \frac{\beta}{\omega\mu}BK_l(qr)e^{il\phi}\exp[i(\omega t - \beta z)] \\
H_z &= \frac{q}{\omega\mu}\frac{B}{2}[K_{l+1}(qr)e^{i(l+1)\phi} - K_{l-1}(qr)e^{i(l-1)\phi}]\exp[i(\omega t - \beta z)]
\end{aligned}
\tag{3.3-25}
$$

In arriving at Equations (3.3-24) and (3.3-25), we again made the assumption that $\beta \cong n_1 k_0 \cong n_2 k_0$ because of Equation (3.3-1b). We note that E_x and H_y are again the dominant field components in this solution. Therefore the mode is again nearly transverse and linearly polarized along the x direction. The constant B is again given by Equation (3.3-20) to ensure the continuity of E_x ($E_\phi \propto E_x$) at the core boundary $r = a$.

We have obtained the field expressions for two types of guided modes whose transverse fields are polarized orthogonally to each other. Based on the circular symmetry of the fiber, we can be sure that these two guided modes must have the same propagation constant and the same intensity and power distribution. The field expressions for the linearly polarized modes are solutions of Maxwell's equations, provided the tangential components of the field vectors are continuous at the dielectric interface $r = a$. The continuity of E_ϕ at $r = a$ leads to Equation (3.3-20). The H_ϕ components are proportional to the E_ϕ components, according to the field expressions (3.3-18), (3.3-19), (3.3-24), and (3.3-25) in this approximation. Therefore the continuity of E_ϕ results in the continuity of H_ϕ. We now consider the continuity of E_z at $r = a$. Since the continuity condition must hold for all azimuth angles ϕ, we must equate the coefficients of $\exp[i(l + 1)\phi]$ and $\exp[i(l - 1)\phi]$ separately. Using the field expressions (3.3-18) and (3.3-19) and (3.3-20), we obtain the following mode conditions:

$$
ha\frac{J_{l+1}(ha)}{J_l(ha)} = qa\frac{K_{l+1}(qa)}{K_l(qa)} \tag{3.3-26}
$$

and

$$
ha\frac{J_{l-1}(ha)}{J_l(ha)} = -qa\frac{K_{l-1}(qa)}{K_l(qa)} \tag{3.3-27}
$$

The same equations result from the continuity of H_z. In addition, if we use the field expressions (3.3-24) and (3.3-25) for the x-polarized mode, we will arrive at the same mode conditions (3.3-26) and (3.3-27). This means that these two transversely orthogonal modes are degenerate in the propagation constant β. The mode condition (3.3-27) is mathematically equivalent to condition (3.3-26) if we use the recurrence relation of the Bessel functions (3.3-17).

The mode condition (3.3-26) obtained in this approximation is much simpler than the exact expression for *HE* and *EH* modes in Appendix B. The exact mode condition (B-11) has twice as many solutions as the simple one (3.3-26) because condition (B-11) is quadratic $J_l'(ha)/J_l(ha)$. This indicates that each solution of Equation (3.3-26) is really twofold degenerate. In fact, the propagation constants of the exact $HE_{l+1,m}$ and $EH_{l-1,m}$ modes are nearly

degenerate [6]. They become exactly the same in the limit $n_2 \rightarrow n_1$. This can also be seen from the expressions of the field components E_z and H_z in Equations (3.3-18), (3.3-19), (3.3-24), and (3.3-25). Comparison of the linearly polarized mode expressions with the exact modes (B-6)–(B-9) shows that the linearly polarized modes are actually a superposition of $HE_{l+1,m}$ and $EH_{l-1,m}$ modes [6]. Two independent linear superpositions lead to the x-polarized and y-polarized modes. The total number of modes is the same in both theories. The eigenvalues obtained from Equation (3.3-26) are labeled β_{lm} with $l = 0, 1, 2, 3, \ldots$ and $m = 1, 2, 3, \ldots$, where the subscript m indicates the mth root of the transcendental Equation (3.3-26). The modes are designated LP_{lm}. The lowest order mode is LP_{01} with a propagation constant labeled β_{01}. This mode corresponds to the HE_{11} mode of the exact solutions. It is important to note that LP_{x01} and LP_{y01} are degenerate and have the same propagation constant β_{01}. This is consistent with the circular symmetry of the waveguide. In practical fibers, the cores are not exactly circular due to production imperfection. Such an asymmetry breaks the degeneracy and leads to $\beta_{x01} \neq \beta_{y01}$. In other words, the x-polarized LP mode and the y-polarized LP mode are propagating at a slightly different propagation constant and thus a slightly different group velocity. This leads to a phenomenon known as polarization mode dispersion (PMD), which will be described in Chapter 7.

The mode conditions for those linearly polarized waves (3.3-26) or (3.3-27) can also be solved numerically. Here we examine the case of $l = 0$. For the purpose of discussion we define an important parameter

$$V = k_0 a(n_1^2 - n_2^2)^{1/2} = \frac{2\pi a}{\lambda}(n_1^2 - n_2^2)^{1/2} = \sqrt{(ha)^2 + (qa)^2} \tag{3.3-28}$$

where a is the core radius. As we will see later, this parameter determines how many confined modes can be supported by the step-index circular fiber. If we replace qa in Equation (3.3-26) with $\sqrt{V^2 - (ha)^2}$, we can then plot both sides of Equation (3.3-26), for a given V, as functions of ha. The intersections are the solutions. Figure 3.11 shows such a graphical

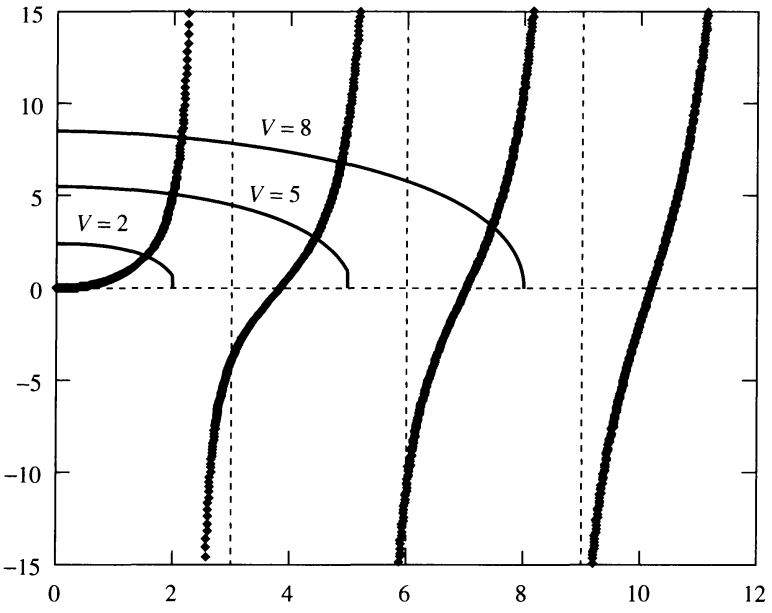


Figure 3.11 Graphical method of determining ha of LP modes with $l = 0$. The left side of Equation (3.3-26) is plotted as the dotted lines. The right side is plotted for $V = 2$, 5, and 8. For $V = 2$, only one solution exists (LP_{01} mode). For $V = 8$, there are three solutions: LP_{01}, LP_{02}, and LP_{03} modes.

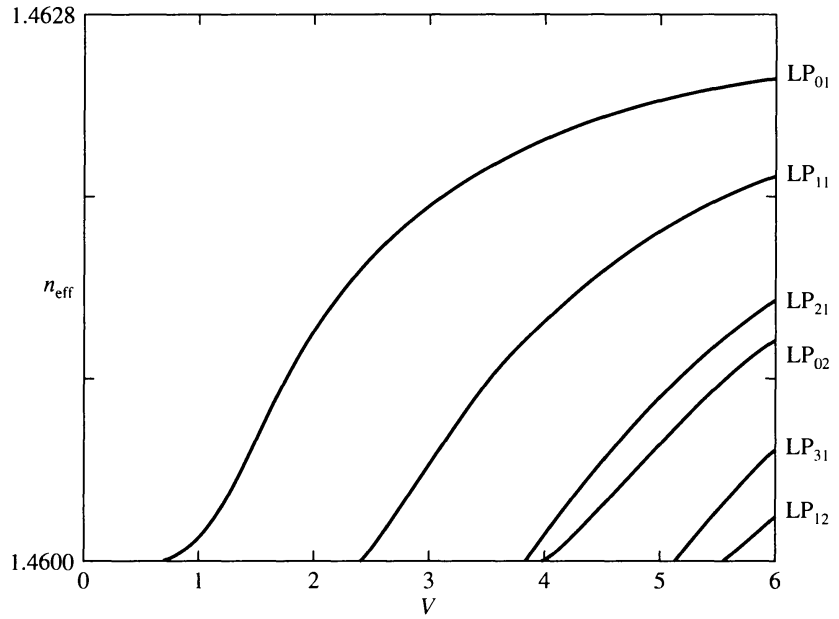


Figure 3.12 Normalized propagation constant n_{eff} as function of normalized frequency V for some of the guided modes of the optical fiber, $n_{eff} = \beta/k_0$. The fiber parameters are $n_1 = 1.4628$, $n_2 = 1.4600$, and $a = 4.7$ μm.

method of determining the solution for $l = 0$ for the cases of $V = 2$, 5, and 8. For the left side, the denominator vanishes at roots of $J_0(ha)$. Thus discontinuity at $\pm\infty$ occurs at $ha = 2.405$, 5.520, 8.654, ... The numerator vanishes at roots of $J_1(ha)$, which occur at $ha = 0$, 3.832, 7.016, ... So, near $ha = 0$, the left side is a positive increasing function of ha starting from zero and reaching $+\infty$ at $ha = 2.405$. For the right side, we note that both $K_0(qa)$ and $K_1(qa)$ are positive functions. Near $ha = 0$, the right side starts from a positive value of $VK_1(V)/K_0(V)$ and decreases as a function of ha, reaching zero at $ha = V$. As a result, there is at least one intersection even if V is very small. As V increases, the number of intersections increases.

Figure 3.12 shows the normalized propagation constant (or effective index) of several modes as a function of the normalized frequency V. We note that for $V < 2.405$, there is only one mode (LP_{01}). For $V = 6$, there are six modes: LP_{01}, LP_{02}, LP_{11}, LP_{12}, LP_{21}, and LP_{31}.

We note that the LP_{01} mode always exists regardless of the fiber parameter V. As V increases beyond $V = 2.405$, the LP_{11} mode starts to appear. Thus we call $V = 2.405$ the cutoff value for the LP_{11} mode. As V increases beyond $V = 3.832$, LP_{21} and LP_{02} modes start to appear. Thus we call $V = 3.832$ the cutoff value for LP_{21} and LP_{02} modes. In general, the mode cutoff corresponds to the condition $q = 0$, which, according to Equation (3.3-27), leads to the condition

$$J_{l-1}(V) = 0 \tag{3.3-29}$$

where $V = k_0 a(n_1^2 - n_2^2)^{1/2} = 2\pi a(n_1^2 - n_2^2)^{1/2}/\lambda$ is the fiber parameter defined in Equation (3.3-28). It follows that the lowest order mode, characterized by $l = 0$, has a cutoff given by the lowest root of the equation

$$J_{-1}(V) = -J_1(V) = 0 \tag{3.3-30}$$

Hence $V = 0$. In other words, the lowest order mode does not have a cutoff. This is the HE_{11} mode and is now labeled LP_{01}. The next mode of the type $l = 0$ cuts off when $J_1(V)$ next equals zero, that is, when $V \sim 3.832$. This mode is labeled LP_{02}. The cutoff values of V for some low-order LP_{lm} modes are given in Table 3.1.

TABLE 3.1 Cutoff Values of V for Some Low-Order LP Modes

V_{cutoff}	$m = 1$	$m = 2$	$m = 3$	$m = 4$
$l = 0$	0	3.832	7.016	10.173
$l = 1$	2.405	5.520	8.654	11.792
$l = 2$	3.832	7.016	10.173	13.323
$l = 3$	5.136	8.417	11.620	14.796
$l = 4$	6.379	9.760	13.017	16.224

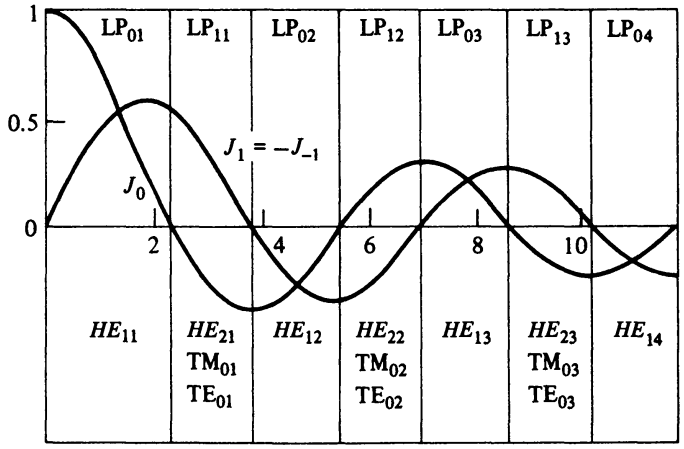


Figure 3.13 The regions of the parameter V for modes of order $l = 0, 1$.

All these values are zeros of the Bessel function. For high-order modes, the cutoff value of V is given approximately according to Equations (3.3-29) and (A-13)

$$V(\mathrm{LP}_{lm}) \cong m\pi + \left(l - \frac{3}{2}\right)\frac{\pi}{2} \tag{3.3-31}$$

Figure 3.13 shows the regions in which a given mode is the highest one allowed for a given l value group, labeled in LP mode designation. Also shown in the figure are the associated *HE*, *EH*, TE, and TM mode notations that are the exact modes. Figure 3.14 shows the field distribution of the LP_{11} modes [6]. We note that the intensity distribution consists of two lobes inside the core. This is distinctly different from that of the fundamental mode (the LP_{01} mode), which has a radially symmetric field distribution $J_0(hr)$ in the core.

One of the most important advantages of using the linearly polarized mode is that the modes are almost transversely polarized and are dominated by one transverse electric field component (E_x or E_y) and one transverse magnetic field component (H_y or H_x). The **E** vector can be chosen to be along any arbitrary radial direction with the **H** vector along a perpendicular radial direction. Once this mode is chosen, there exists another independent mode with **E** and **H** orthogonal to the first pair.

Power Flow and Power Density

We now derive expressions for the Poynting vector and the power flow in the core and cladding. The time-averaged Poynting vector along the waveguide is, acccording to Equation (1.3-17)

$$S_z = \tfrac{1}{2}\mathrm{Re}[E_x H_y^* - E_y H_x^*] \tag{3.3-32}$$

Substituting the field components from Equations (3.3-18) and (3.3-19) or (3.3-24) and (3.3-25) into Equation (3.3-32), we obtain

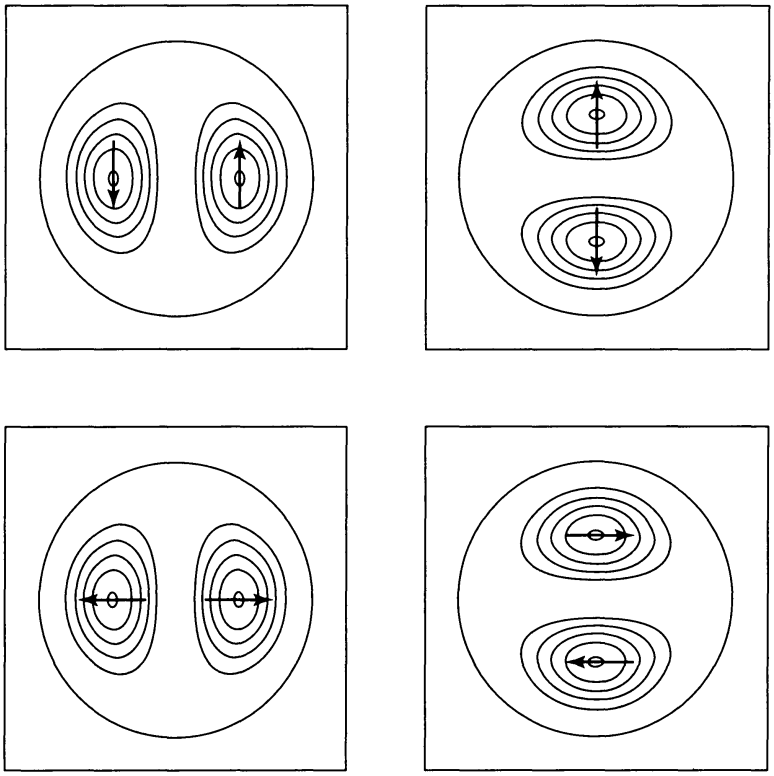


Figure 3.14 Sketch of the four possible field distributions of LP_{11} modes inside the core of the fiber. The arrows indicate the polarization of the electric field vector.

$$S_z = \begin{cases} \dfrac{\beta}{2\omega\mu}|A|^2 J_l^2(hr), & r < a \\ \dfrac{\beta}{2\omega\mu}|B|^2 K_l^2(qr), & r > a \end{cases} \tag{3.3-33}$$

Note that the intensity distribution is cylindrically symmetric (i.e., no ϕ dependence). It is important to note that a combination of the degenerate modes can lead to intensity distribution with azimuth (ϕ) dependence. For example, a linear combination of LP_{11x} mode with $\exp(i\phi)$ and $\exp(-i\phi)$ can lead to an intensity distribution of $\cos^2\phi$ or $\sin^2\phi$ (see Figure 3.14).

The amount of power that is contained in the core and the cladding is given, respectively, by

$$P_{\text{core}} = \int_0^{2\pi}\int_0^a S_z r\, dr\, d\phi \tag{3.3-34}$$

$$P_{\text{clad}} = \int_0^{2\pi}\int_a^{\infty} S_z r\, dr\, d\phi \tag{3.3-35}$$

Using the following integrals of Bessel functions [7] (see also Problem 3.15),

$$\int_0^a rJ_l^2(hr)\, dr = \frac{a^2}{2}\left[J_l^2(ha) - J_{l-1}(ha)J_{l+1}(ha)\right] \tag{3.3-36a}$$

$$\int_a^{\infty} rK_l^2(hr)\, dr = \frac{a^2}{2}\left[-K_l^2(qa) + K_{l-1}(qa)K_{l+1}(qa)\right] \tag{3.3-36b}$$

the powers P_{core} and P_{clad} can be written, respectively,

$$P_{\text{core}} = \frac{\beta}{2\omega\mu}\pi a^2|A|^2[J_l^2(ha) - J_{l-1}(ha)J_{l+1}(ha)] \tag{3.3-37}$$

$$P_{\text{clad}} = \frac{\beta}{2\omega\mu}\pi a^2|B|^2[-K_l^2(qa) + K_{l-1}(qa)K_{l+1}(qa)] \tag{3.3-38}$$

By using Equation (3.3-20) for B and the mode conditions (3.3-26) and (3.3-27), the power P_{clad} can be written

$$P_{\text{clad}} = \frac{\beta}{2\omega\mu}\pi a^2|A|^2\left[-J_l^2(ha) - \left(\frac{h}{q}\right)^2 J_{l-1}(ha)J_{l+1}(ha)\right] \tag{3.3-39}$$

For those ha values that are allowed by the mode condition (3.3-26) or (3.3-27), $J_{l-1}(ha)J_{l+1}(ha)$ is always negative, so that P_{core} is always positive. The negativeness of $J_{l-1}(ha)J_{l+1}(ha)$ can be seen from Equations (3.3-26) and (3.3-27), since the $K_l(qa)$ terms are always positive. According to Equations (3.3-37) and (3.3-39), the total power flow is thus given by

$$P = \frac{\beta}{2\omega\mu}\pi a^2|A|^2\left(1 + \frac{h^2}{q^2}\right)[-J_{l-1}(ha)J_{l+1}(ha)] \tag{3.3-40}$$

The ratio of cladding power to the total power, $\Gamma_2 = (P_{\text{clad}}/P)$, which measures the fraction of mode power flowing in the cladding layer, is given, according to Equations (3.3-39) and (3.3-40), by

$$\Gamma_2 = \frac{P_{\text{clad}}}{P} = \frac{1}{V^2}\left((ha)^2 + \frac{(qa)^2 J_l^2(ha)}{J_{l-1}(ha)J_{l+1}(ha)}\right) \tag{3.3-41}$$

where we used $(ha)^2 + (qa)^2 = k_0^2 a^2(n_1^2 - n_2^2) = V^2$. The confinement factor Γ_1 is given by

$$\Gamma_1 = \frac{P_{\text{core}}}{P} = 1 - \Gamma_2 = \frac{1}{V^2}\left((qa)^2 - \frac{(qa)^2 J_l^2(ha)}{J_{l-1}(ha)J_{l+1}(ha)}\right) \tag{3.3-42}$$

Figure 3.15 shows the confinement factor $\Gamma_1 = P_{\text{core}}/P$ for several modes as a function of the normalized frequency V. Note that the fundamental mode LP_{01} is best confined. Generally, the confinement factor for a given mode $\Gamma_1 = P_{\text{core}}/P$ increases with V.

Generally, the confinement factor for confined modes is zero at cutoff ($qa = 0$). We have seen this in previous sections for dielectric slab waveguides. This is also true for LP modes with $l = 0$ and 1. For high-order LP modes with $l > 1$, the confinement factor at cutoff is obtained by taking the limit $qa \to 0$. This leads to the following limiting value of confinement factor at cutoff:

$$\Gamma_1(qa \to 0) = 1 - \frac{1}{l} \quad (\text{for } l > 1 \text{ at cutoff}) \tag{3.3-43}$$

For example, the confinement factor of the LP_{31} mode at cutoff is $\frac{2}{3}$. Further discussion of the confinement factor for high-order LP modes can be found in Problem 3.11.

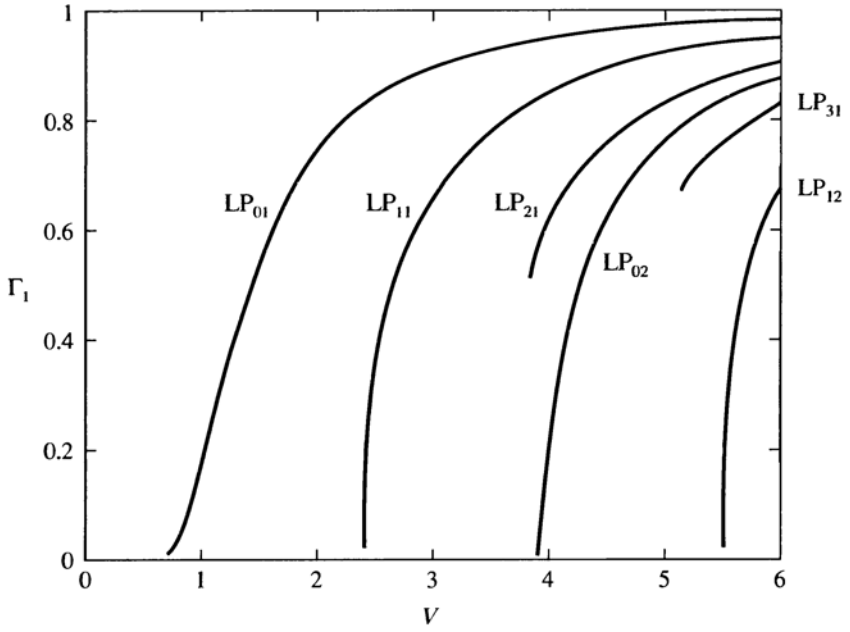


Figure 3.15 Confinement factor Γ_l versus V. Note that, at cutoff, the confinement factor is zero for LP modes with $l = 0$, 1; whereas the confinement factor is $\frac{1}{2}$ for $l = 2$; $\frac{2}{3}$ for $l = 3$; $\frac{3}{4}$ for $l = 4$; and so on.

3.4 EFFECTIVE INDEX THEORY

Although simple layered structures such as dielectric slabs can be used for waveguiding purposes, the confinement of energy is only limited to one dimension. In practice, more complicated waveguide structures are used. For example, the waveguides used in integrated optics or guided-wave optics (also known as planar light circuits (PLCs)) are usually two-dimensional waveguides (e.g., channel waveguides, ridge waveguides). Figure 3.16 shows examples of such two-dimensional waveguides. Exact analytical treatment of these waveguide structures is not possible, except for some special cases. Although numerical solutions can be obtained by various methods, there are several approximate analytical approaches. Here, we will introduce one of the simplest approaches, the *effective index theory*.

Referring to Figure 3.16c, we consider the guiding of electromagnetic radiation in a ridge waveguide. The rectangular waveguide shown in Figure 3.16b can be considered as a special case of this ridge waveguide by taking $d = 0$, where d is the thickness of the guiding layer on both sides of the ridge. The thickness of the guiding layer at the ridge is t, which is chosen to be greater than d because we are interested in the confinement of electromagnetic radiation at

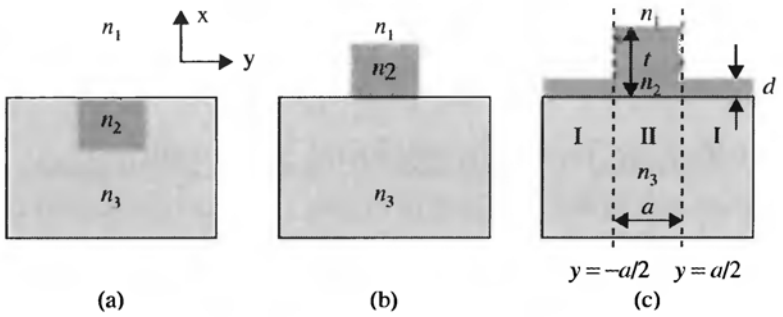


Figure 3.16 Schematic drawing of several two-dimensional waveguides: (a) a rectangular strip of dielectric medium embedded in another dielectric medium of lower index of refraction; (b) ridge waveguide structure; and (c) another ridge structure.

the ridge. The width of the ridge is a. For confined propagation, the index of refraction n_2 is greater than either n_1 or n_3.

We now divide the structure into three regions for the purpose of introducing the concept of effective index theory. These regions are $y < -\frac{1}{2}a$, $-\frac{1}{2}a < y < \frac{1}{2}a$, and $\frac{1}{2}a < y$. In each of the three regions, the confinement along the vertical direction (x) is exactly identical to that of the asymmetric slab waveguide. For simplicity, we assume that only one TE mode is supported by all three regions of the dielectric structure. The propagation constant of the confined mode inside the ridge is, however, different from those on both sides of the ridge because of the difference in layer thickness. The effective index of refraction (or the normalized propagation constant) defined in Equation (3.1-20) for these three segments is

$$n_{\text{eff}}(y) = \begin{cases} n_{\text{I}}, & y < -\frac{1}{2}a \\ n_{\text{II}}, & -\frac{1}{2}a < y < \frac{1}{2}a \\ n_{\text{I}}, & \frac{1}{2}a < y \end{cases} \tag{3.4-1}$$

where n_{I} and n_{II} are the effective indices of refraction of the asymmetric slab waveguides, respectively.

In the theory of effective index, the lateral waveguiding is treated by taking Equation (3.4-1) as a "slab" waveguide structure along the y direction. Thus confinement of electromagnetic radiation requires that $n_{\text{II}} > n_{\text{I}}$. This is always true when the thickness t is greater than d because, according to Figure 3.7, the effective index of refraction of any confined mode is an increasing function of thickness. To illustrate further the lateral guiding, we will consider the following example.

GaAs Ridge Waveguide

We consider a ridge waveguide made of GaAs layers and AlGaAs substrate. Let the indices of refraction be $n_1 = 1$, $n_2 = 3.5$, and $n_3 = 3.2$. The thicknesses are $t = 0.40\lambda$ and $d = 0.25\lambda$. According to a numerical solution of Equation (3.2-5), the effective indices are

$$n_{\text{I}} = 3.301 \quad \text{and} \quad n_{\text{II}} = 3.388$$

We note that a step height of 0.15λ at the ridge gives rise to an increase in the effective index of 0.087. Such an index difference is sufficient to provide lateral waveguiding. In fact, if we take $a = 0.5\lambda$ as the width of the ridge and solve for the confined TE modes of such a symmetric waveguide, we obtain a single TE mode with a normalized propagation constant of 3.348. The wavefunction of such a mode is similar to the fundamental mode shown in Figure 3.3. This wavefunction shows the lateral confinement.

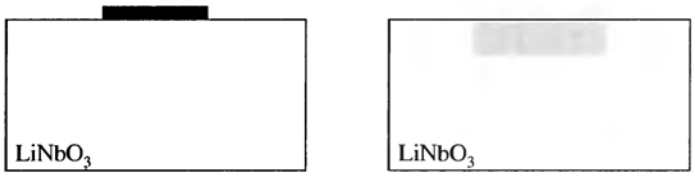
It is important to note that the effective index theory is a good approximation, provided the index difference between the core and cladding is small so that the scalar wave approximation is valid. In the scalar approximation, we ignore the vector nature of the electromagnetic waves.

Ridge Waveguides or Two-Dimensional Waveguides

The ridge waveguide structures described in Figure 3.16 can be obtained by several different approaches, including etching and diffusion. In the following, we describe some examples of channel waveguides that are important in optical communications.

Thermal Indiffusion

A channel waveguide in a $LiNbO_3$ crystal can be obtained by using a conventional thermal indiffusion process. Prior to the diffusion process, a thin strip of metallic Ti layer about 100 nm



(a) Masking (b) Thermal diffusion

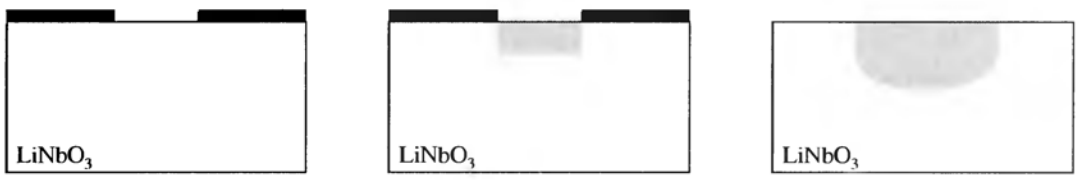
Figure 3.17 Fabrication steps in Ti-diffused $LiNbO_3$ channel waveguide. (a) First, a thin strip of Ti metal is deposited on the optically polished surface of the $LiNbO_3$ crystal wafer. Such a strip or a more complicated planar optical circuit pattern can be obtained by a photolithographic method that defines the channel waveguide circuit. (b) The wafer is then heat treated at an elevated temperature. As a result of thermal diffusion, Ti atoms migrate into the crystal wafer. This leads to an optical circuits with a higher refractive index at the core of the waveguide circuit.

in thickness is deposited on the surface of a $LiNbO_3$ wafer by a vacuum evaporation technique. A planar optical circuit can be patterned by a conventional photolithography technique. The wafer is then heat treated at around 1000 °C for a period of time (e.g., 20 hours). As a result of the heating-assisted diffusion, the Ti atoms migrate into the crystal. The presence of Ti atoms in the $LiNbO_3$ crystal increases both the ordinary and extraordinary index of refraction. This creates a channel waveguide in the region of Ti-atom concentration. Figure 3.17 illustrates the steps in a Ti-diffused $LiNbO_3$ channel waveguide.

A channel waveguide in silica (SiO_2) wafer can be obtained by a similar thermal indiffusion process using Pb or other metals. The presence of Pb atoms in silica increases the index of refraction and thus creates a channel waveguide.

Proton Exchange in Lithium Niobate

A channel waveguide in a lithium niobate wafer can also be obtained by using the proton exchange process (see Figure 3.18). First, a patterned material is applied to the optically polished surface of the $LiNbO_3$ crystal. The pattern can be obtained by a photolithographic method that defines the channel waveguide optical circuit. The wafer is then immersed in a bath of liquid acid that provides a source of protons that exchanges with lithium ions, creating a thin layer of protons at the surface. The protons at the $LiNbO_3$ surface can penetrate further into the crystal via diffusion during an anneal process at high temperature. Once cooled, the channel waveguide is extremely stable. As a result, the extraordinary index of refraction is increased in the proton-exchanged regions and the ordinary index is reduced. This leads to a channel waveguide that supports confined propagation with the **E** field polarized along the c axis of the crystal. There are other techniques that can be employed to increase the index of refraction and to produce optical waveguides. These include ion implantation and the sol-gel process.



(a) Masking (b) Proton exchange (c) Anneal

Figure 3.18 Fabrication steps in proton-exchange process in a $LiNbO_3$ channel waveguide.

3.5 WAVEGUIDE DISPERSION IN OPTICAL FIBERS

So far in this chapter, we have discussed various optical waveguides, including dielectric slab waveguides and step-index optical fibers. These waveguides can be described in terms of the index of refraction as a function of space. Confined propagation in these waveguides is governed by the wave equation, which provides the wavefunction and the propagation constant once the index distribution of the waveguide is given. The wave equation is often written

$$\left\{\nabla_\perp^2 + \frac{\omega^2}{c^2} n^2(x, y)\right\} E_m(x, y) = \beta_m^2 E_m(x, y) \tag{3.5-1}$$

where $\nabla_\perp^2$ is the transverse Laplacian, $E_m(x, y)$ is the wavefunction of the mth mode, and β_m is the propagation constant. Equation (3.5-1) may be viewed as an eigenvalue problem with β_m^2 as the eigenvalue. Consider a general mode of propagation

$$E(x, y, z, t) = E_m(x, y) \exp[i(\omega t - \beta_m z)] \tag{3.5-2}$$

In multimode waveguides, a general wave of propagation can be written as a linear combination of all the modes.

For the convenience of discussion, we define an effective index of a mode as

$$n_{\text{eff}} = \frac{\beta_m}{\omega/c} \tag{3.5-3}$$

It is important to note that both the wavefunction and the propagation constant β_m depend on the waveguide geometry $n^2(x, y)$ as well as the angular frequency ω. For confined modes, the effective index is always between the core index and the clad index. In other words,

$$n_{\text{clad}} < n_{\text{eff}} < n_{\text{core}} \tag{3.5-4}$$

For multimode waveguides, each confined mode has a distinct effective index. The fundamental mode has the largest effective index. Generally, each mode has its own phase velocity and group velocity. As a result, modal dispersion occurs in multimode waveguides. This was discussed in Section 2.9.

In a multimode waveguide, if an optical wave (e.g., a sequence of pulses) is represented by a superposition of a number of modes, then the pulses will spread as they propagate in the waveguide. The spread can be due to (a) modal dispersion, (b) waveguide dispersion, and (c) material dispersion. In single-mode waveguides, the pulse spreading is due to (a) waveguide dispersion and (b) material dispersion. In this section, we discuss waveguide dispersion and the pulse spreading in single-mode optical fibers.

The effective index of a step-index fiber (or slab waveguide) depends on the frequency of light ω, even if the indices of refraction of the fiber are independent of the frequency of light ω. This is a result of the different wavefunctions and ray paths inside the waveguide at different frequencies. The dispersion figures in Sections 3.1–3.3 (e.g., Figure 3.12) show the dependence of n_{eff} on frequency ω. Thus we may write

$$n_{\text{eff}} = n_{\text{eff}}[n_1(\omega), n_2(\omega), \omega] \tag{3.5-5}$$

where the effective index depends on the frequency directly due to waveguide dispersion and indirectly due to material dispersion through $n_1(\omega)$ and $n_2(\omega)$.

We now examine these two contributions of the dispersion. Using the method of variation in perturbation theory, we apply a small variation of the frequency $\delta\omega$ to the wave equation (3.5-1). Let $\delta\beta_m^2$, δn_1^2, and δn_2^2 be the corresponding variations in the propagation constant of the mth mode, and the refractive indices, respectively. We obtain from the wave equation (3.5-1)

$$\delta\beta_m^2 = \Gamma_1 \frac{\omega^2}{c^2}\delta n_1^2 + \Gamma_2 \frac{\omega^2}{c^2}\delta n_2^2 + \Gamma_1 \frac{2\omega\,\delta\omega}{c^2} n_1^2 + \Gamma_2 \frac{2\omega\,\delta\omega}{c^2} n_2^2 \tag{3.5-6}$$

where the first two terms on the right side are due to the changes of the index of refraction δn_1^2 and δn_2^2, while the last two terms are due to a change of the frequency $\delta\omega$; Γ_1 and Γ_2 are confinement factors given by

$$\Gamma_\alpha = \frac{\iint_{n_\alpha} \mathbf{E}_m^* \cdot \mathbf{E}_m \, dx\, dy}{\iint_{\infty} \mathbf{E}_m^* \cdot \mathbf{E}_m \, dx\, dy} \quad (\alpha = 1, 2) \tag{3.5-7}$$

where the integral in the numerator is over the region where the index of refraction is n_α.

Equation (3.5-6) can be rewritten as

$$\frac{d\beta}{d\omega} = \Gamma_1 \frac{\omega}{c}\frac{n_1}{n_{\text{eff}}}\frac{\partial n_1}{\partial\omega} + \Gamma_2 \frac{\omega}{c}\frac{n_2}{n_{\text{eff}}}\frac{\partial n_2}{\partial\omega} + \frac{1}{cn_{\text{eff}}}(\Gamma_1 n_1^2 + \Gamma_2 n_2^2) \tag{3.5-8}$$

where we have dropped the subscript m for simplicity. Formula (3.5-8) can be employed for each mode. Using Equation (3.5-3) for β, the above equation can further be written

$$\frac{d}{d\omega} n_{\text{eff}} = \Gamma_1 \frac{n_1}{n_{\text{eff}}}\frac{\partial n_1}{\partial\omega} + \Gamma_2 \frac{n_2}{n_{\text{eff}}}\frac{\partial n_2}{\partial\omega} + \frac{1}{\omega n_{\text{eff}}}(\Gamma_1 n_1^2 + \Gamma_2 n_2^2 - n_{\text{eff}}^2) \tag{3.5-9}$$

where the first two terms on the right side are due to material dispersion, and the last term is due to waveguide dispersion, which represents the variation of the effective index of the guided mode with respect to the frequency by keeping both the core and cladding index constant. In other words, the dispersion of a guided wave consists of the sum of an intrinsic material dispersion and the waveguide dispersion. Thus we can write

$$\frac{d}{d\omega} n_{\text{eff}} = \left(\frac{\partial n_{\text{eff}}}{\partial\omega}\right)_{\text{material}} + \left(\frac{\partial n_{\text{eff}}}{\partial\omega}\right)_{\text{waveguide}} \tag{3.5-10}$$

where the material dispersion is given by

$$\left(\frac{\partial n_{\text{eff}}}{\partial\omega}\right)_{\text{material}} = \Gamma_1 \frac{n_1}{n_{\text{eff}}}\frac{\partial n_1}{\partial\omega} + \Gamma_2 \frac{n_2}{n_{\text{eff}}}\frac{\partial n_2}{\partial\omega} \tag{3.5-11}$$

and the waveguide dispersion is given by

$$\left(\frac{\partial n_{\text{eff}}}{\partial\omega}\right)_{\text{waveguide}} = \frac{1}{\omega n_{\text{eff}}}(\Gamma_1 n_1^2 + \Gamma_2 n_2^2 - n_{\text{eff}}^2) \tag{3.5-12}$$

For single-mode silica fibers used in modern optical communications, the clad and core are both made of silica with $n_1 \approx n_2$, so Equations (3.5-10)–(3.5-12) can be written approximately

$$\left(\frac{\partial n_{\text{eff}}}{\partial\omega}\right)_{\text{material}} = \frac{\partial n}{\partial\omega} \tag{3.5-13}$$

$$\left(\frac{\partial n_{\text{eff}}}{\partial \omega}\right)_{\text{waveguide}} = \frac{1}{\omega n_{\text{eff}}}(n_2^2 + \Gamma_1[n_1^2 - n_2^2] - n_{\text{eff}}^2) \tag{3.5-14}$$

where n is the index of refraction of the fiber material (silica), and we have used $n_2 \approx n_{\text{eff}} \approx n_1$ and $\Gamma_2 = 1 - \Gamma_1$. We note that the term on the right side of Equation (3.5-13) is purely material dispersion, while Equation (3.5-14) addresses the waveguide dispersion.

Using the confinement factor expression in Section 3.3, the waveguide dispersion can be written (see Problem 3.12)

$$\left(\frac{\partial n_{\text{eff}}}{\partial \omega}\right)_{\text{waveguide}} = \frac{1}{\omega n_{\text{eff}}}(n_{\text{eff}}^2 - n_2^2)\left(\frac{-J_m^2(ha)}{J_{m-1}(ha)J_{m+1}(ha)}\right) \tag{3.5-15}$$

Since $J_{m-1}(ha)J_{m+1}(ha)$ is always negative for the confined-modes that satisfy the mode conditions (3.3-26) and (3.3-27), we note that the right side of Equation (3.5-15) is always positive. In other words, the effective index is an increasing function of frequency. This is consistent with the dispersion curves shown in Figure 3.12.

In optical networks, the dispersion of a mode of propagation is often written

$$\frac{dn_{\text{eff}}}{d\lambda} = \left(\frac{\partial n}{\partial \lambda}\right)_{\text{material}} + \left(\frac{\partial n_{\text{eff}}}{\partial \lambda}\right)_{\text{waveguide}} \tag{3.5-16}$$

and

$$\frac{d^2 n_{\text{eff}}}{d\lambda^2} = \left(\frac{\partial^2 n}{\partial \lambda^2}\right)_{\text{material}} + \left(\frac{\partial^2 n_{\text{eff}}}{\partial \lambda^2}\right)_{\text{waveguide}} \tag{3.5-17}$$

Consider the propagation of an optical pulse in a single-mode waveguide of length L. Let n_{eff} be the effective index of the mode of propagation. The phase shift of the beam at output due to propagation is $\phi = n_{\text{eff}}(\omega/c)L$. It can be shown that the group delay (flight time) is

$$\tau = \frac{d\phi}{d\omega} = L\left(\frac{n_{\text{eff}}}{c} + \frac{\omega}{c}\frac{dn_{\text{eff}}}{d\omega}\right) = L\left(\frac{n_{\text{eff}}}{c} - \frac{\lambda}{c}\frac{dn_{\text{eff}}}{d\lambda}\right) \tag{3.5-18}$$

The group velocity dispersion (in units of picaseconds per nanometer of bandwidth per kilometer of fiber) for optical fibers is defined as

$$D = \frac{d}{d\lambda}\left(\frac{\tau}{L}\right) \quad \text{(ps/nm-km)} \tag{3.5-19}$$

Using Equations (3.5-18) and (3.5-19), we obtain

$$D = -\frac{1}{c\lambda}\left(\lambda^2 \frac{d^2 n_{\text{eff}}}{d\lambda^2}\right) \tag{3.5-20}$$

If the waveguide material is dispersive, the group velocity dispersion parameter D can be written, according to Equation (3.5-17),

$$D = -\frac{1}{c\lambda}\left(\lambda^2 \frac{\partial^2 n}{\partial \lambda^2}\right)_{\text{material}} - \frac{1}{c\lambda}\left(\lambda^2 \frac{\partial^2 n_{\text{eff}}}{\partial \lambda^2}\right)_{\text{waveguide}} \tag{3.5-21}$$

Note that the terms inside the parentheses are dimensionless. The parameter D has a dimension of s/m-m. For practical application, D is often expressed in units of ps/nm-km.

EXAMPLE: PULSE SPREADING

Consider the transmission of 10 Gb/s signals at $\lambda = 1500$ nm in a single-mode fiber of 100 km, with a group velocity dispersion of $D = 17$ ps/nm-km. The pulse spreading after a transmission of distance L can be written

$$\Delta\tau = DL\,\Delta\lambda \tag{3.5-22}$$

where $\Delta\lambda$ is the spectral bandwidth of the signals. For 10 Gb/s, the pulse width is $\tau_0 = 100$ ps. The spectral bandwidth is approximately $\Delta\nu = 1/\tau_0 = 10$ GHz. In terms of nanometers, the bandwidth is

$$\Delta\lambda = \frac{\lambda^2}{c}\Delta\nu = 0.075 \text{ nm}$$

The pulse spreading is thus given by, using Equation (3.5-22), $\Delta\tau = 128$ ps

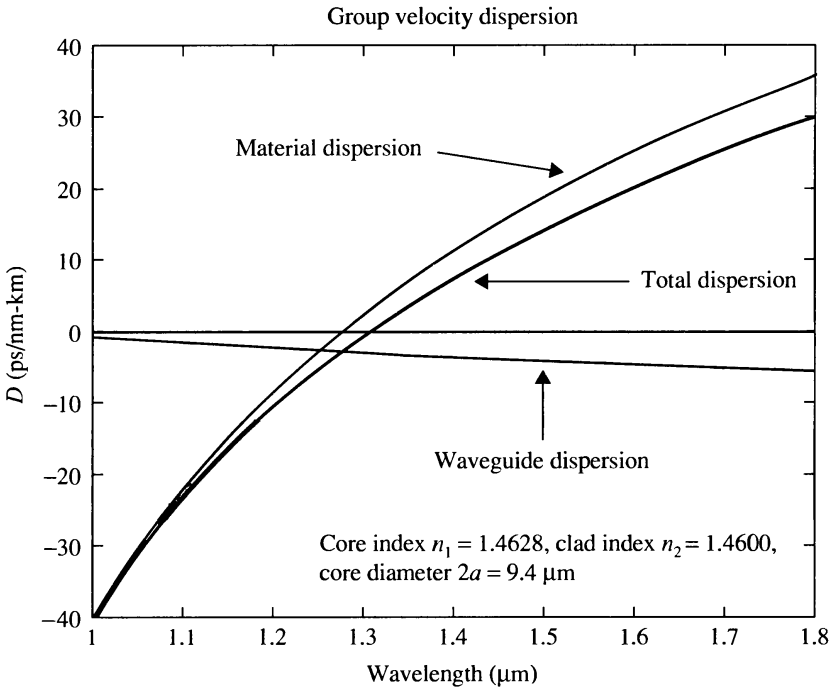


Figure 3.19 The material dispersion of silica, the waveguide dispersion of a single-mode GeO_2-doped silica fiber, and the resultant total dispersion in the spectral region between 1 and 1.8 μm.

Figure 3.19 shows the material dispersion of silica, the waveguide dispersion of a single-mode GeO_2-doped silica fiber with a core diameter of 9.4 μm, and the resultant total dispersion. Not that the total dispersion is zero at $\lambda = 1.3$ μm [8, 9]. It is important to note that the waveguide dispersion depends on core diameter a as well as the core and clad indices n_1 and n_2. It is possible to tailor the zero-dispersion wavelength by balancing the positive material dispersion against the negative waveguide dispersion [10]. Thus, by choosing a proper core diameter a between 4 and 5 μm, and a relative refractive index difference of $(n_1 - n_2)/n_1 > 0.004$, the wavelength of zero dispersion can be shifted to the 1.5–1.6 μm region, where the propagation loss is minimum [11–16].

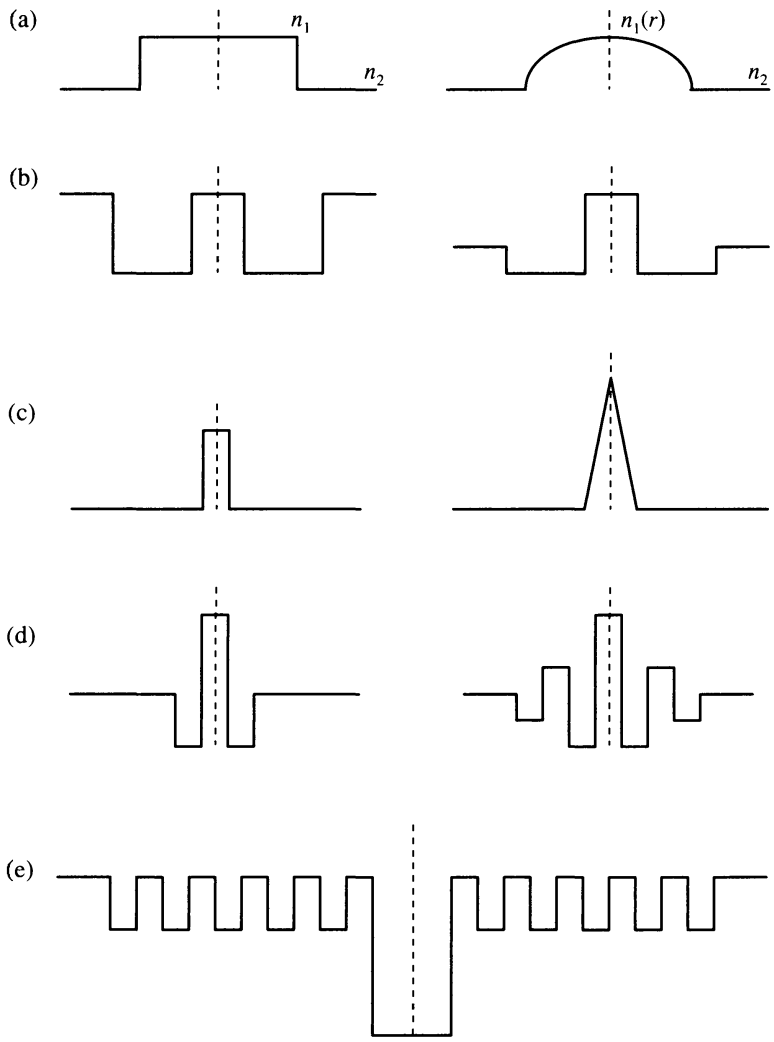


Figure 3.20 Various index profiles of fibers. (a) Multimode fibers with a core diameter of about 50 μm, and a clad diameter of 120 μm. (b) Single-mode fibers with a core diameter of about 10 μm. (c) Dispersion-shifted fibers with a core diameter of about 5 μm. (d) Dispersion-flattened fibers with double clad or triple clad and a core diameter of about 6 μm. (e) Bragg fibers with a core of low index of refraction, or a hollow core.

We have so far discussed the step-index circular fibers. Closed-form solutions for the modes are available. There are optical fibers with different refractive index profiles. These fibers are designed to provide minimum loss or minimum dispersion in the spectral region of interest. Some of the fibers can provide near-zero dispersion over a broad spectral region of interest. Figure 3.20 shows some of the index profiles. Some of the profiles involve additional cladding layers of different refractive indices [17–21]. These index profiles offer additional degrees of freedom in the design of fibers for specific applications.

According to Figure 3.19, we note that at $\lambda = 1550$ nm the chromatic dispersion is mostly due to material dispersion. If a fiber can be fabricated with an air core, the material dispersion can be virtually eliminated. To support a confined propagation in a low-index region requires a multilayer cladding with a periodic variation of the clad index in the radial direction. High reflection via Bragg scattering in the radial direction provides the confinement (see Figure 3.20e). The discussion of such fiber is beyond the scope of this book. Interested readers are referred to Reference [22].

3.6 ATTENUATION IN SILICA FIBERS

Propagation attenuation exists in virtually all waveguides. As a result of attenuation, signal power $P(z)$ will decay exponentially according to the following equation:

$$P(L) = P(0)\exp(-\alpha L) \tag{3.6-1}$$

where $P(0)$ is the power at the input, L is the distance of propagation, and α is the linear attenuation coefficient. In optical communications, attenuation is often measured in units of dB, and the attenuation coefficient is measured in units of dB/km. Thus signal attenuation (dB) per unit length is defined as

$$\alpha = \frac{10}{L}\log_{10}\left(\frac{P(L)}{P(0)}\right) \tag{3.6-2}$$

There are many sources of propagation attenuation in optical fibers. In the early days of silica fiber development, propagation attenuation was mainly due to the presence of impurities that absorb optical energy. Trace metal impurities, such as iron, nickel, and chromium, are introduced into the fiber during the fabrication. The electronic energy levels of these metal ions are broadened by the random atomic environment in the silica. Some of the electronic transitions fall in the spectral regime of the propagating optical beam. Photons in the optical beam are absorbed by these electronic transitions. The absorption also occurs when a small amount of water is present in the silica glass. Water in silica glass forms a silicon-hydroxyl (Si—OH) bond, which can cause absorption in the spectral regime around $\lambda = 1.4$ μm.

Although the intrinsic absorption due to the vibration of the SiO_2 is relatively insignificant in the spectral regime of optical communications, propagation attenuation is still present in pure silica fibers. The attenuation is mainly due to manufacturing imperfection of the fiber structure, as well as the fundamental process of Rayleigh scattering. Probably the single most important factor responsible for the emergence of silica glass optical fiber as the premium information transmission medium is the low optical propagation losses in such fibers. Figure 3.21 shows the measured losses as a function of wavelength of a high-quality, GeO_2-doped single-mode fiber. The loss peak at around 1.4 μm is due to residual OH contamination of the glass. A low value of loss ~ 0.2 dB/km obtains near $\lambda = 1.55$ μm. In 1986, a low transmission attenuation of 0.154 dB/km was realized [23]. Recently, a pure silica core fiber, with the index profile shown in Figure 3.20d, has been fabricated with a low transmission loss of 0.1484 dB/km at $\lambda = 1570$ nm [20, 21]. Consequently, this region of the spectrum is now favored for long-distance optical communications. Recent experiments have taken advantage of the small pulse spreading near the zero group velocity dispersion wavelength and the low losses to demonstrate high-data-rate transmission (data rate exceeding 400 Mb/s) over a propagation path exceeding 100 km [24, 25] at $\lambda \sim 1.55$ μm. Longer distance and higher data rates are possible provided optical amplifiers are employed to boost the energy attenuation due to propagation as well as chromatic dispersion compensation to eliminate the pulse distortion. For a more detailed discussion of propagation effects in optical fibers, the student can consult Reference [26].

Another source of attenuation in single-mode fibers is the bending loss. This is usually not a problem in the transmission optical networks where the curvature of bending is very small. In optical fiber sensors such as fiberoptic gyros, a long segment of fibers (kilometers) must be coiled in a small box of a few centimeters. The radius of curvature in this case is in the range of centimeters. In Dense Wavelength Division Multiplexing(DWDM) optical networks, there are situations when many passive or active components must be connected by fibers inside a

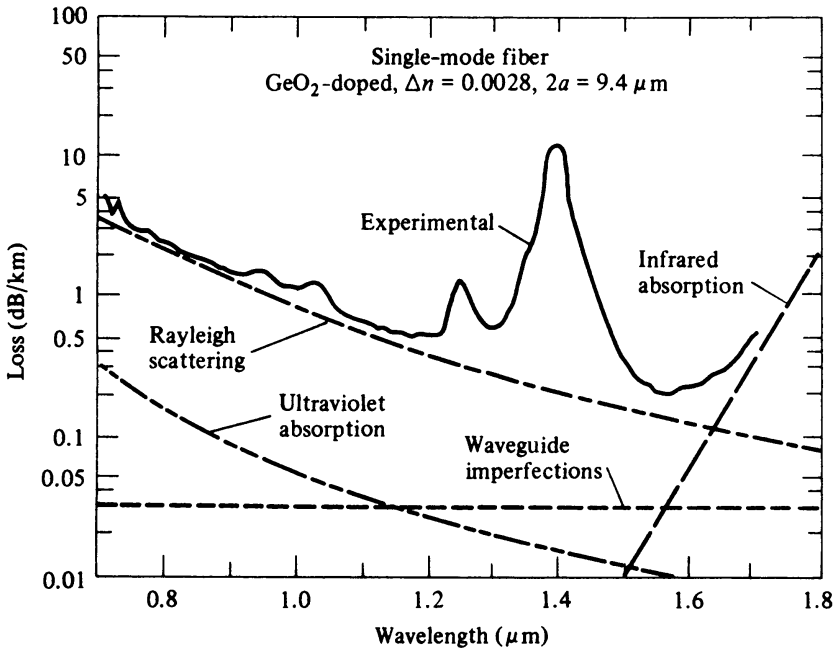


Figure 3.21 Observed loss spectrum of a germanosilicate single-mode fiber. Estimated loss spectra for various intrinsic material effects and waveguide imperfections are also shown. (From Reference [24].)

small box. Bending of fibers is unavoidable in these situations. To minimize the loss due to bending, the radius of curvature must be kept as large as possible inside the box. In this section we briefly discuss the loss due to bending of fibers.

Referring to Figure 3.22, we consider the propagation of a LP_{01} mode inside a single-mode fiber. In the straight fiber, the field distribution will remain unchanged as it propagates along the fiber. When the fiber is bent with a radius of curvature of R, we may assume that the fiber mode pattern remains the same, provided the radius of curvature of bending is much larger than the fiber core radius a (i.e., $a \ll R$). It is known that bent waveguides are intrinsically leaky. The source of leak can be explained as follows. To maintain the same mode pattern in the bent waveguide, the planar wavefront of the mode must be pivoted around the center of curvature O. This leads to a velocity mismatch problem. To keep up with the mode, the wavefront on the outer evanescent tail must travel faster than the wavefront at the center of the

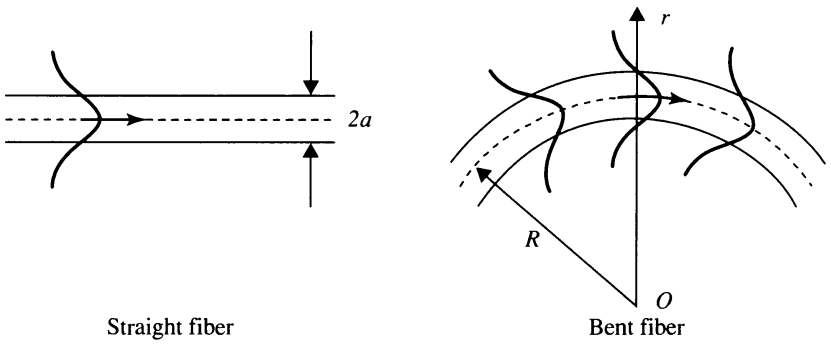


Figure 3.22 Schematic drawing of a straight fiber and a bent fiber. To maintain the mode pattern in the bent fiber, the wavefront of the outer tail of the mode must travel faster than the wavefront of the mode at the center of the fiber core. R is the radius of curvature of the bending and r is the radial distance measured from the center of curvature O.

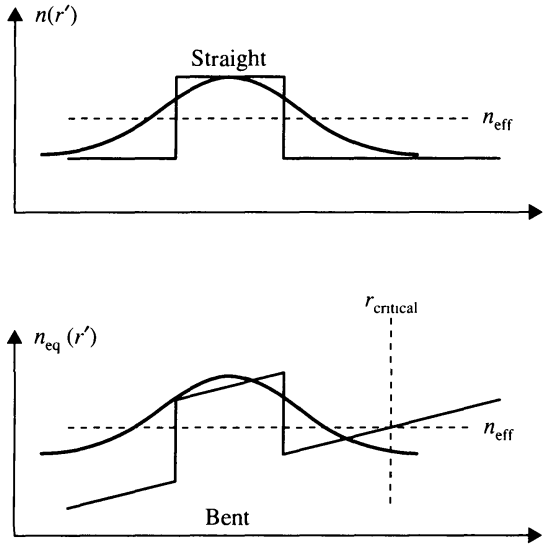


Figure 3.23 Index profile and mode wavefunction of the straight waveguide and the equivalent index profile of the bent waveguide. The horizontal dashed line indicates the effective index of the fiber mode, n_{eff}.

core. In fact, the speed of the wavefront at location r' from the center of the fiber core must be $(1 + r'/R)$ times c/n_{eff}. At some critical distance from the core of the fiber, this speed can be greater than c/n_2. Since this is impossible, the field beyond this critical distance is coupled to the radiation modes. Thus part of the mode energy will break away from the guide. This leads to propagation attenuation.

Based on the above argument and the schematic drawing shown in Figure 3.23, the wave propagation in a bent waveguide can be described by the wave propagation in an equivalent waveguide. Since the outer side of the mode must travel more distance and experience more phase shift for a given propagation constant β, the index profile of the equivalent waveguide can be written

$$n_{\text{eq}}(r') = n(r')\left(1 + \frac{r'}{R}\cos\phi'\right) \tag{3.6-3}$$

where r' is the radial distance measured from the center of the fiber core, and ϕ' is the azimuthal angle measured in the cross-sectional plane of the fiber. Note that the radial coordinate r is measured from the center of curvature of the bending O. Figure 3.23 is a sketch of the equivalent index profile. The horizontal dashed line indicates the effective index of the fiber mode, n_{eff}. Based on the discussion earlier in the chapter, the effective index of a confined mode must be greater than the index of the clad. Beyond the critical radius (r_{critical}), the equivalent index in the clad is higher than the effective index n_{eff}. This is the region where the guide mode is coupled to the radiation mode. A guided mode in the fiber can thus tunnel through the triangular barrier between the outer edge of the core and the critical radius.

As a result of the equivalent index profile shown in Figure 3.23, all modes of propagation in the bent waveguide are leaky. Solving the wave equation with the equivalent index profile given by Equation (3.6-3) would require extensive numerical techniques. The attenuation coefficient due to the bending of the fibers is usually obtained by using various different approaches [27–41]. Here we describe a result computed with the help of scalar diffraction theory [27]. In this approach, we consider the propagation of light inside the core of a circular ring of fiber. The radius of the ring R is much greater than the core radius of the fiber

a ($a \ll R$). A guided mode of the fiber is propagating in the ring. We can regard the field of the guided mode at the surface of the ring of fiber as the source of a field whose values at all points in space can be obtained by the well-known scalar diffraction theory. This approach is particularly useful in this case, as we are interested in the radiation of power into the far field. Furthermore, the scalar diffraction integral is reduced to the simple Fraunhofer diffraction integral. Based on scalar diffraction theory, the field $\psi(\mathbf{r})$ at an arbitrary point of observation $\mathbf{r}$ in the far field can be written

$$\psi(\mathbf{r}) = \frac{e^{-ikr}}{4\pi r} \int_S e^{i\mathbf{k}\cdot\mathbf{r}'} [i\mathbf{k} \cdot \mathbf{n}\psi(\mathbf{r}') - \mathbf{n} \cdot \nabla'\psi(\mathbf{r}')]\, da' \tag{3.6-4}$$

where $\mathbf{k}$ is the wavevector in the direction of observation, $\mathbf{r}'$ is the coordinate of an arbitrary point on the surface of the ring of fiber S "seen" by an observer at point $\mathbf{r}$, $\mathbf{n}$ is an outward normal to the surface S, and $\psi(\mathbf{r}')$ is the field of the mode of propagation at point $\mathbf{r}'$ on the surface S [42]. For optical fibers with, $n_{\text{core}} - n_{\text{clad}} \ll 1$, $\psi(\mathbf{r}')$ is simply the LP modes derived in Section 3.3. For planar surfaces in the xy plane, Equation (3.6-4) reduces to the Fraunhofer approximation of Equation (2.12-32).

The power flow of the radiated energy can be written in terms of the field (3.6-4):

$$\Delta S_r = \frac{1}{2}\text{Re}\left(\psi^* \frac{i}{\omega\mu}\frac{\partial}{\partial r}\psi\right) = \frac{n_2 k_0}{2\omega\mu}|\psi(\mathbf{r})|^2 \tag{3.6-5}$$

where we assume that the cladding of the fiber is infinite (i.e., $b = \infty$ in Figure 3.10), $n_2 = n_{\text{clad}}$ is the clad index, and the position $\mathbf{r}$ is very far from the ring of fiber (i.e., $R \ll r$). The attenuation coefficient is then obtained from the following equation:

$$\alpha = \frac{1}{2\pi RP} \int_{r\to\infty} \Delta S_r r^2\, d\Omega \tag{3.6-6}$$

where $d\Omega$ is an element of solid angle, r is the distance of observation measured from the center of the ring, R is the radius of the ring, and P is the mode power in the fiber as given by Equation (3.3-40). It can easily be shown that the Poynting power flow ΔS_r at the far field drops like $1/r$. So the integral converges as $r \to \infty$. Using the method of steepest descent (the saddle point method) as well as the assumption of $1 \ll qa \ll qR$, an approximate expression for the attenuation coefficient is obtained [27]:

$$\alpha = \frac{e^{2qa}}{\sqrt{\pi qa}}\left(\frac{ha}{n_2 V}\right)^2 \frac{1}{\sqrt{Ra}} \exp\left(-\frac{2(qa)^3}{3(\beta a)^2}\frac{R}{a}\right) \tag{3.6-7}$$

where R is the radius of curvature of the bending, a is the core radius of the fiber, and q, h, V, and β are the fiber and mode parameters defined in Section 3.3. We note that qa, ha, V, and βa are all dimensionless parameters. According to Equation (3.6-7), the attenuation coefficient due to bending decays exponentially as a function of the radius of curvature of the bending. As an example, Figure 3.24 is a plot of Equation (3.6-7) as a function of the radius of curvature of bending of a single-mode fiber. We note that the bending loss is significant only when the radius of bending is less than a few centimeters.

According to Equation (3.6-7), fiber modes with a larger qa will have a smaller bending loss. This can be achieved by designing fibers with a larger index step $(n_1 - n_2)$. In fiberoptic gyros or planar light circuits (PLCs), where the radius of curvature of bending is in the range of centimeters, it is important to use waveguides with a larger index step to minimize the bending loss.

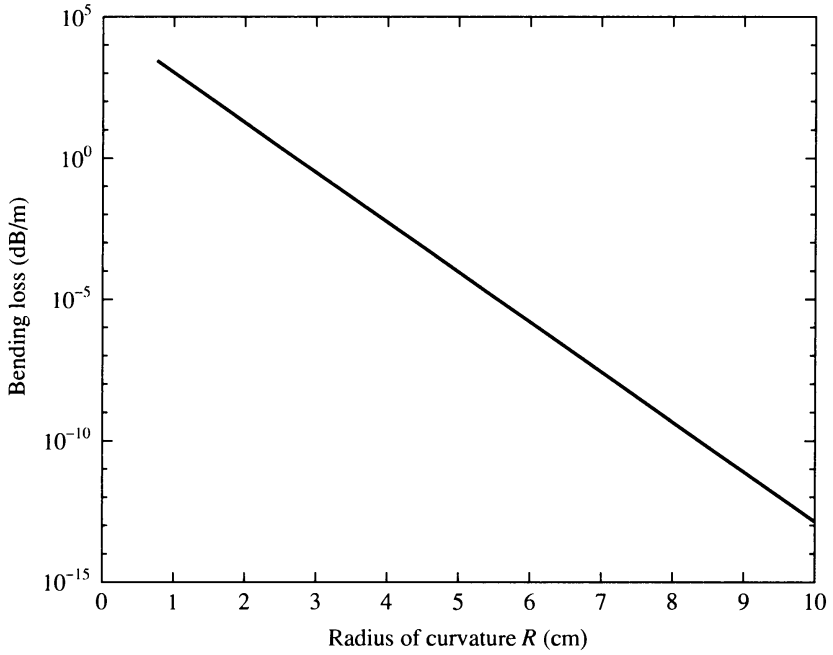


Figure 3.24 Attenuation coefficient due to bending as a function of the radius of curvature of bending, calculated using Equation (3.6-7). The fiber parameters used are: core index $n_1 = 1.4628$, clad index $n_2 = 1.4600$, core radius $a = 5.49$ μm, $\lambda = 1.30$ μm. These parameters lead to $V = 2.400$, $n_{\text{eff}} = 1.4614$, $ha = 1.6978$, $qa = 1.6969$, and $\beta a = 38.777$.

PROBLEMS

3.1

(a) Show that in the layers where $\beta^2 > n^2(\omega/c)^2$, the wavefunction $E(x)$ can have, at most, one zero crossing. [Hint: Use the continuous nature of $E(x)$ and $(1/E)(\partial^2 E/\partial x^2) > 0$.]

(b) Show that local maxima of $|E(x)|$ can only occur in the layers where $\beta^2 < n^2(\omega/c)^2$.

3.2 Let Γ_i be the fraction of power flowing in medium i ($i = 1, 2, 3$) of a slab waveguide. In particular, Γ_2 is the fraction of mode power flowing in the guiding layer n_2 and is often called the mode confinement factor. If the mode is normalized to a power of 1 W, the Γ values are defined as

$$\Gamma_i = \tfrac{1}{2}\text{Re}\int_i (\mathbf{E} \times \mathbf{H}^*) \cdot \hat{\mathbf{z}}\, dx, \quad i = 1, 2, 3$$

where the integral is over the region occupied by medium i where the index of refraction is n_i.

(a) Show that, for TE modes,

$$\Gamma_1 = \frac{1/q}{t + 1/q + 1/p}\frac{h^2}{h^2 + q^2}$$

$$\Gamma_2 = \frac{t}{t + 1/q + 1/p} + \frac{1/p}{t + 1/q + 1/p}\frac{p^2}{h^2 + p^2} + \frac{1/q}{t + 1/q + 1/p}\frac{q^2}{h^2 + q^2}$$

$$\Gamma_3 = \frac{1/p}{t + 1/q + 1/p}\frac{h^2}{h^2 + p^2}$$

Note that $\Gamma_1 + \Gamma_2 + \Gamma_3 = 1$. This proves Equation (3.2-7).

(b) Show that, for TM modes,

$$\Gamma_1 = \frac{1/q'}{t' + 1/q' + 1/p'}\frac{h^2}{h^2 + q^2}$$

$$\Gamma_2 = \frac{t'}{t' + 1/q' + 1/p'} + \frac{1/p'}{t' + 1/q' + 1/p'}\frac{p^2}{h^2 + p^2} + \frac{1/q'}{t' + 1/q' + 1/p'}\frac{q^2}{h^2 + q^2}$$

$$\Gamma_3 = \frac{1/p'}{t' + 1/q' + 1/p'}\frac{h^2}{h^2 + p^2}$$

where

$$t' = \frac{t}{n_2^2}, \quad \frac{1}{q'} = \frac{q^2 + h^2}{\tilde{q}^2 + h^2}\frac{1}{n_1^2 q}, \quad \frac{1}{p'} = \frac{p^2 + h^2}{\tilde{p}^2 + h^2}\frac{1}{n_3^2 p}$$

(c) Show that Γ_2 increases as β increases, and

$$\Gamma_2 \to 1, \quad \text{as} \quad \beta \to \frac{n_2\omega}{c}$$

Thus lower-order modes are more confined than higher-order modes because lower-order modes have a larger propagation constant. [Hint: Use $h^2 + p^2 = (n_2^2 - n_3^2)(\omega/c)^2$ and $h^2 + q^2 = (n_2^2 - n_1^2)(\omega/c)^2$.]

(d) Compare the mode confinement factor Γ_2 for TE_m and TM_m modes.

(e) Show that in the limit of well confinement (i.e., $h \to 0$),

$$\frac{\Gamma_3^{TM}}{\Gamma_3^{TE}} = \frac{1 + p/q + pt}{1 + (n_1^2/n_3^2)(p/q) + (n_2^2/n_3^2)pt}$$

where $p = (\omega/c)(n_2^2 - n_3^2)^{1/2}$ and $q = (\omega/c)(n_2^2 - n_1^2)^{1/2}$.

3.3 The interface between a metal and a dielectric can support the propagation of a surface wave, known as a surface plasmon, provided the metal has a negative dielectric constant.

(a) If the metal is viewed as a semi-infinite free electron gas with a dielectric constant given by

$$\varepsilon = \varepsilon_0\left(1 - \frac{\omega_p^2}{\omega^2}\right)$$

where ω_p is the *plasma frequency*, show that *surface plasmon* modes exist only when $\omega^2 < \frac{1}{2}\omega_p^2$. Here, we assume that $n_1^2 = 1$ for the dielectric. For an electron gas with an electron number density of N, the plasma frequency is $\omega_p^2 = Ne^2/m\varepsilon_0$, where m is the electron mass.

(b) Show that the propagation constant β is given by

$$\beta = \left(\frac{n_1^2 n_3^2}{n_1^2 + n_3^2}\right)^{1/2}\frac{\omega}{c}$$

[Hint: Use Equation (3.2-11). Show that β is always greater than $n_1 k_0$ (i.e., $\beta > n_1 k_0$), provided $n_1^2 > 0$ and $n_1^2 + n_3^2 < 0$. Here, $k_0 = \omega/c$.]

(c) Show that $\beta^2 = pq$ and that the polarization states of the **E** vector in the two media are mutually orthogonal.

(d) Obtain an expression for the z component of the Poynting vector. Show that the Poynting power flow in the positive and negative dielectric media are opposite in direction. Thus a surface plasmon wave propagating along the silver surface along the $+z$ direction will have a negative Poynting power flow in the silver and a positive Poynting power flow in the air.

(e) The complex refractive index of gold at $\lambda = 10\ \mu m$ is $n - i\kappa = 7.4 - i53.4$. Find the propagation constant $\beta^{(0)}$ and the attenuation coefficient α of the surface plasmon wave.

3.4 The number of confined modes that can be supported by a circular dielectric waveguide depends on the refractive index profile and the wavelength.

(a) Using the cutoff value for the LP_{lm} mode, show that the mode subscripts (l, m) for a step-index fiber must satisfy the condition

$$m\pi + \left(l - \frac{3}{2}\right)\frac{\pi}{2} \le V$$

where $V = k_0 a(n_1^2 - n_2^2)^{1/2}$. Show that each LP_{lm} mode is fourfold degenerate.

(b) By counting the allowed mode subscripts (l, m), show that the total number of confined modes that can be supported by a step-index fiber is

$$N \cong \frac{4}{\pi^2}V^2 \cong \frac{1}{2}V^2$$

(c) Using Equation (2.9-15), show that for a truncated quadratic-index fiber ($n(r) = n_2$, for $r > a$)

$$N \cong \tfrac{1}{4}V^2$$

Note that the total number of modes in a truncated quadratic-index fiber is one-half that of a step-index fiber.

(d) Estimate the number of confined modes in a multimode step-index fiber with $a = 50\ \mu m$, $n_1 = 1.52$, and $n_2 = 1.50$ at a carrier wavelength of $\lambda = 1\ \mu m$.

(e) In a general, truncated graded-index fiber with a core radius a and a cladding index n_2, it is convenient to define an effective V number such that

$$V_{\text{eff}}^2 = 2k_0^2\int_0^a [n^2(r) - n_2^2]r\,dr$$

and the number of confined modes is approximately given by

$$N \cong \tfrac{1}{2}V_{\text{eff}}^2$$

Show that this approximation agrees with (b) and (c) for step-index and quadratic-index fibers, respectively.

(f) Show that, according to (e), the number of confined modes in a power-law (power p) graded-index fiber is given by

$$N = \frac{1}{2(1+2/p)}V^2 = \frac{k_0^2 a^2}{2(1+2/p)}(n_1^2 - n_2^2)$$

Show that this expression again agrees with the results obtained in (b) for step-index fibers ($p = \infty$) and in (c) for quadratic-index fibers ($p = 2$).

3.5 The numerical aperture (NA) is a measure of the light-gathering capability of a fiber. It is defined as the sine of the maximum external angle of the entrance ray (measured with respect to the axis of the fiber) that is trapped in the core by total internal reflection.

(a) Show that

$$NA = n_1 \sin\theta_1 = (n_1^2 - n_2^2)^{1/2}$$

(b) Show that the solid acceptance angle in air is

$$\Omega = n\pi(n_1^2 - n_2^2)^{1/2} = \pi(NA)^2$$

(c) Show that the solid angle (in air) for a single electromagnetic radiation mode leaving or entering the core aperture is

$$\Omega_{\text{mode}} = \frac{\lambda^2}{\pi a^2}$$

(d) The total number of modes that the fiber can support, couple to, and radiate into air is therefore

$$N = 2\frac{\Omega}{\Omega_{\text{mode}}}$$

where the factor of 2 accounts for the two independent polarizations in air. Show that this estimate agrees with Problem 3.4.

(e) Find the numerical aperture of a multimode fiber with $n_1 = 1.52$ and $n_2 = 1.50$.

3.6 A single-mode step-index fiber must have a V number less than 2.405; that is,

$$V = k_0 a(n_1^2 - n_2^2)^{1/2} < 2.405$$

(a) Show that the expression derived in Problem 3.4(b) ($N \cong 4V^2/\pi^2$) still applies, provided we realize that a single-mode fiber supports two independently polarized HE_{11} modes (or LP_{01} modes).

(b) With $a = 5\ \mu m$, $n_2 = 1.50$, and $\lambda = 1\ \mu m$, find the maximum core index for a single-mode fiber. [Answer: $n_1 = 1.50195$.]

(c) With $n_1 = 1.501$, $n_2 = 1.500$, and $\lambda = 1\ \mu m$, find the maximum core radius for a single-mode fiber. [Answer: $a = 7\ \mu m$.]

(d) Show that the confinement factor for a single-mode fiber is

$$\Gamma_1 = \frac{P_{\text{core}}}{P} = \frac{(qa)^2}{V^2}\left(1 + \frac{J_0^2(ha)}{J_1(ha)}\right)$$

where ha satisfies the mode condition (3.3-26)

$$ha\frac{J_1(ha)}{J_0(ha)} = qa\frac{K_1(qa)}{K_0(qa)}$$

(e) Show that, by using the table of Bessel functions, $ha = 1.647$ is an approximate solution to the mode condition for $V = 2.405$. Evaluate the confinement factor Γ_1 for the LP_{01} mode of this single-mode fiber [Answer: $\Gamma_1 = 83\%$.] Note that this is the maximum confinement factor for a single-mode fiber. Compare this value with the curves in Figure 3.12.

3.7

(a) Derive the mode condition for step-index fibers (Equation (B-11)).

(b) Derive the expressions for the constants B, C, and D in terms of A (Equation (B-12)).

(c) Derive the mode condition for TE and TM modes (Equations (B-17a) and (B-17b)).

(d) Show that $E_z = E_r = 0$ for TE modes and $H_z = H_r = 0$ for TM modes.

(e) Show that in the limit $n_1 - n_2 \ll n_1$, TE and TM modes become identical.

3.8

(a) Derive Equations (3.3-6) and (3.3-7).

(b) Derive Equations (3.3-18) and (3.3-19).

(c) Derive the field components (3.3-23).

(d) Derive Equations (3.3-24) and (3.3-25).

3.9 There are several approximations for an analytic form of the effective index of the fundamental mode of a symmetric slab waveguide. Using the definition $u = hd/2$ and $v = qd/2$, the mode condition for the fundamental mode can be written $\tan u = v/u$.

(a) Show that the mode condition can also be written

$$u = V\cos u, \quad \text{or} \quad v = V\sin u, \quad \text{or} \quad \tan u = \frac{\sin u}{u}V$$

where $V = \sqrt{n_2^2 - n_1^2}\,(\pi d/\lambda)$ for the slab waveguide.

(b) If $u \ll 1$, then $u = \tan^{-1}V$ is an approximation. Show that the following is also an approximation:

$$u = \tan^{-1}\left(\frac{V}{\tan^{-1}V}\frac{V}{\sqrt{1+V^2}}\right)$$

(c) The following is also a useful approximation [43]:

$$u^2 = \ln\sqrt{1+2V^2}$$

Compare the approximations in (b) and (c) with the exact mode condition by plotting u versus V. Show that for $V < \pi/2$, the approximation in (c) yields more accurate results.

3.10 The propagation β of the fundamental mode (LP_{01}) satisfies the following mode condition:

$$ha\frac{J_1(ha)}{J_0(ha)} = qa\frac{K_1(qa)}{K_0(qa)}$$

In the following, we find asymptotic solutions for β in the regions $V \ll 1$ and $V \gg 1$.

(a) Using the asymptotic forms of the Bessel functions, show that, for $V \ll 1$, the mode condition reduces to

$$qa = \frac{2}{\Gamma}e^{-2/V^2}$$

or equivalently,

$$n_{\text{eff}}^2 = \left(\frac{\beta}{k_0}\right)^2 = n_2^2 + \frac{4(n_1^2 - n_2^2)}{V^2}e^{-2(2/V^2+\gamma)}$$

where γ is Euler's constant $\gamma = 0.5772$, and $\Gamma = \exp(\gamma) = 1.782$. We note that n_{eff} approaches the clad index n_2 very quickly as V approaches zero (see Figure 3.12).

(b) Using the asymptotic forms of the Bessel functions, show that, for $V \gg 1$, the mode condition reduces to

$$ha = x_1(1 - 1/V)$$

$$n_{\text{eff}}^2 = \left(\frac{\beta}{k_0}\right)^2 = n_1^2 - \frac{x_1^2}{V^2}(n_1^2 - n_2^2)\left(1 - \frac{1}{V}\right)^2$$

where $x_1 = 2.405$ is the first root of $J_0(x) = 0$. [Hint: See Figure 3.11. ha approaches x_1 for LP_{01} mode at large V.]

3.11 The confinement factor Γ_1 is defined as $\Gamma_1 = 1 - \Gamma_2$, where

$$\Gamma_2 = \frac{P_{\text{clad}}}{P} = \frac{1}{V^2}\left((ha)^2 + \frac{(qa)^2 J_l^2(ha)}{J_{l-1}(ha)J_{l+1}(ha)}\right)$$

(a) Using the asymptotic forms of the Bessel functions, show that, at cutoff ($qa \to 0$), the confinement factor is

$$\Gamma_1 = \frac{l-1}{l} = 1 - \frac{1}{l} \quad \text{for} \quad 1 \le l$$

$$\Gamma_1 = 0 \quad \text{for} \quad l = 0, 1$$

(b) At cutoff ($qa = 0$), the solutions of the wave equation in the cladding region are no longer Bessel functions. According to Equation (A-7), the wave equation for the field in the cladding region at cutoff ($qa = 0$ and $\beta = n_2 k_0$) becomes

$$\frac{\partial^2\psi}{\partial r^2} + \frac{1}{r}\frac{\partial\psi}{\partial r} + \left(-\frac{l^2}{r^2}\right)\psi = 0$$

Show that the solution for the y-polarized LP mode can be written

$$E_y = \begin{cases} AJ_l(hr)e^{il\phi}\exp[i(\omega t - \beta z)], & r < a \\ Br^{-l}e^{il\phi}\exp[i(\omega t - \beta z)], & r > a \end{cases}$$

where $B = a^l AJ_l(ha) = a^l AJ_l(V)$ due to continuity at the core boundary.

(c) Using $S_z = (\beta/2\omega\mu)|B|^2 r^{-2l}$ for the cladding region, evaluate the following integral:

$$P_{\text{clad}} = \int_0^{2\pi}\int_a^{\infty} S_z r\, dr\, d\phi$$

Show that the integral diverges for the case when $l = 0$ and 1. This is consistent with the zero confinement factor at cutoff. For $l > 1$, the integral converges to a finite value, leading to a finite confinement factor at cutoff.

3.12 The mode condition for the LP modes is given by Equation (3.3-26),

$$ha\frac{J_{l+1}(ha)}{J_l(ha)} = qa\frac{K_{l+1}(qa)}{K_l(qa)}$$

(a) Differentiating both sides of the mode condition, show that

$$ha\left(\frac{J_{l+1}(ha)J_{l-1}(ha)}{J_l^2(ha)} - 1\right)\frac{d(ha)}{dV}$$

$$+ qa\left(\frac{K_{l+1}(qa)K_{l-1}(qa)}{K_l^2(qa)} - 1\right)\frac{d(qa)}{dV} = 0$$

[Note: V is proportional to frequency ω.]

(b) Using $(qa)^2 + (ha)^2 = V^2$ and the results in (a), show that

$$qa\frac{d(qa)}{dV} = \Gamma_1 V \quad \text{and} \quad ha\frac{d(ha)}{dV} = \Gamma_2 V$$

where Γ_1 is the confinement factor and $\Gamma_2 = 1 - \Gamma_1$:

$$\Gamma_1 = \frac{(qa)^2}{V^2}\left(1 - \frac{J_l^2(ha)}{J_{l-1}(ha)J_{l+1}(ha)}\right),$$

$$\Gamma_2 = \frac{1}{V^2}\left((ha)^2 + \frac{(qa)^2 J_l^2(ha)}{J_{l-1}(ha)J_{l+1}(ha)}\right)$$

(c) Using the results in (a) and (b) and the definition of the effective index $n_{\text{eff}} = \beta/(\omega/c)$, show that

$$\left(\frac{\partial n_{\text{eff}}}{\partial \omega}\right)_{\text{waveguide}} = \frac{1}{\omega n_{\text{eff}}}(\Gamma_1 n_1^2 + \Gamma_2 n_2^2 - n_{\text{eff}}^2)$$

It is important to note that the right side is always positive.

(d) Show that

$$\left(\frac{\partial n_{\text{eff}}}{\partial \omega}\right)_{\text{waveguide}} = \frac{1}{\omega n_{\text{eff}}}(n_{\text{eff}}^2 - n_2^2)\left(-\frac{J_m^2(ha)}{J_{m-1}(ha)J_{m+1}(ha)}\right)$$

Show that the right side is always positive for the modes of propagation.

3.13

(a) Solve for the mode of propagation along the z axis by assuming $n = 1$ (vacuum). The partial differential equation is separable. You may assume a solution of the form

$$E_z(r, t) = AJ_l(hr)\exp[i(\omega t + l\phi - \beta z)]$$

$$H_z(r, t) = BJ_l(hr)\exp[i(\omega t + l\phi - \beta z)]$$

and use

$$E_r = \frac{-i\beta}{\omega^2\mu\varepsilon - \beta^2}\left(\frac{\partial}{\partial r}E_z + \frac{\omega\mu}{\beta}\frac{\partial}{r\,\partial\phi}H_z\right),$$

$$E_\phi = \frac{-i\beta}{\omega^2\mu\varepsilon - \beta^2}\left(\frac{\partial}{r\partial\phi}E_z - \frac{\omega\mu}{\beta}\frac{\partial}{\partial r}H_z\right)$$

$$H_r = \frac{-i\beta}{\omega^2\mu\varepsilon - \beta^2}\left(\frac{\partial}{\partial r}H_z - \frac{\omega\varepsilon}{\beta}\frac{\partial}{r\partial\phi}E_z\right),$$

$$H_\phi = \frac{-i\beta}{\omega^2\mu\varepsilon - \beta^2}\left(\frac{\partial}{r\partial\phi}H_z + \frac{\omega\varepsilon}{\beta}\frac{\partial}{\partial r}E_z\right)$$

(b) Obtain expressions for the energy density and Poynting vector. Consider TE, TM, and hybrid modes.

(c) Sketch the energy density and the Poynting vector components (S_ϕ, S_z) as functions of r, for $l = 0, \pm1, \pm2$. Show that for $l \neq 0$, the Poynting power flows spirally around the z axis, with a handedness that depends on the sign of l.

(d) Using the angular momentum defined as $\mathbf{L} = \mathbf{r} \times \mathbf{P}$, where $\mathbf{P}$ is the linear momentum defined in Problem 1.5, calculate the time-averaged component of the angular momentum along the direction of propagation ($+z$). Show that this component of the angular momentum is $l\hbar$ provided the energy of the electromagnetic wave is normalized to $\hbar\omega$.

(e) What are the mode orthogonal relationships?

3.14

Using Equations (3.1-14) and (3.1-15), the mode condition can be written

$$[h\tan(\tfrac{1}{2}hd) - q][h\cot(\tfrac{1}{2}hd) + q] = 0$$

Show that the above equation reduces to $\tan(hd) = 2hq/(h^2 - q^2)$. [Hint: $\tan 2\theta = 2\tan\theta/(1 - \tan^2\theta)$.]

3.15

(a) Using Equations (3.3-16) and (3.3-17), show that

$$\int rJ_l^2(\alpha r)\,dr = \frac{r^2}{2}[J_l^2(\alpha r) - J_{l-1}(\alpha r)J_{l+1}(\alpha r)]$$

(b) Similarly, show that

$$\int rK_l^2(\alpha r)\,dr = \frac{r^2}{2}[K_l^2(\alpha r) - K_{l-1}(\alpha r)K_{l+1}(\alpha r)]$$

where α is an arbitrary constant. [Hint: Differentiate both sides with respect to r.]

REFERENCES

1. Yariv, A., and P. Yeh, *Optical Waves in Crystals*. Wiley, New York, 1984.
2. Yeh, P., *Optical Waves in Layered Media*. Wiley, New York, 1988.
3. Kao, C. K., and T. W. Davies, Spectroscopic studies of ultra low loss optical glasses. *J. Sci. Instrum.* **1**(2):1063 (1968).
4. Kapron, F. P., D. B. Keck, and R. D. Maurer, Radiation losses in glass optical waveguides. *Appl. Phys. Lett.* **17**:423 (1970).
5. Gloge, D., Weakly guiding fibers. *Appl. Opt.* **10**:2252 (1971).
6. See, for example, D. Marcuse, *Theory of Dielectric Optical Waveguides*. Academic Press, New York, 1974.
7. See, for example, I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. Academic Press, New York, 1965, p. 634, Equation (5.54-2).

8. Payne, D. N., and W. A. Gambling, Zero material dispersion in optical fibers. *Electron. Lett.* **11**:176 (1975).
9. Cohen, L. G., and C. Lin, Pulse delay measurements in the zero material dispersion wave length region for optical fibers. *Appl. Opt.* **12**:3136 (1977).
10. Li, T., Structures, parameters, and transmission properties of optical fibers. *Proc. IEEE* **68**:1175 (1980).
11. Cohen, L. G., C. L. Mammel, and H. M. Presby, Correlation between numerical predications and measurements of single-mode fiber dispersion characteristics. *Appl. Opt.* **19**:2007 (1980).
12. Tsuchiya, H., and N. Imoto, Dispersion-free single mode fibers in 1.5 μm wavelength region. *Electron. Lett.* **15**:476 (1979).
13. Cohen, L. G., C. Lin, and W. G. French, Tailoring zero chromatic dispersion into the 1.5–1.6 μm low-loss spectral region of single-mode fibers. *Electron. Lett.* **15**:334 (1979).
14. White, K. I., and B. P. Nelson, Zero total dispersion in step-index monomode fibers at 1.30 and 1.55 μm. *Electron. Lett.* **15**:396 (1979).
15. Gambling, W. A., H. Matsumara, and C. M. Ragdale, Zero total dispersion in graded-index single-mode fibers. *Electron. Lett.* **15**:474 (1979).
16. Bloom, D. M., L. F. Mollenauer, Chinlon Lin, and A. M. Del Gaudio, Demonstration of pulse propagation in km-length fibers. *Opt. Lett.* **4**:297 (1979).
17. Monroe, M., Propagation in doubly clad single mode optical fibres. *IEEE J. Quantum Electron.* **18**(4):535 (1992).
18. Li, Y., C. D. Hussey, and T. A. Birks, Triple-clad single-mode fibers for dispersion shifting. *IEEE J. Lightwave Technol.* **11**:1812 (1993).
19. Li, Y. and C. D. Hussey, Triple-clad single-mode fibers for dispersion flattening. *Opt. Eng.* **33**:3999 (1994).
20. Nagayama, K., T. Saitoh, M. Kakui, K. Kawasaki, M. Matsui, H. Takamizawa, H. Miyaki, Y. Ooga, I. Tsuchiya, and Y. Chigusa, Ultra low loss pure silica core fiber and its impact on submarine transmission systems. *Tech. Digest of OFC2002*, PD FA10 (2002).
21. Nagayama, K., M. Kakui, M. Matsui, T. Saitoh, and Y. Chigusa, Ultra low loss pure silica core fiber and extension of transmission distance. *Electron. Lett.* **38**:1168 (2002).
22. Yeh, P., A. Yariv, and E. Marom, Theory of Bragg fibers. *J. Opt. Soc. Am.* **68**(9):1196 (1978).
23. Kanamori, H., H. Yokota, G. Tanaka, M. Watanabe, Y. Ishiguro, I. Yoshida, T. Kakii, S. Itou, Y. Asano, and S. Tanaka, Transmission characteristics and reliability of pure silica core single-mode fibers. *J. Lightwave Technol.* **4**(8):1144 (1986).
24. Miya, T., Y. Terunuma, T. Hosaka, and T. Miyashita, Ultimate low-loss single-mode fiber at 1.55 μm. *Electron. Lett.* **15**:106 (1979).
25. Suematsu, Y., Long wavelength optical fiber communication. *Proc. IEEE* **71**:692 (1983).
26. Miller, S., and I. P. Kaminow (eds.), *Optical Fiber Telecommunication II*. Academic Press, San Diego, 1988, Chapters 2 and 3.
27. Marcuse, D., Bend loss of slab and fiber modes computed with diffraction theory. *IEEE J. Quantum Electron.* **29**:2957 (1993).
28. Arnaud, J. A., Transverse coupling in fiber optics, Part III: Bending loss. *Bell Syst. Tech. J.* **53**:1379 (1974).
29. Marcatili, E. A. J., Bends in optical dielectric guides. *Bell Syst. Tech. J.* **48**:2103 (1969).
30. Marcuse, D., Curvature loss formula for optical fibers. *J. Opt. Soc. Am.* **66**:216 (1976).
31. Tsao, C., *Optical Fiber Waveguide Analysis*. Oxford University Press, Oxford, UK, 1992, Chapter 11.
32. Lewin, L., Radiation from curved dielectric slabs and fibers. *IEEE Trans. Microwave Theory Tech.* **22**:718 (1974).

33. Gambling, W. A., D. N. Payne, and H. Matsumura, Radiation from curved single-mode fibers. *Electron. Lett.* **12**:567 (1976).
34. Sakai, J., and T. Kimura, Bending loss of propagation modes in arbitrary-index profile optical fibers. *Appl. Opt.* **17**:1499 (1978).
35. Snyder, A. W., and J. D. Love, *Optical Waveguide Theory*. Chapman & Hall, New York, 1983.
36. Kawakami, S., M. Miyagi, and S. Nishida, Bending losses of of dielectric slab optical waveguide with double or multiple cladding: theory. *Appl. Opt.* **14**:2588 (1975).
37. Vassallo, C., Scalar-field theory and 2-D ray theory for bent single-mode weakly guiding optical fibers. *J. Lightwave Technol.* **3**:416 (1985).
38. Murakami, Y., and H. Tsuchiya, Bending losses of coated single-mode optical fibers. *IEEE J. Quantum Electron.* **14**:495 (1978).
39. Morgan, R., J. S. Barton, P. G. Harper, and J. D. C. Jones, Wavelength dependence of bending loss in monomode fibers: effect of the buffer coating. *Opt. Lett.* **15**:947 (1990).
40. Valiente, I., and C. Vassallo, New formalism for bending losses in coated single-mode optical fibers. *Electron. Lett.* **25**:1544 (1989).
41. Renner, H., Bending losses of coated single-mode fibers: a simple approach. *J. Lightwave Technol.* **10**:544 (1992).
42. See, for example, J. D. Jackson, *Classical Electrodynamics*, 3rd ed. Wiley, New York, 1999, p. 491, Equation (10-108).
43. Chen, K.-L., and S. Wang, An approximate expression for the effective index in symmetric DH lasers. *IEEE J. Quantum Electron.* **19**:1354 (1983).

ADDITIONAL READING

Collin, R. E., *Field Theory of Guided Waves*. McGraw-Hill, New York, 1960.

CHAPTER 4

OPTICAL RESONATORS

4.0 INTRODUCTION

Optical resonators, like their microwave counterparts, are used primarily for the purpose of generating monochromatic optical beams with high intensities (e.g., lasers). These resonators, also known as optical cavities, consist in most cases of two or more mirrors that are aligned to provide multiple reflections and refocusing of light waves. Under appropriate conditions, the energy of an optical beam can be "trapped" inside the resonator. In this case, the optical beam becomes a mode of the resonator. A universal measure of this property is the quality factor Q of the resonator. Q is defined by the relation [1, 2]

$$Q = \omega \times \frac{\text{field energy stored by resonator}}{\text{power dissipated by resonator}} \tag{4.0-1}$$

where ω is the frequency of the wave. As an example, consider the case of a simple resonator formed by bouncing a plane wave between two perfectly reflecting planes of separation L so that the field inside is

$$E(z, t) = E_0 \sin \omega t \sin kz \tag{4.0-2}$$

where E_0 is a constant, the z axis is perpendicular to the reflecting planes, ω is the angular frequency, and k is the wavenumber. According to the discussion in Chapter 1, the average electric energy stored in the resonator is

$$U_{\text{electric}} = \frac{A\varepsilon}{2T} \int_0^L \int_0^T E^2(z, t)\, dz\, dt \tag{4.0-3}$$

where A is the cross-sectional area, ε is the dielectric constant, and $T = 2\pi/\omega$ is the period. Using Equation (4.0-2) we obtain

$$U_{\text{electric}} = \tfrac{1}{8}\varepsilon E_0^2 V \tag{4.0-4}$$

where $V = AL$ is the resonator volume. Since the average magnetic energy stored in a resonator is equal to the electric energy [1], the total stored energy is

$$U = \tfrac{1}{4}\varepsilon E_0^2 V \tag{4.0-5}$$

Thus, recognizing that in the steady state the input power is equal to the dissipated power, and designating the power input to the resonator by P, we obtain from Equation (4.0-1)

$$Q = \frac{\omega \varepsilon E_0^2 V}{4P}$$

The peak electric field is given by

$$E_0 = \sqrt{\frac{4QP}{\omega \varepsilon V}} \tag{4.0-6}$$

We note that the electric field amplitude is proportional to the square root of Q. At a given power input, a high-Q resonator can provide a high electric field amplitude, or a high field intensity.

Mode Density in Optical Resonators

For the purpose of frequency definition, typical dimensions of resonators (or cavities) are of the order of the wavelength of the electromagnetic waves. This is important to ensure that the resonators possess a very small number (ideally only one) of high-Q modes in a given spectral region. For microwaves, the wavelengths are in the ranges of centimeters. Resonators of these dimensions can easily be obtained. In optical regimes, the resonator's dimensions need to be of order of micrometers. These micrometer optical resonators can be a manufacturing challenge.

EXAMPLE: ONE-DIMENSIONAL RESONATOR

We consider the case of a one-dimensional resonator consisting of two mirrors (also known as the Fabry–Perot resonator). The Fabry–Perot resonator is widely used as a multiple-beam interferometer, an instrument first constructed in the early 1800s by Charles Fabry and Alfred Perot [3]. The Fabry–Perot interferometer has an extremely high resolving power—about 10 times better than a grating spectrometer (which is already at least an order of magnitude better than a prism spectrometer). When the mirrors are aligned perfectly parallel to each other, the reflections of the light waves between the two mirrors can interfere constructively, giving rise to a standing wave pattern as described by Equation (4.0-2) between the mirror surfaces, just like standing waves on a string. This occurs when the mirror separation (also known as the cavity spacing or resonator length L) is an exact integer multiple of half a wavelength of light. If the resonator is made of dielectric mirrors of partial reflection, then an incident beam traveling along the axis can be "trapped" by the multiple reflections inside the resonator, provided the resonator length is an integral number of the half-wavelength of the beam. The "trapping" leads to a build up of a large field intensity inside the resonator. This phenomenon is often called resonance. The resonant frequencies are determined by requiring that the field vanishes at $z = 0$ and at the location $z = L$ of the second reflector. This happens when

$$kL = m\pi, \quad \text{with } m = 1, 2, 3, \ldots$$

Using

$$k = \frac{2\pi\nu}{c} n = \frac{2\pi}{\lambda} n$$

where n is the index of refraction and λ is the wavelength of light in vacuum, the resonance condition can be written

$$L = m\frac{\lambda}{2n}$$

where m is an integer ($m = 1, 2, 3, \ldots$). In terms of the frequency of light (ν), the resonance condition is often written

$$\nu = m\frac{c}{2nL}$$

The resonance condition ensures that the multiple reflections inside the cavity add up constructively. The frequency spacing between adjacent resonances of a one-dimensional resonator is $c/2nL$. Thus the total number of modes between 0 and ν is $4nL\nu/c$ (a factor of 2 is added to account for 2 orthogonal polarization states). For single-mode operation (actually two polarization modes), we obtain $L = \lambda/2n$. In other words, the *linear dimension needs to be comparable to the wavelength* (in the medium).

Mode control in the optical regime would thus seem to require that we construct resonators with volume $\sim \lambda^3(\sim 10^{-12}\ \text{cm}^3$ at $\lambda = 1\ \mu\text{m})$. This is not easily achievable. An alternative is to build large ($L \gg \lambda$) resonators but to use a geometry that endows only a small fraction of these modes with low losses (a high Q). In our two-mirror example, any mode that does not travel normally to the mirror will "walk off" after a few bounces and thus will possess a low Q factor. We will show later that when the resonator contains an amplifying (inverted population) medium, oscillation will occur preferentially at high-Q modes, so that the strategy of modal discrimination by controlling Q is sensible. We shall also find that further modal discrimination is due to the fact that the atomic medium is capable of amplifying radiation only within a limited frequency region so that modes outside this region, even if possessing high Q, do not oscillate.

One question asked often is the following: Given a large optical resonator ($L \gg \lambda$), how many of its modes will have their resonant frequencies in a given frequency interval, say, between ν and $\nu + \Delta\nu$? To answer this question, consider a large, perfectly reflecting box resonator with sides a, b, c along the x, y, z directions. Resonators of different shapes will differ in detail, but for large resonators ($L \gg \lambda$), the results are similar. Without going into modal details, it is sufficient for our purpose to take the amplitude field solution in the form [2]

$$E(x, y, z) = E_0 \sin k_x x \sin k_y y \sin k_z z \tag{4.0-7}$$

where the components of the wavevector must satisfy the following condition:

$$k_x^2 + k_y^2 + k_z^2 = \left(\frac{\omega}{c}n\right)^2 \tag{4.0-8a}$$

where n is the index of refraction of the cavity medium. For the field to vanish at the boundaries, we thus need to satisfy

$$k_x = \frac{r\pi}{a},\quad k_y = \frac{s\pi}{b},\quad k_z = \frac{t\pi}{c},\quad r, s, t = \text{any integers} \tag{4.0-8b}$$

With each such mode, we may thus associate a propagation vector $k = \hat{\mathbf{x}}k_x + \hat{\mathbf{y}}k_y + \hat{\mathbf{z}}k_z$. The triplet r, s, t defines a mode. Since replacing any integer with its negative does not, according to Equation (4.0-7), generate an independent mode, we will restrict, without loss of generality, r, s, t to positive integers. It is convenient to describe the modal distribution in $\mathbf{k}$ space, as in

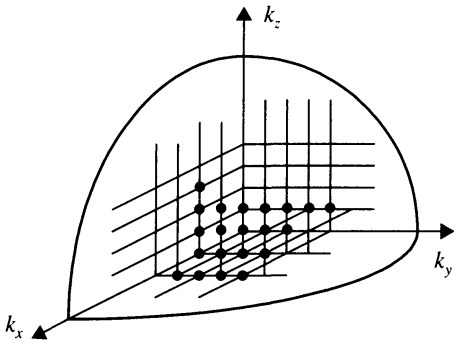


Figure 4.1 The **k**-space description of modes. Every positive triplet of integers r, s, t defines a unique mode. We can thus associate a primitive volume π^3/abc in **k** space with each mode.

Figure 4-1. Since each lattice point (r, s, t) in **k** space represents an independent mode, we can associate with each mode an elemental volume in **k** space.

$$V_{\text{mode}} = \frac{\pi^3}{abc} = \frac{\pi^3}{V} \tag{4.0-9}$$

where V is the physical volume of the resonator. We recall that the length of the vector **k** satisfies Equation (4.0-8a), rewritten here as

$$k(r, s, t) = \frac{2\pi\nu(r, s, t)}{c} n \tag{4.0-10}$$

To find the total number of modes with wavenumber between 0 and k, we divide the corresponding volume in **k** space by the volume per mode:

$$N(k) = \frac{\left(\frac{1}{8}\right)\frac{4\pi}{3}k^3}{\pi^3/V} = \frac{k^3V}{6\pi^2}$$

where the factor $\frac{1}{8}$ is due to the restriction of r, s, $t > 0$.

We next use Equation (4.0-10) to obtain the number of modes with resonant frequencies between 0 and ν:

$$N(k) = \frac{4\pi\nu^3 n^3 V}{3c^3}$$

The mode density, that is, the number of modes per unit ν near ν in a resonator with volume V ($\gg\lambda^3$), is thus

$$p(\nu) = \frac{dN(\nu)}{d\nu} = \frac{8\pi\nu^2 n^3 V}{c^3} \tag{4.0-11}$$

where we multiplied the final result by 2 to account for the two independent orthogonally polarized modes that are associated with each r, s, t triplet.

The number of modes that fall within the interval $d\nu$ centered on ν is thus

$$N \cong \frac{8\pi\nu^2 n^3 V}{c^3} d\nu \tag{4.0-12}$$

where V is the volume of the resonator. For the case of $V = 1\ \text{cm}^3$, $\nu = 3 \times 10^{14}$ Hz, and $d\nu = 3 \times 10^{10}$, as an example, Equation (4.0-12) yields $N \sim 2 \times 10^9$ modes. If the resonator were closed, all these modes would have similar values of Q. This situation is to be avoided in the case of lasers, since it will cause the atoms to emit power (thus causing oscillation) into

a large number of modes, which may differ in their frequencies as well as in their spatial characteristics.

This objection is overcome to a large extent by the use of open resonators, which consist essentially of a pair of opposing flat or curved reflectors. In such resonators the energy of the vast majority of the modes does not travel at right angles to the mirrors and will thus be lost in essentially a single traversal. These modes will consequently possess a very low Q. If the mirrors are curved, the transverse confinement is obtained automatically by the Gaussian modes: thus the diffraction losses caused by the open sides can be made small compared with other loss mechanisms such as mirror transmission. In arriving at Equation (4.0-12), we ignored the dependence of the index of refraction n on the optical frequency. As discussed in previous chapters, the index of refraction depends on the frequency ν. When the dispersion cannot be ignored, the density of optical modes becomes

$$p(\nu) = \frac{dN(\nu)}{d\nu} = \frac{8\pi\nu^2 n^2 n_g V}{c^3} \tag{4.0-13}$$

where n_g is the group index

$$n_g = n + \nu\frac{\partial n}{\partial \nu} \tag{4.0-14}$$

In this chapter, we discuss some important properties of optical resonators that are related to optical electronics. Starting with a simple plane wave analysis of the optical properties of Fabry–Perot resonators, we discuss the reflection and transmission of light through a Fabry–Perot etalon. We then discuss the property of asymmetric Fabry–Perot etalons, including Gires–Tournois etalons and ring resonators. We also discuss optical properties of coupled ring resonators, as well as multicavity etalons. To take diffraction loss into account, we consider Gaussian beams inside optical resonators consisting of spherical mirrors. The mode stability condition is then derived. Such a mode stability condition is essential in the design of stable optical resonators for laser applications. At the end of this chapter, we discuss an important concept of mode coupling in resonators.

4.1 FABRY–PEROT ETALON

The Fabry–Perot etalon, or interferometer, named after its inventors [3], can be considered as the best example of an optical resonator. It consists of a plane-parallel plate of thickness l and refractive index n that is immersed in a medium of index n'. Generally, an etalon can be obtained by spacing two partially reflecting mirrors at a distance l apart so that $n = n' = 1$. Another common form of etalon is produced by grinding two plane-parallel (or curved) faces on a transparent solid and then evaporating a metallic or dielectric layer (or layers) on the surfaces. The term etalon is often reserved for a plane-parallel plate of solid transparent material with reflecting surfaces on both sides, while interferometers (or cavities) are reserved for structures that consist of two parallel mirrors with an empty space in between.

Let a plane wave be incident on the etalon at an angle θ' to the normal, as shown in Figure 4.2. We can treat the problem of transmission (and reflection) of the plane wave through the etalon by considering the infinite number of partial waves produced by multiple reflections at the two end surfaces. The phase delay between two partial waves—which is attributable to one additional round trip—is given, according to Figure 4.3, by

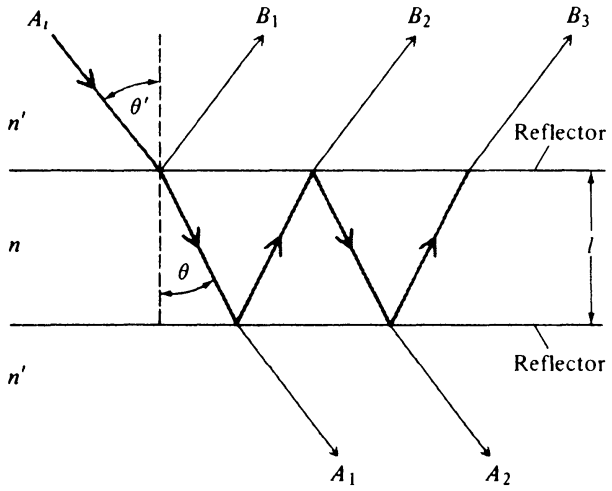


Figure 4.2 Multiple-reflection model for analyzing the Fabry–Perot etalon.

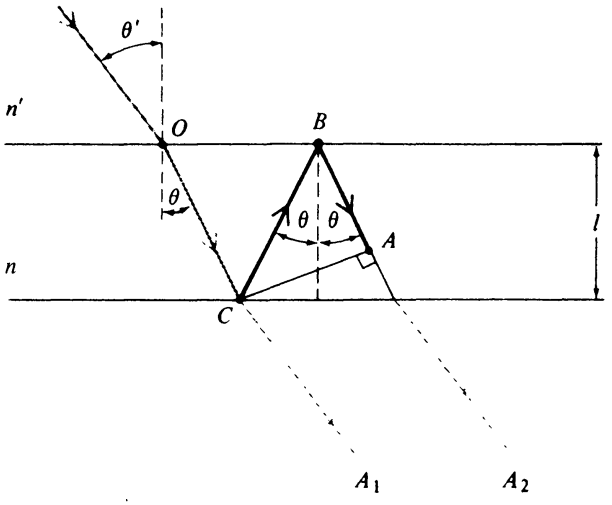


Figure 4.3 Path difference between successive reflections inside the Fabry–Perot etalon. Their path difference ΔL is given by $\Delta L = AB + BC = l(\cos 2\theta/\cos\theta) + l/\cos\theta = 2l\cos\theta$, $\rightarrow \delta = (2\pi(\Delta L)n)/\lambda = (4\pi nl\cos\theta)/\lambda$

$$\delta = \frac{4\pi nl\cos\theta}{\lambda} = 2k_x l \tag{4.1-1}$$

where λ is the vacuum wavelength of the incident wave, θ is the internal angle of incidence, and k_x is the x component of the wavevector (x axis is perpendicular to the mirrors). At normal incidence, this phase shift is simply $\delta = 2kl$, where k is the wavenumber of propagation in the medium. Thus the phase shift is often called the round-trip phase shift. If the complex amplitude of the incident wave is taken as A_i, then the partial reflection amplitudes, B_1, B_2, and so forth, are given by

$$B_1 = rA_i, \quad B_2 = tt'r'A_i e^{-i\delta}, \quad B_3 = tt'r'^3 A_i e^{-i2\delta}, \quad \ldots$$

where r is the amplitude reflection coefficient of the interface (ratio of reflected to incident amplitude), t is the amplitude transmission coefficient for waves incident from n' toward n, and r' and t' are the corresponding quantities for waves traveling from n toward n'. The complex amplitude of the (total) reflected wave is $A_r = B_1 + B_2 + B_3 + \cdots$, or

$$A_r = [r + tt'r'e^{-i\delta}(1 + r'^2 e^{-i\delta} + r'^4 e^{-2i\delta} + \cdots)]A_i \tag{4.1-2}$$

For the transmitted wave, the amplitudes of the partial waves are

$$A_1 = tt'A_i e^{-i\delta/2}, \quad A_2 = tt'r'^2 e^{-i\delta} A_i e^{-i\delta/2}, \quad A_3 = tt'r'^4 e^{-2i\delta} A_i e^{-i\delta/2}, \quad \ldots$$

where the phase factor $e^{-i\delta/2}$ corresponds to a single traversal of the plate and is common to all the terms. Adding up the A terms, we obtain

$$A_t = A_i tt'(1 + r'^2 e^{-i\delta} + r'^4 e^{-2i\delta} + \cdots)e^{-i\delta/2} \tag{4.1-3}$$

for the complex amplitude of the total transmitted wave. We notice that the terms within the parentheses in Equations (4.1-2) and (4.1-3) form an infinite geometric series; adding them, we get

$$A_r = \frac{(1 - e^{-i\delta})\sqrt{R}}{1 - Re^{-i\delta}} A_i \tag{4.1-4}$$

and

$$A_t = \frac{Te^{-i\delta/2}}{1 - Re^{-i\delta}} A_i \tag{4.1-5}$$

where we used the fact that $r' = -r$ for the dielectric interface, the conservation-of-energy relation that applies to lossless mirrors,

$$r^2 + tt' = 1$$

as well as the definitions

$$R \equiv r^2 = r'^2 \quad \text{and} \quad T \equiv tt'$$

R and T are, respectively, the fraction of the intensity reflected and transmitted at each interface and will be referred to in the following discussion as the mirrors' reflectance and transmittance. It is important to note $r' = -r$ and $r^2 + tt' = 1$ are valid only for the interface between two dielectric media. For mirrors with dielectric coatings, similar relationships exist. They will be discussed in Section 4.9.

If the incident intensity (watts per square meter) is taken as $A_i A_i^*$, we obtain from Equation (4.1-4) the following expression for the fraction of the incident intensity that is reflected by the etalon:

$$\frac{I_r}{I_i} = \frac{A_r A_r^*}{A_i A_i^*} = \frac{4R \sin^2(\delta/2)}{(1 - R)^2 + 4R \sin^2(\delta/2)} \tag{4.1-6}$$

Moreover, from Equation (4.1-5),

$$\frac{I_t}{I_i} = \frac{A_t A_t^*}{A_i A_i^*} = \frac{(1 - R)^2}{(1 - R)^2 + 4R \sin^2(\delta/2)} \tag{4.1-7}$$

for the transmitted fraction. Our basic model contains no loss mechanisms, so conservation of energy requires that $I_r + I_t$ be equal to I_i, as is indeed the case.

Let us consider the transmission characteristics of a Fabry–Perot etalon. According to Equation (4.1-7) the transmission is unity whenever

$$\delta = \frac{4\pi nl \cos\theta}{\lambda} = 2m\pi, \quad m = \text{any integer} \tag{4.1-8}$$

The condition (4.1-8) for maximum transmission can be written

$$\nu_m = m\frac{c}{2nl \cos\theta}, \quad m = \text{any integer} \tag{4.1-9}$$

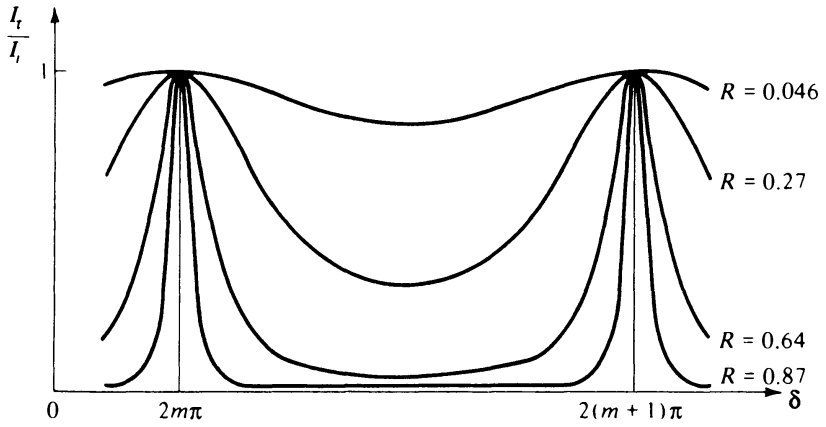


Figure 4.4 Transmission characteristics (theoretical) of a Fabry–Perot etalon. (After Reference [4].)

where $c = \nu\lambda$ is the velocity of light in vacuum and ν is the optical frequency. For a fixed l and θ, Equation (4.1-9) defines the unity transmission (resonance) frequencies of the etalon. These are separated by the *free spectral range*,

$$\Delta\nu \equiv \nu_{m+1} - \nu_m = \frac{c}{2nl\cos\theta} \tag{4.1-10}$$

Equation (4.1-10) is obtained by assuming that the index of refraction n is a constant. In the case when the spectral dispersion cannot be neglected, the free spectral range becomes

$$\Delta\nu \equiv \nu_{m+1} - \nu_m = \frac{c}{2n_g l\cos\theta} = \frac{v_g}{2l\cos\theta} \tag{4.1-11}$$

where v_g is the group velocity of light in the medium and n_g is the group index defined in Equation (4.0-14).

Theoretical transmission plots of a Fabry–Perot etalon are shown in Figure 4.4. The maximum transmission is unity, as stated previously. The minimum transmission, on the other hand, approaches zero as the mirror reflectivity R approaches unity.

If we allow for the existence of loss or gain in the etalon medium, we find that the peak transmission is no longer unity. Let the intensity gain (or loss) per pass be G, defined as $G = I_{\text{output}}/I_{\text{input}}$. We find that the intensity transmission becomes

$$\frac{I_t}{I_i} = \frac{A_t A_t^*}{A_i A_i^*} = \frac{(1-R)^2 G}{(1-GR)^2 + 4GR\sin^2(\delta/2)} \tag{4.1-12}$$

The proof of Equation (4.1-12) is left as an exercise (Problem 4.2). The maximum transmission can be greater than unity when gain is present. In fact, the maximum transmission can be much greater than unity when the product GR is approaching unity. According to Equation (4.1-12), the maximum transmission can be written

$$\left(\frac{I_t}{I_i}\right)_{\max} = \frac{(1-R)^2 G}{(1-GR)^2} \tag{4.1-13}$$

In the case of loss, G is less than 1, and the maximum transmission is less than unity.

If the loss is due to linear absorption with an absorption coefficient α in the medium, then G can be written

$$G = \exp(-\alpha l) \tag{4.1-14}$$

An experimental transmission plot of a Fabry–Perot etalon is shown in Figure 4.5.

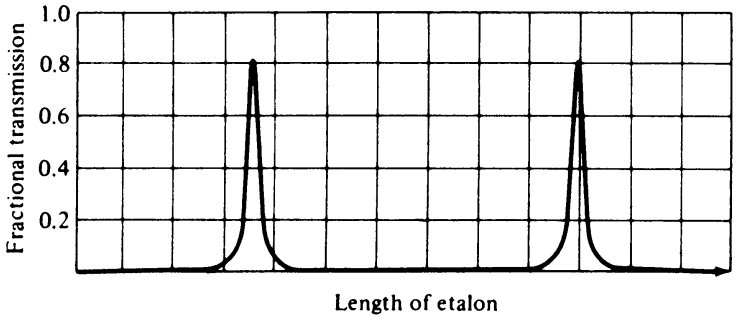


Figure 4.5 Experimental transmission characteristics of a Fabry–Perot etalon at $\lambda = 632.8$ nm as a function of the etalon optical length with $R = 0.9$ and $G = 0.98$. The two peaks shown correspond to a change in the optical length $\Delta(nl) = \lambda/2$. (After Reference [5].)

Intracavity field

We mentioned earlier that the field intensity inside the cavity can be built up significantly at resonance. We now examine the field inside the cavity. Let I_i be the intensity of the incident beam, and I_o be the intensity of the transmitted beam. Inside the Fabry–Perot etalon, the electric field can be written as the sum of a right-traveling wave with an intensity I_1 (traveling along the $+x$ direction) and a left-traveling wave with an intensity I_2 (traveling along the $-x$ direction). If the cavity medium is lossless, the Poynting vector inside the cavity can be written as the algebraic sum of the Poynting vector of these two traveling waves. Let I_1 and I_2 be the magnitudes of the Poynting vectors of these two waves, respectively. If we further assume that the mirrors are lossless, then these intensities are related by

$$I_o = \frac{(1-R)^2}{(1-R)^2 + 4R\sin^2(\delta/2)} I_i \tag{4.1-15}$$

$$I_o = TI_1 = (1-R)I_1 \tag{4.1-16}$$

$$I_2 = RI_1 \tag{4.1-17}$$

where we recall that R is the mirror reflectivity, and $T = 1 - R$ is the mirror transmittance. According to these equations, the intracavity intensities at resonance ($\delta = 2\pi, 4\pi, \ldots$) are given by

$$I_1 = \frac{1}{(1-R)} I_i, \quad I_2 = \frac{R}{(1-R)} I_i \tag{4.1-18}$$

The right-traveling wave and the left-traveling wave form an interference (standing wave) pattern inside the cavity (see Figure 4.6). The intensity pattern in the cavity can be written

$$I(x) = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(2kx + \alpha_0) \tag{4.1-19}$$

where k is the wavenumber in the cavity, and α_0 is a constant phase due to possible phase shifts upon reflection from the cavity mirrors. It can easily be shown that the average intensity inside the cavity can be written, according to Equations (4.1-18) and (4.1-19),

$$I_{\text{ave}} = \frac{1+R}{1-R} I_i \tag{4.1-20}$$

For a mirror reflectivity of $R = 0.99$, the average intracavity intensity in such a Fabry–Perot etalon is 199 times that of the incident beam at resonance when maximum transmission occurs.

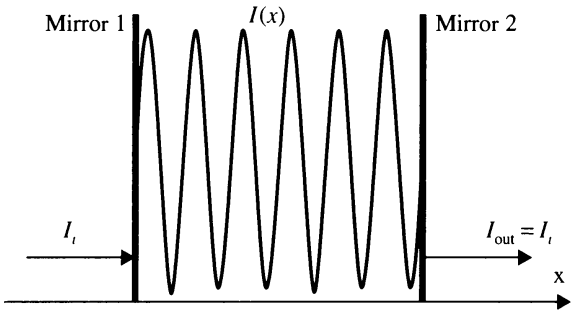


Figure 4.6 Intensity distribution inside the cavity at resonance.

As shown in Figure 4.4, the transmitted beam intensity is spectrally modulated. The transmission is maximum at resonance when the cavity length is an integral number of half-wavelengths. In addition to spectral modulation, the phase shift of the transmitted beam is also being spectrally modulated. The phase dispersion is especially strong at resonance. This is related to the build up of field intensity inside the cavity. We define ϕ as the one-pass phase shift inside the cavity (i.e., $\phi = \delta/2$). The etalon transmission coefficient can be written, according to Equation (4.1-5),

$$t_e = \frac{Te^{-i\phi}}{1 - Re^{-2i\phi}} = |t_e|e^{-i\psi} \tag{4.1-21}$$

where ψ is the phase shift of the transmitted beam. It can be shown that

$$\psi = \phi + \tan^{-1}\left(\frac{R \sin 2\phi}{1 - R \cos 2\phi}\right) \tag{4.1-22}$$

We now examine the derivative of ψ with respect to the angular frequency ω:

$$\frac{d\psi}{d\omega} = \frac{d\psi}{d\phi}\frac{d\phi}{d\omega} \tag{4.1-23}$$

Physically, the frequency derivative of the phase shift is the flight time (or group delay). If we write

$$\tau_e = \frac{d\psi}{d\omega} \quad \text{and} \quad \tau = \frac{d\phi}{d\omega} \tag{4.1-24}$$

where τ is the one-pass flight time inside the cavity medium and τ_e is the flight time through the etalon, then we obtain

$$\tau_e = \frac{d\psi}{d\phi}\tau \tag{4.1-25}$$

Thus $d\psi/d\phi$ may be viewed as an enhancement factor of the time of flight due to the etalon. By taking the derivative, we obtain

$$\frac{d\psi}{d\phi} = \left(\frac{1 + R}{1 - R}\right) \quad \text{when } \phi = \pi, 2\pi, 3\pi, \ldots \tag{4.1-26}$$

Combining the above two equations, we obtain

$$\tau_e = \left(\frac{1 + R}{1 - R}\right)\tau \quad \text{when } \phi = \pi, 2\pi, 3\pi, \ldots \tag{4.1-27}$$

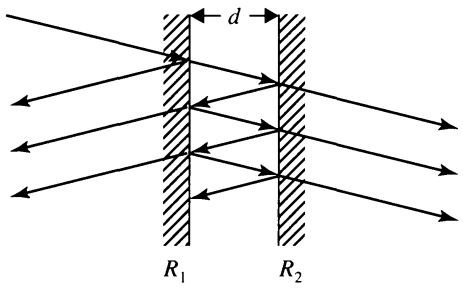


Figure 4.7 Schematic drawing of an asymmetric Fabry–Perot etalon with mirror reflectivities R_1 and R_2, and a cavity spacing of d.

We note that the time of flight (group delay) is enhanced by a factor of $(1 + R)/(1 - R)$ at resonance. For a mirror reflectivity of $R = 0.99$, the group delay at resonance is enhanced by a factor of 199. This is in agreement with the increase in the intracavity energy density at resonance. In a lossless system, the Poynting vector is a constant. If we write the Poynting vector as the product of energy density and the speed of light, then the effective speed of light must be low in the region where the energy density is high.

Asymmetric Fabry–Perot Etalons

Most Fabry–Perot etalons employed in spectral analysis are made of two identical dielectric mirrors facing each other. These mirrors are usually made of multilayer dielectric coatings on optically flat dielectric substrates. The symmetric etalons can provide a high finesse as well as a maximum transmission of 100% at specific wavelengths. Generally, an etalon can be made of two mirrors of different mirror reflectivities. Figure 4.7 shows a schematic drawing of an asymmetric Fabry–Perot etalon, with mirror reflectivities R_1 and R_2. An extreme case of the asymmetric Fabry–Perot etalon is the Gires–Tournois etalon, where $R_2 = 1$.

Using a similar analysis leading to Equations (4.1-2) and (4.1-3), we obtain

$$A_r = \{r_1 + t_1 t_1' r_2' e^{-i\delta}[1 + r_1' r_2' e^{-i\delta} + (r_1' r_2' e^{-i\delta})^2 + \cdots]\}A_i \tag{4.1-28}$$

$$A_t = t_1 t_2' e^{-i\delta/2}[1 + r_1' r_2' e^{-i\delta} + (r_1' r_2' e^{-i\delta})^2 + \cdots]A_i \tag{4.1-29}$$

with

$$\delta = \frac{4\pi n d \cos\theta}{\lambda} = 2k_x d \tag{4.1-30}$$

where n is the index of refraction of the cavity medium, k_x is the x component of the wave-vector, d is the cavity spacing, and r_1, r_1', r_2, r_2', t_1, t_1', t_2 and t_2' are reflection and transmission coefficients of the mirrors. The prime indicates that the mirror reflection and transmission coefficients are defined for the case when the incident beam is launched from inside the etalon. The four coefficients associated with each mirror are, in general, complex numbers. They are related by the fundamental principles of physics, including reciprocity and time-reversal symmetry (see Problem 4.16). Physically, δ is the round-trip propagation phase shift inside the etalon.

By summing up the infinite geometric series, we obtain the etalon reflection coefficient and transmission coefficient, respectively:

$$r \equiv \frac{A_r}{A_i} = r_1 + \frac{t_1 t_1' r_2' e^{-i\delta}}{1 - r_1' r_2' e^{-i\delta}} = \frac{r_1 + (t_1 t_1' - r_1 r_1') r_2' e^{-i\delta}}{1 - r_1' r_2' e^{-i\delta}} \tag{4.1-31}$$

$$t \equiv \frac{A_t}{A_i} = \frac{t_1 t_2' e^{-i\delta/2}}{1 - r_1' r_2' e^{-i\delta}} \tag{4.1-32}$$

For dielectric mirrors made of lossless materials, the reflection coefficients and transmission coefficients are related by the Stokes relationships [6] (see Problems 4.16 and 4.18). Using the Stokes relationships and Equations (4.1-31) and (4.1-32), the intensity transmission and reflection coefficient of the etalon can be written

$$T_{\text{etalon}} = \frac{|A_t|^2}{|A_i|^2} = \frac{(1-R_1)(1-R_2)}{(1-\sqrt{R_1}\sqrt{R_2})^2 + 4\sqrt{R_1}\sqrt{R_2}\sin^2(\phi)} \tag{4.1-33}$$

$$R_{\text{etalon}} = \frac{|A_r|^2}{|A_i|^2} = \frac{(\sqrt{R_1}-\sqrt{R_2})^2 + 4\sqrt{R_1}\sqrt{R_2}\sin^2(\phi)}{(1-\sqrt{R_1}\sqrt{R_2})^2 + 4\sqrt{R_1}\sqrt{R_2}\sin^2(\phi)} \tag{4.1-34}$$

where R_1 and R_2 are the mirror reflectivities, and the phase ϕ is given by

$$2\phi = \delta + \rho_1' + \rho_2' \tag{4.1-35}$$

where ρ_1' and ρ_2' are the phase shifts of reflections from mirrors 1 and 2, respectively (i.e., $r_1' = |r_1'|\exp(-i\rho_1')$ and $r_2' = |r_2'|\exp(-i\rho_2')$). We note that both the etalon reflectivity and transmittance are periodic functions of ϕ, which is a linear function of the frequency ω. For lossless mirrors with $R_1 + T_1 = 1$ and $R_2 + T_2 = 1$, it can easily be shown that

$$T_{\text{etalon}} + R_{\text{etalon}} = 1 \tag{4.1-36}$$

which is in agreement with the conservation of energy. Note that zero etalon reflection is possible only when $R_1 = R_2$, according to Equation (4.1-34). In other words, zero reflection (or unity maximum transmission) occurs only in symmetric Fabry–Perot etalons at resonance.

Absorption in Fabry–Perot Etalons

Unity maximum transmission is possible only when the mirrors as well as the cavity medium are lossless. In practice, the transmission of light through an etalon can be reduced due to material absorption in the cavity, scattering loss in the cavity or in the mirror, and material absorption in the mirror. Let α be the intensity attenuation coefficient (or gain if $\alpha < 0$) in the cavity medium. The etalon transmission and reflection coefficient become

$$T_{\text{etalon}} = \frac{|A_t|^2}{|A_i|^2} = \frac{(1-R_1)(1-R_2)e^{-\alpha L}}{(1-\sqrt{R_1}\sqrt{R_2}e^{-\alpha L})^2 + 4\sqrt{R_1}\sqrt{R_2}e^{-\alpha L}\sin^2(\phi)} \tag{4.1-37}$$

$$R_{\text{etalon}} = \frac{|A_r|^2}{|A_i|^2} = \frac{(\sqrt{R_1}-\sqrt{R_2}e^{-\alpha L})^2 + 4\sqrt{R_1}\sqrt{R_2}e^{-\alpha L}\sin^2(\phi)}{(1-\sqrt{R_1}\sqrt{R_2}e^{-\alpha L})^2 + 4\sqrt{R_1}\sqrt{R_2}e^{-\alpha L}\sin^2(\phi)} \tag{4.1-38}$$

where L is the length of the medium in which the intensity attenuation is present. Equations (4.1-37) and (4.1-38) are general forms of intensity transmission and reflection of an asymmetric Fabry–Perot etalon. They can be used to describe some special cases of interest.

If we define $1 - A$ as the one-way passage loss in the cavity, then $A = \exp(-\alpha L)$. The transmission at resonance is given by, according to Equation (4.1-37),

$$(T_{\text{etalon}})_{\text{resonance}} = \frac{|A_t|^2}{|A_i|^2} = \frac{(1-R_1)(1-R_2)A}{(1-\sqrt{R_1}\sqrt{R_2}A)^2} \tag{4.1-39}$$

which is in agreement with Equation (4.1-13).

Gires–Tournois Etalons

A very special class of asymmetric Fabry–Perot etalons is known as the Gires–Tournois etalons (GTEs). In these etalons, the rear mirror has a mirror reflectivity of 100% ($R_2 = 1$).

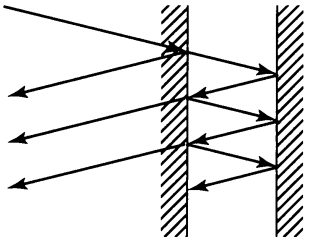


Figure 4.8 Schematic drawing of a Gires–Tournois etalon (GTE).

In this case, all wavelengths of light are reflected, as nothing can transmit through the rear mirror, provided the structure is lossless ($\alpha = 0$). Figure 4.8 shows a schematic drawing of a Gires–Tournois etalon.

In the case of lossless structures, a Gires–Tournois etalon is an all-pass reflector, as nothing can be transmitted through the rear mirror. Without loss of generality, we let $r_1' = -r_1 = \sqrt{R}$, $r_2' = 1$, and $t_1 t_1' - r_1 r_1' = 1$ in Equation (4.1-31) and obtain

$$r_{\text{etalon}} = \exp(-i\Phi) = \frac{-\sqrt{R} + e^{-2i\phi}}{1 - \sqrt{R}e^{-2i\phi}} \tag{4.1-40}$$

where the phase shift ϕ is given by Equation (4.1-35). In cases where the mirror phase shifts can be ignored, the phase can be written (for the case of normal incidence)

$$\phi = \frac{2\pi}{\lambda} nd = \frac{\omega}{c} nd \tag{4.1-41}$$

where d is the space between the mirrors and n is the index of refraction of the medium. It can be shown that the phase shift of a Gires–Tournois etalon can be written, according to Equation (4.1-40),

$$\Phi = 2 \tan^{-1}\left(\frac{1 + \sqrt{R}}{1 - \sqrt{R}} \tan \phi\right) \equiv 2 \tan^{-1}(\sigma \tan \phi) \tag{4.1-42}$$

where

$$\sigma = \frac{1 + \sqrt{R}}{1 - \sqrt{R}} \tag{4.1-43}$$

Figure 4.9 shows the etalon phase shift as a function of ϕ, which is proportional to the frequency. We note that strong phase dispersion occurs at resonance when $\phi = \pi, 2\pi, 3\pi, \ldots$.

By taking the derivative with respect to ω, we obtain

$$\frac{d\Phi}{d\omega} = \frac{d\Phi}{d\phi}\frac{d\phi}{d\omega} = \frac{2\sigma}{1 + (\sigma^2 - 1)\sin^2\phi}\frac{d\phi}{d\omega} = \frac{\sigma}{1 + (\sigma^2 - 1)\sin^2\phi}\tau_0 \tag{4.1-44}$$

where $\tau_0 = 2\, d\phi/d\omega$ is the round-trip group delay of the etalon when the front mirror reflectivity is zero. If we ignore the phase shift dispersion of the rear reflector, then $\tau_0 = 2d/v_g = 2n_g d/c$, where v_g is the group velocity of light in the medium between the mirrors, and n_g is the corresponding group index. Equation (4.1-44) is an expression of the time delay (group delay) due to reflection from the etalon when the front mirror reflectivity is finite. The dimensionless factor $\sigma/[1 + (\sigma^2 - 1)\sin^2\phi]$ in Equation (4.1-44) is exactly one-half of the slope of the etalon phase shift as shown in Figure 4.9. We note that the slope achieves a maximum of 2σ at resonance when $\phi = \pi, 2\pi, 3\pi, \ldots$ and reaches a minimum of $2/\sigma$ when $\phi = m\pi + \pi/2$, m = integer. According to Equation (4.1-43), σ is an increasing function of the front mirror

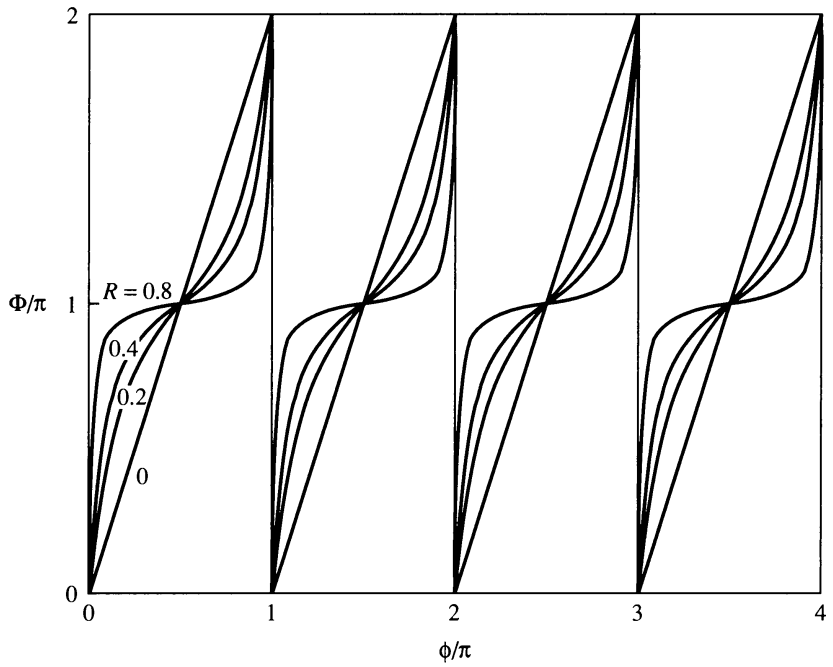


Figure 4.9 Etalon phase shift as a function of ϕ, which is proportional to the frequency. The front mirror reflectivities are $R = 0, 0.2, 0.4, 0.8$.

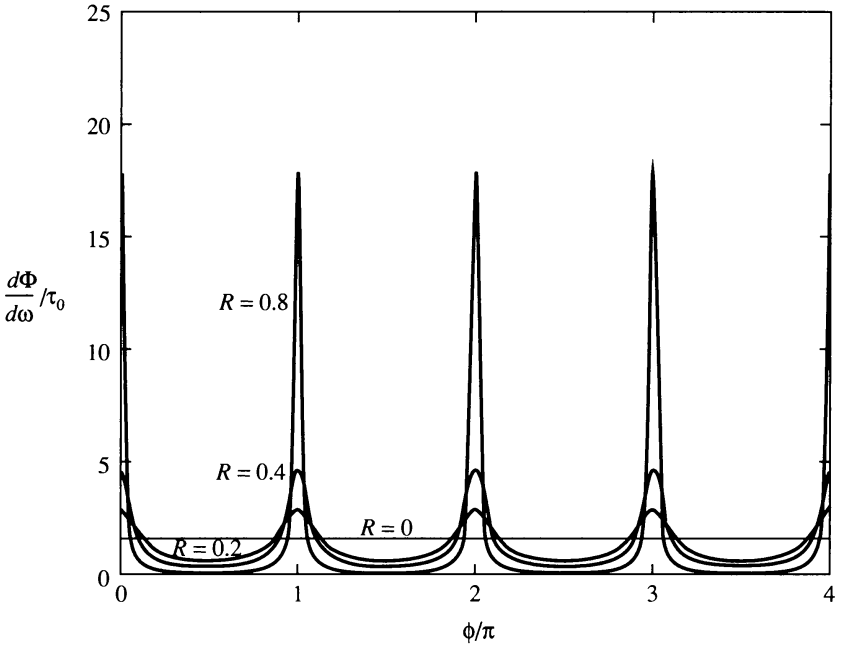


Figure 4.10 Etalon group delay in units of τ_0 as a function of ϕ, which is proportional to the frequency. The front mirror reflectivities are $R = 0, 0.2, 0.4, 0.8$.

reflectivity R. At resonance, the group delay of the etalon becomes $\sigma\tau_0$ and the field energy builds up inside the cavity. Figure 4.10 shows the group delay of the etalon as a function of ϕ, which is proportional to the frequency. For a front mirror reflectivity of $R = 0.8$, the group delay becomes about $18\tau_0$.

When $\phi = m\pi + \pi/2$, m = integer, the etalon group delay becomes τ_0/σ. This is the situation of antiresonance when destructive interference occurs in the cavity.

4.2 FABRY–PEROT ETALONS AS OPTICAL SPECTRUM ANALYZERS

According to Equation (4.1-8), the maximum transmission of a Fabry–Perot etalon occurs when

$$\frac{2nl\cos\theta}{\lambda} = m \tag{4.2-1}$$

Taking, for simplicity, the case of normal incidence ($\theta = 0°$), we obtain the following expression for the change $d\nu$ in the resonance frequency of a given transmission peak due to a length variation dl:

$$\frac{d\nu}{\Delta\nu} = -\frac{dl}{(\lambda/2n)} \tag{4.2-2}$$

where $\Delta\nu$ is the intermode frequency separation as given by Equation (4.1-10). According to Equation (4.2-2), we can tune the peak transmission frequency of the etalon by $\Delta\nu$ by changing its length by half a wavelength. This property is utilized in operating the etalon as a scanning interferometer. The optical signal to be analyzed passes through the etalon as its length is being swept. If the width of the transmission peaks is small compared to that of the spectral detail in the incident optical beam signal, the output of the etalon will constitute a replica of the spectral profile of the signal. In this application it is important that the spectral width of the signal beam be smaller than the intermode spacing of the etalon ($c/2nl$) so that the ambiguity due to simultaneous transmission through more than one transmission peak is avoided. For the same reason, the total length scan is limited to $dl < \lambda/2n$. Figure 4.11 demonstrates the operation of a scanning Fabry–Perot etalon; Figure 4.12 shows intensity versus frequency data obtained by analyzing the output of a multimode He–Ne laser oscillating near 6328 Å. The peaks shown correspond to longitudinal laser modes, which will be discussed in Section 4.5.

It is clear from the foregoing that when operating as a spectrum analyzer the etalon resolution—that is, its ability to distinguish details in the spectrum—is limited by the finite width of its transmission peaks. If we take, somewhat arbitrarily,[1] the limiting resolution of

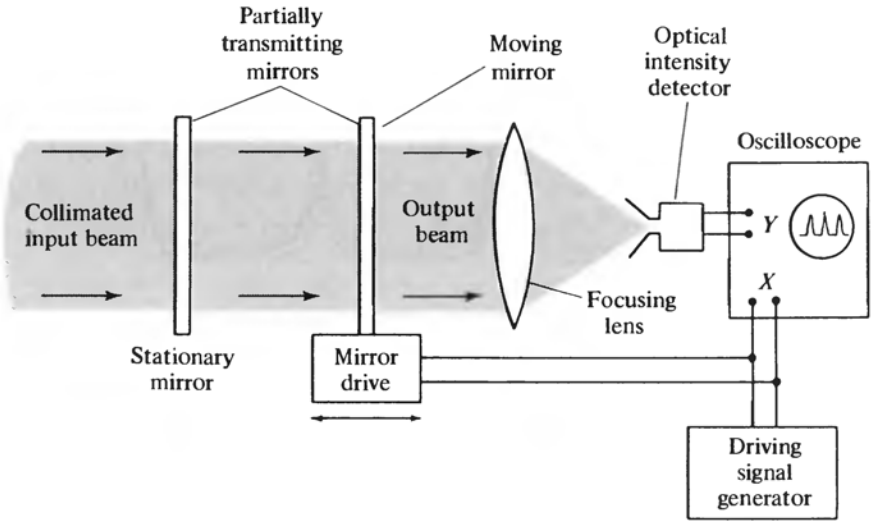


Figure 4.11 Typical scanning Fabry–Perot interferometer experimental arrangement.

[1] For a more complete discussion concerning the definition of resolution, see Reference [4].

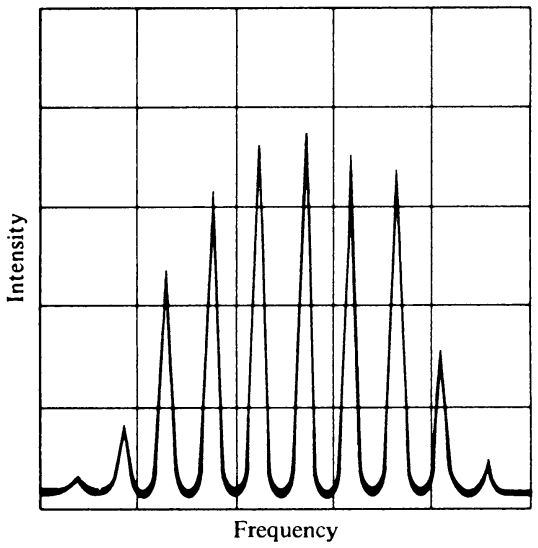


Figure 4.12 Intensity versus frequency analysis of the output of a He–Ne 6328 Å laser obtained with a scanning Fabry–Perot etalon. The horizontal scale is 250 MHz per division.

the etalon as the separation $\Delta\nu_{1/2}$ between the two frequencies at which the transmission is down to half of its peak value, then from Equation (4.1-7) we obtain

$$\sin^2\left(\frac{\delta_{1/2} - 2m\pi}{2}\right) = \frac{(1-R)^2}{4R} \tag{4.2-3}$$

where $\delta_{1/2}$ is the value of δ corresponding to the two half-power points—that is, the value of δ at which the denominator of Equation (4.1-7) is equal to $2(1-R)^2$. If we assume $(\delta_{1/2} - 2m\pi) \ll \pi$, so that the width of the high-transmission regions in Figure 4.4 is small compared to the separation between the peaks, we obtain

$$\Delta\nu_{1/2} = \frac{c}{2\pi nl\cos\theta}(\delta_{1/2} - 2m\pi) \simeq \frac{c}{2\pi nl\cos\theta}\frac{1-R}{\sqrt{R}} \tag{4.2-4}$$

or using Equation (4.1-10) and defining the etalon *finesse* as

$$F \equiv \frac{\pi\sqrt{R}}{1-R} \tag{4.2-5}$$

we obtain

$$\Delta\nu_{1/2} = \frac{\Delta\nu}{F} = \frac{c}{2nl\cos\theta F} \tag{4.2-6}$$

for the limiting resolution. The finesse F (which is used as a measure of the resolution of a Fabry–Perot etalon) is, according to Equation (4.2-6), the ratio of the separation between peaks to the width of a transmission bandpass. This ratio can be read directly from the transmission characteristics such as those of Figure 4.5, for which we obtain $F \simeq 26$.

EXAMPLE: DESIGN OF A FABRY–PEROT ETALON

Consider the problem of designing a scanning Fabry–Perot etalon to be used in studying the mode structure of a He–Ne laser with the following characteristics: $l_{\text{laser}} = 100$ cm and the region of oscillation = $\Delta\nu_{\text{gain}} \simeq 1.5 \times 10^9$ Hz.

The free spectral range of the etalon (i.e., its intermode spacing) must exceed the spectral region of interest, so from Equation (4.1-10) we obtain

$$\frac{c}{2nl_{\text{etal}}} \geq 1.5 \times 10^9 \text{ Hz} \quad \text{or} \quad 2nl_{\text{etal}} \leq 20 \text{ cm} \tag{4.2-7}$$

The separation between longitudinal modes of the laser oscillation is $c/2nl_{\text{laser}} = 1.5 \times 10^8$ Hz (here we assume $n = 1$). We choose the resolution of the etalon to be a tenth of this value, so spectral details as narrow as 1.5×10^7 Hz can be resolved. According to Equation (4.2-6), this resolution can be achieved if

$$\Delta\nu_{1/2} = \frac{c}{2nl_{\text{etal}}F} \leq 1.5 \times 10^7 \text{ Hz} \quad \text{or} \quad 2nl_{\text{etal}}F \geq 2 \times 10^3 \text{ cm} \tag{4.2-8}$$

To satisfy condition (4.2-7), we choose $2nl_{\text{etal}} = 20$ cm; thus condition (4.2-8) is satisfied when

$$F \geq 100 \tag{4.2-9}$$

A finesse of 100 requires, according to Equation (4.2-5), a mirror reflectivity of approximately 97%.

As a practical note we may add that the finesse, as defined by the first equality in Equation (4.2-6), depends not only on R but also on the mirror flatness and the beam angular spread. These points are taken up in Problems 4.3 and 4.4.

Another important mode of optical spectrum analysis performed with Fabry–Perot etalons involves the fact that a noncollimated monochromatic beam incident on the etalon will emerge simultaneously, according to Equation (4.1-8), along many directions θ,[2] which corresponds to the various orders m. If the output is then focused by a lens, each such direction θ will give rise to a circle in the focal plane on the lens, and, therefore, each frequency component present in the beam leads to a family of circles. This mode of spectrum analysis is especially useful under transient conditions where scanning etalons cannot be employed. Further discussion of this topic is included in Problem 4.6.

4.3 OPTICAL RESONATORS WITH SPHERICAL MIRRORS

In this section we study the properties of optical resonators formed by two opposing spherical mirrors; see References [7] and [8]. We will show that the field solutions inside the resonators are those of the propagating Gaussian beams, which were considered in Chapter 2. Consequently, it is useful to start by reviewing the properties of these beams.

The field distribution corresponding to the (l, m) transverse mode is given, according to Equation (2.8-1), by

$$E_{l,m}(\mathbf{r}) = E_0 \frac{\omega_0}{\omega(z)} H_l\left(\sqrt{2}\frac{x}{\omega(z)}\right) H_m\left(\sqrt{2}\frac{y}{\omega(z)}\right) \times \exp\left(-\frac{x^2+y^2}{\omega^2(z)} - ik\frac{x^2+y^2}{2R(z)} - ikz + i(l+m+1)\eta\right) \tag{4.3-1}$$

where the spot size $\omega(z)$ is

$$\omega(z) = \omega_0\left[1 + \left(\frac{z}{z_0}\right)^2\right]^{1/2}, \quad z_0 = \frac{\pi\omega_0^2 n}{\lambda} \tag{4.3-2}$$

[2] Each direction θ corresponds in three dimensions to the surface of a cone with a half-apex angle θ.

and where ω_0, the minimum spot size, is a parameter characterizing the beam. The radius of curvature of the wavefront is

$$R(z) = z\left[1 + \left(\frac{\pi\omega_0^2 n}{\lambda z}\right)^2\right] = \frac{1}{z}[z^2 + z_0^2] \tag{4.3-3}$$

and the phase factor η is as follows:

$$\eta = \tan^{-1}\left(\frac{\lambda z}{\pi\omega_0^2 n}\right) \tag{4.3-4}$$

The sign of $R(z)$ is taken as positive when the center of curvature is to the left of the wavefront, and vice versa. According to Equations (4.3-1) and (4.3-2), the loci of the points at which the beam intensity (watts per square meter) is a given fraction of its intensity on the axis are the hyperboloids

$$x^2 + y^2 = \text{const.} \times \omega^2(z) \tag{4.3-5}$$

The hyperbolas generated by the intersection of these surfaces with planes that include the z axis are shown in Figure 4.13. These hyperbolas are normal to the phase fronts and thus correspond to the local direction of energy flow. The hyperboloid $x^2 + y^2 = \omega^2(z)$ is, according to Equation (4.3-1), the locus of the points where the exponential factor in the field amplitude is down to e^{-1} from its value on the axis. The quantity $\omega(z)$ is thus defined as the *mode spot size* at the plane z.

Given a beam of the type described by Equation (4.3-1), we can form an optical resonator merely by inserting at points z_1 and z_2 two reflectors with radii of curvature that match those of the propagating beam spherical phase fronts at these points. Since the surfaces are normal to the direction of energy propagation as shown in Figure 4.13, the reflected beam retraces itself; thus, if the phase shift between the mirrors is some multiple of 2π radians, a *self-reproducing stable field* configuration results.

Alternatively, given two mirrors with spherical radii of curvature R_1 and R_2 and some distance of separation l, we can, under certain conditions to be derived later, adjust the position $z = 0$ and the parameter ω_0 so that the mirrors coincide with two spherical wavefronts of the propagating beam defined by the position of the waist ($z = 0$) and ω_0. If, in addition, the mirrors can be made large enough to intercept the majority (99%, say) of the incident beam energy in the fundamental ($l = m = 0$) transverse mode, we may expect this mode to have a larger Q than higher-order transverse modes, which, according to Figure 2.7, have fields extending farther from the axis and consequently lose a larger fraction of their energy by "spilling" over the mirror edges (diffraction losses).

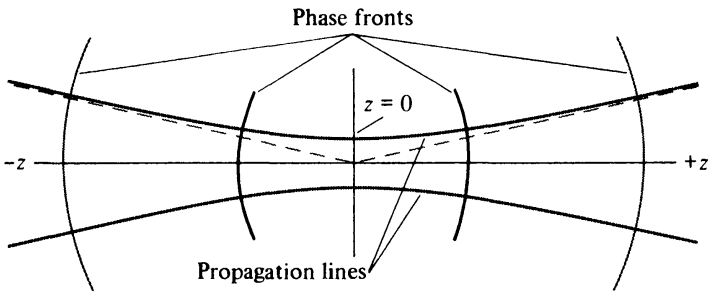


Figure 4.13 Hyperbolic curves corresponding to the local directions of propagation. The nearly spherical phase fronts represent possible positions for reflectors. Any two reflectors form a resonator with a transverse field distribution given by Equation (4.3-1).

Optical Resonator Algebra

As mentioned in the preceding paragraphs, we can form an optical resonator by using two reflectors, one at z_1 and the other at z_2, chosen so that their radii of curvature are the same as those of the beam wavefronts at the two locations. The propagating beam mode (4.3-1) is then reflected back and forth between the reflectors without a change in the transverse profile. The requisite radii of curvature, determined by Equation (4.3-3), are

$$R_1 = z_1 + \frac{z_0^2}{z_1}$$

$$R_2 = z_2 + \frac{z_0^2}{z_2}$$

from which we get

$$z_1 = \frac{R_1}{2} \pm \frac{1}{2}\sqrt{R_1^2 - 4z_0^2}$$

$$z_2 = \frac{R_2}{2} \pm \frac{1}{2}\sqrt{R_2^2 - 4z_0^2} \tag{4.3-6}$$

For a given minimum spot size $\omega_0 = (\lambda z_0/\pi n)^{1/2}$, we can use (4.3-6) to find the positions z_1 and z_2 at which to place mirrors with curvatures R_1 and R_2, respectively.

In practice, we often start with given mirror curvatures R_1 and R_2 and a mirror separation l. The problem is then to find the minimum spot size ω_0, its location with respect to the reflectors, and the mirror spot sizes ω_1 and ω_2. Taking the mirror spacing as $l = z_2 - z_1$, we can solve (4.3-6) for z_0^2, obtaining

$$z_0^2 = \frac{l(-R_1 - l)(R_2 - l)(R_2 - R_1 - l)}{(R_2 - R_1 - 2l)^2} \tag{4.3-7}$$

where z_2 is to the right of z_1 (so that $l = z_2 - z_1 > 0$) and the mirror curvature is taken as positive when the center of curvature is to the left of the mirror.

The minimum spot size $\omega_0 = (\lambda z_0/\pi n)^{1/2}$ and its position is next determined from (4.3-6). The mirror spot sizes are then calculated by the use of Equation (4.3-2).

The Symmetrical Mirror Resonator

The special case of a resonator with symmetrically (about $z = 0$) placed mirrors merits a few comments. The planar phase front at which the minimum spot size occurs is, by symmetry, at $z = 0$. Putting $R_2 = -R_1 = R$ in Equation (4.3-7) gives

$$z_0^2 = \frac{(2R - l)l}{4} \tag{4.3-8}$$

and

$$\omega_0 = \left(\frac{\lambda z_0}{\pi n}\right)^{1/2} = \left(\frac{\lambda}{\pi n}\right)^{1/2}\left(\frac{l}{2}\right)^{1/4}\left(R - \frac{l}{2}\right)^{1/4} \tag{4.3-9}$$

which, when substituted in Equation (4.3-2) with $z = l/2$, yields the following expression for the spot size at the mirrors:

$$\omega_{1,2} = \left(\frac{\lambda l}{2\pi n}\right)^{1/2}\left(\frac{2R^2}{l(R - l/2)}\right)^{1/4} \tag{4.3-10}$$

A comparison with Equation (4.3-9) shows that, for $R \gg l$, $\omega_{1,2} \simeq \omega_0$ and the beam spread inside the resonator is small.

The value of R (for a given l) for which the mirror spot size is a minimum, is readily found from Equation (4.3-10) to be $R = l$. When this condition is fulfilled, we have what is called a symmetrical *confocal resonator*, since the two foci, occurring at a distance of $R/2$ from the mirrors, coincide. From Equation (4.3-9) we obtain

$$(\omega_0)_{\text{conf}} = \left(\frac{\lambda l}{2\pi n}\right)^{1/2} \tag{4.3-11}$$

whereas from Equation (4.3-10) we get

$$(\omega_{1,2})_{\text{conf}} = (\omega_0)_{\text{conf}}\sqrt{2} \tag{4.3-12}$$

so the beam spot size increases by $\sqrt{2}$ between the center and the mirrors.

EXAMPLE: DESIGN OF A SYMMETRICAL RESONATOR

Consider the problem of designing a symmetrical resonator for $\lambda = 10^{-4}$ cm with a mirror separation $l = 2m$. If we were to choose the confocal geometry with $R = l = 2$m, the minimum spot size (at the resonator center) would be, from Equation (4.3-11) and for $n = 1$,

$$(\omega_0)_{\text{conf}} = \left(\frac{\lambda l}{2\pi n}\right)^{1/2} = 0.0564 \text{ cm}$$

whereas, using Equation (4.3-12), the spot size at the mirrors would have the value

$$(\omega_{1,2})_{\text{conf}} = \omega_0\sqrt{2} \simeq 0.0798 \text{ cm}$$

Assume next that a mirror spot size $\omega_{1,2} = 0.3$ cm is desired. Using this value in Equation (4.3-10) and assuming $R \gg l$, we get

$$\frac{\omega_{1,2}}{(\lambda l/2\pi n)^{1/2}} = \frac{0.3}{0.056} = \left(\frac{2R}{l}\right)^{1/4}$$

Thus

$$R \simeq 400l \simeq 800 \text{ m}$$

so that the assumption $R \gg l$ is valid. The minimum beam spot size ω_0 is found, through Equations (4.3-2) and (4.3-8), to be

$$\omega_0 = 0.994\omega_{1/2} \simeq 0.3 \text{ cm}$$

Thus to increase the mirror spot size from its minimum (confocal) value of 0.0798 cm to 0.3 cm, we must use exceedingly plane mirrors ($R = 800$ m). This also shows that even small mirror curvatures (i.e., large R) give rise to "narrow" beams.

The numerical example we have worked out applies equally well to the case in which a plane mirror is placed at $z = 0$. The beam pattern is equal to that existing in the corresponding half of the symmetric resonator in the example, so the spot size on the planar reflector is ω_0.

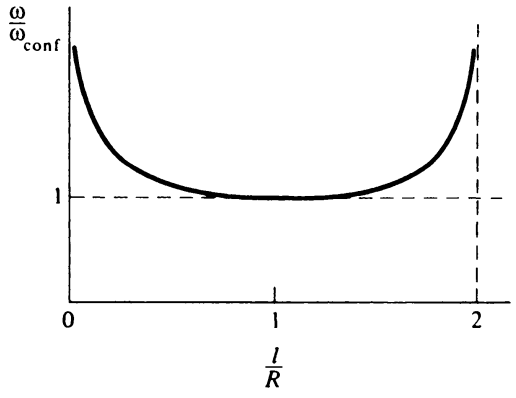


Figure 4.14 Ratio of beam spot size at the mirrors of a symmetrical resonator to its confocal ($l/R = 1$) value.

4.4 MODE STABILITY CRITERIA

The ability of an optical resonator to support low (diffraction) loss[3] modes depends on the mirrors' separation l and their radii of curvature R_1 and R_2. To illustrate this point, consider first the symmetric resonator with $R_2 = -R_1 = R$.

The ratio of the mirror spot size at a given l/R to its minimum confocal ($l/R = 1$) value, given by the ratio of Equation (4.3-10) to (4.3-12), is

$$\frac{\omega_{1,2}}{(\omega_{1,2})_{\text{conf}}} = \left(\frac{1}{(l/R)[2 - (l/R)]}\right)^{1/4} \tag{4.4-1}$$

This ratio is plotted in Figure 4.14. For $l/R = 0$ (plane-parallel mirrors) and for $l/R = 2$ (two concentric mirrors), the spot size becomes infinite. It is clear that the diffraction losses from these cases are very high, since most of the beam energy "spills over" the reflector edges. Since, according to Table 2.1, the reflection of a Gaussian beam from a mirror with a radius of curvature R is formally equivalent to its transmission through a lens with a focal length $f = R/2$, the problem of the existence of stable confined optical modes in a resonator is formally the same as that of the existence of stable solutions for the propagation of a Gaussian beam in a biperiodic lens sequence, as shown in Figure 4.15. This problem was considered in Section 2.1 and led to the stability condition (2.1-16).

If, in Equation (2.1-16), we replace f_1 by $R_1/2$ and f_2 by $R_2/2$, we obtain the stability condition for optical resonators[4]:

$$0 \leq \left(1 - \frac{l}{R_1}\right)\left(1 - \frac{l}{R_2}\right) \leq 1 \tag{4.4-2}$$

A convenient representation of the stability condition (4.4-2) is by means of the diagram [8] shown in Figure 4.16. From this diagram, for example, it can be seen that the symmetric concentric ($R_1 = R_2 = l/2$), the confocal ($R_1 = R_2 = l$), and the plane-parallel ($R_1 = R_2 = \infty$) resonators are all on the verge of instability and thus may become extremely lossy by small deviations of the parameters in the direction of instability.

---

[3] By diffraction loss we refer to the fact that due to the beam spread (see Equation (2.5-18)), a fraction of the Gaussian beam energy 'misses" the mirror and is not reflected and is thus lost.

[4] This causes the sign convention of R_1 and R_2 to be different from that used in preceding sections. The sign of R is the same as that of the focal length of the equivalent lens. This makes R_1 (or R_2) positive when the center of curvature of mirror 1 (or 2) is in the direction of mirror 2 (or 1), and negative otherwise.

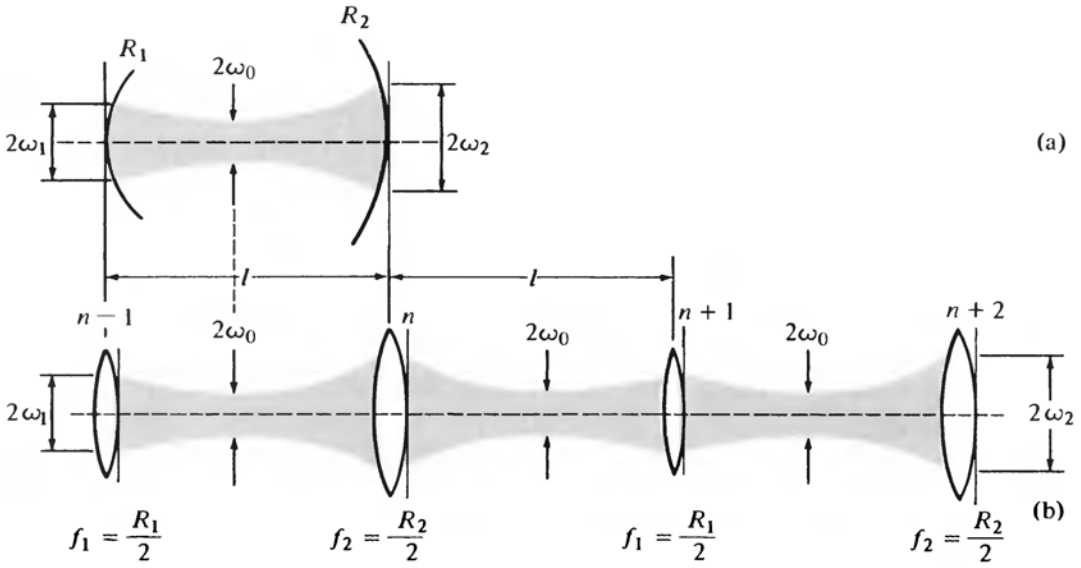


Figure 4.15 (a) Asymmetric resonator ($R_1 \neq R_2$) with mirror curvatures R_1 and R_2. (b) Biperiodic lens system (lens waveguide) equivalent to resonator shown in (a).

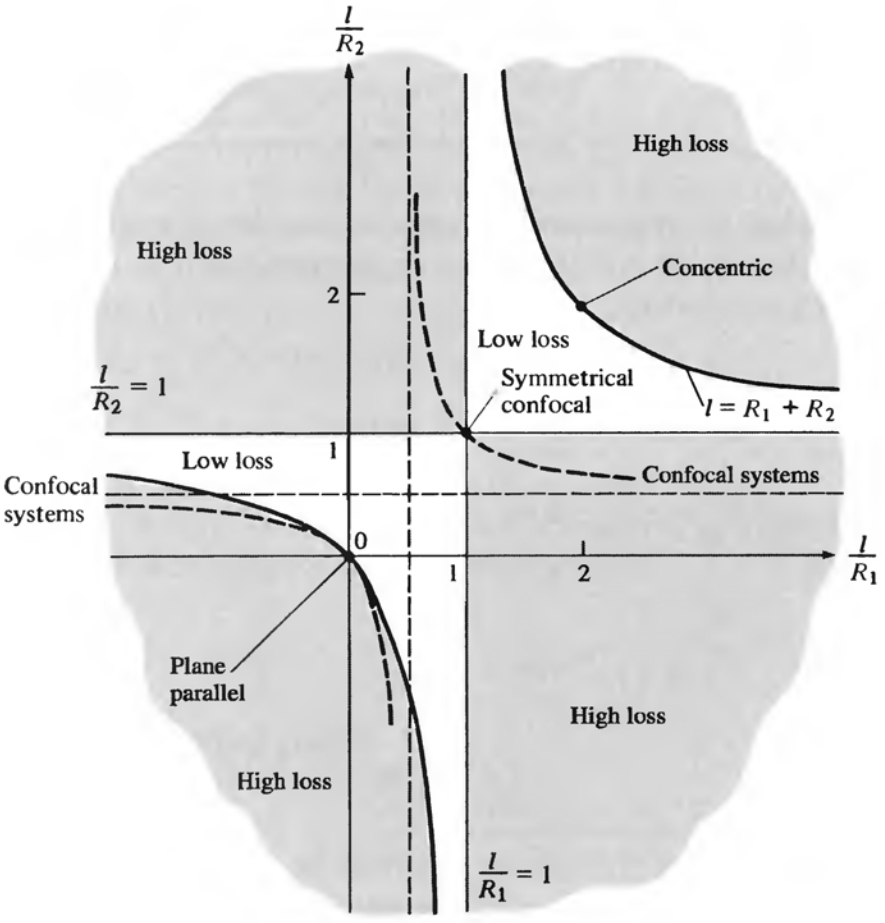


Figure 4.16 Stability diagram of optical resonator. Shaded (high-loss) areas are those in which the stability condition $0 \leq (1 - l/R_1)(1 - l/R_2) \leq 1$ is violated and the clear (low-loss) areas are those in which it is fulfilled. The sign convention for R_1 and R_2 is discussed in footnote 4. (After Reference [8].)

4.5 MODES IN A GENERALIZED RESONATOR—SELF-CONSISTENT METHOD

Up to this point we have treated resonators consisting of two opposing spherical mirrors. We may sometimes wish to consider the properties of more complex resonators made up of an arbitrary number of lenslike elements such as those shown in Table 2.1. A simple case of such a resonator may involve placing a lens between two spherical reflectors or constructing an off-axis three-reflector resonator. Yet another case is that of a traveling wave resonator in which the beam propagates in one sense only.

In either of these cases, we need to find if low-loss (i.e., stable) modes exist in the complex resonator and, if so, to solve for the spot size $\omega(z)$ and the radius of curvature $R(z)$ everywhere.

We apply the self-consistency condition and require that a stable eigenmode of the resonator is one that *reproduces itself after one round trip*. We choose an *arbitrary* reference plane in the resonator, denote the steady-state complex beam parameter at this plane as q_s, and, using the *ABCD* law (2.6-9), require that

$$q_s = \frac{Aq_s + B}{Cq_s + D} \tag{4.5-1}$$

where A, B, C, D are the "ray" matrix elements for one complete round trip—starting and ending at the chosen reference plane.

Solving Equation (4.5-1) for $1/q_s$ gives

$$\frac{1}{q_s} = \frac{(D - A) \pm \sqrt{(D - A)^2 + 4BC}}{2B} \tag{4.5-2}$$

Since the individual elements in the resonator are described by unimodular matrices, that is, $A_iD_i - B_iC_i = 1$ (see Table 2.1), it follows that the matrix A, B, C, D, which is the product of individual matrices, satisfies

$$AD - BC = 1$$

and Equation (4.5-2) can, consequently, be written

$$\frac{1}{q_s} = \frac{D - A}{2B} \pm i\frac{\sqrt{1 - [(D + A)/2]^2}}{B} = \frac{D - A}{2B} + \frac{i \sin \theta}{B} \tag{4.5-3}$$

where

$$\cos \theta = \frac{D + A}{2}$$

$$\theta = \pm\left|\cos^{-1}\left(\frac{D + A}{2}\right)\right| \tag{4.5-4}$$

According to Equation (2.5-14) the condition for a confined Gaussian beam is that the square of the beam spot size ω^2 be a finite positive number. Recalling that q is related to the spot size ω and the radius of curvature R as

$$\frac{1}{q} = \frac{1}{R} - i\frac{\lambda}{\pi\omega^2 n}$$

we find by comparing the last expression to (4.5-3) that the condition for a confined beam is satisfied by choosing θ in Equation (4.5-4), so that $\sin \theta/B < 0$ provided

$$\left|\frac{D+A}{2}\right| < 1 \tag{4.5-5}$$

and the steady-state beam parameter is

$$\frac{1}{q_s} = \frac{D-A}{2B} - i\frac{\sqrt{1-[(D+A)/2]^2}}{|B|} = \frac{D-A}{2B} + \frac{i\sin\theta}{|B|}, \quad \theta < 0 \tag{4.5-6}$$

Equation (4.5-5) can thus be viewed as the generalization of the stability condition (4.4-2) to the case of an arbitrary resonator. When applied to a resonator composed of two spherical reflectors, it reduces to Equation (4.4-2).

The radius of curvature R and the spot size ω at the reference plane are obtained from Equation (4.5-6) by using Equation (2.6-7)

$$\begin{aligned} R &= \frac{2B}{D-A} \\ \omega &= \left(\frac{\lambda}{\pi n}\right)^{1/2} \frac{(|B|)^{1/2}}{[1-[(D+A)/2]^2]^{1/4}} \end{aligned} \tag{4.5-7}$$

The complex beam parameter q, and hence ω and R, at any other plane can be obtained by applying the *ABCD* law (2.6-9) to q_s.

Stability of the Resonator Modes

The treatment just concluded dealt with the existence of steady-state (self-reproducing) resonator modes. Having found that such modes do exist, we need to inquire whether the modes are stable. This can be done by perturbing the steady-state solution $1/q_s$ as given by Equation (4.5-6) and following the evolution of the perturbation with propagation [9].

We start with Equation (2.6-9), which relates the beam parameter q_{out} to the beam parameter q_{in}, after one round trip:

$$q_{\text{out}} = \frac{Aq_{\text{in}} + B}{Cq_{\text{in}} + D}$$

where A, B, C, D are the ray matrix elements for one complete round trip inside the optical resonator. Rewriting the last expression as

$$q_{\text{out}}^{-1} = \frac{C + Dq_{\text{in}}^{-1}}{A + Bq_{\text{in}}^{-1}} \tag{4.5-8}$$

we obtain by differentiation

$$\begin{aligned} \frac{dq_{\text{out}}^{-1}}{dq_{\text{in}}^{-1}} &= \frac{D - [(C + Dq_{\text{in}}^{-1})/(A + Bq_{\text{in}}^{-1})]B}{A + Bq_{\text{in}}^{-1}} \\ &= \frac{D - Bq_{\text{out}}^{-1}}{A + Bq_{\text{in}}^{-1}} \end{aligned} \tag{4.5-9}$$

At steady state, $q_{\text{out}} = q_{\text{in}} \equiv q_s$:

$$\left.\frac{dq_{\text{out}}^{-1}}{dq_{\text{in}}^{-1}}\right|_{q_{\text{in}}=q_s} = \frac{D - Bq_s^{-1}}{A + Bq_s^{-1}} \tag{4.5-10}$$

Using Equation (4.5-6) we obtain

$$D - Bq_s^{-1} = \frac{D + A}{2} - i \sin \theta = e^{-i\theta}$$

$$A + Bq_s^{-1} = \frac{D + A}{2} + i \sin \theta = e^{i\theta}$$

so that

$$\left.\frac{dq_{\text{out}}^{-1}}{dq_{\text{in}}^{-1}}\right|_{q_{\text{in}}=q_s} = e^{-2i\theta} \tag{4.5-11}$$

Because confined modes require, according to Equations (4.5-4) and (4.5-5), that θ be real, it follows from Equation (4.5-11) that a small perturbation $\Delta q_{\text{in}}^{-1}$ of the beam parameter q^{-1} from the steady-state value q_s^{-1} does not decay, since the perturbation after one round trip ($\Delta q_{\text{out}}^{-1}$) satisfies

$$|\Delta q_{\text{out}}^{-1}| = |\Delta q_{\text{in}}^{-1}| \tag{4.5-12}$$

We thus find that the theory predicts that mode perturbations in Gaussian mode resonators do not decay. This does not agree with experience, which shows that the mode characteristics of laser oscillators are highly stable, thus implying a strong perturbational decay, that is, $|\Delta q_{\text{out}}^{-1}| < |\Delta q_{\text{in}}^{-1}|$. The discrepancy is resolved if we include in the analysis leading to Equation (4.5-11) the fact that the resonator mirrors are of finite extent.

4.6 RESONANCE FREQUENCIES OF OPTICAL RESONATORS

Up to this point we have considered only the dependence of the spatial mode characteristics on the resonator mirrors (their radii of curvature and separation). Another important consideration is that of determining the resonance frequency of a given spatial mode.

The frequencies are determined by the condition that the complete round-trip phase delay of a resonant mode be some multiple of 2π. This requirement is equivalent to that in microwave waveguide resonators where the resonator length must be equal to an integral number of half-guide wavelengths [1]. This requirement makes it possible for a stable standing wave pattern to establish itself along the axis with a transverse field distribution equal to that of the propagating mode.

If we consider a spherical mirror resonator with mirrors at z_2 and z_1, the resonance condition for the l, m mode can be written[5]

$$\theta_{l,m}(z_2) - \theta_{l,m}(z_1) = q\pi \tag{4.6-1}$$

where q is some integer and $\theta_{l,m}(z)$, the phase shift, is given according to Equation (2.8-2) by

$$\theta_{l,m}(z) = kz - (l + m + 1) \tan^{-1} \frac{z}{z_0} \quad (z_0 = \pi\omega_0^2 n/\lambda) \tag{4.6-2}$$

[5] In obtaining Equation (4.6-1) we did not allow for the phase shift upon reflection. This correction does not affect any of the results of this section, since these shifts cancel out in the subtraction of Equation (4.6-3).

The resonance condition (4.6-1) is thus

$$k_q d - (l + m + 1)\left(\tan^{-1}\frac{z_2}{z_0} - \tan^{-1}\frac{z_1}{z_0}\right) = q\pi$$

where $d = z_2 - z_1$ is the resonator length. It follows that

$$k_{q+1} - k_q = \frac{\pi}{d}$$

or, using $k = 2\pi\nu n/c$,

$$\nu_{q+1} - \nu_q = \frac{c}{2nd} \tag{4.6-3}$$

for the intermode frequency spacing.

Let us consider, next, the effect of varying the transverse mode indices l and m in a mode with a fixed q. We notice from Equation (4.6-3) that the resonant frequencies depend on the sum $l + m$ and not on l and m separately, so for a given q all the modes with the same value of $l + m$ are degenerate (i.e., they have the same resonance frequencies). Considering Equation (4.6-3) at two different values of $l + m$ gives

$$\begin{aligned} k_1 d - (l + m + 1)_1\left(\tan^{-1}\frac{z_2}{z_0} - \tan^{-1}\frac{z_1}{z_0}\right) &= q\pi \\ k_2 d - (l + m + 1)_2\left(\tan^{-1}\frac{z_2}{z_0} - \tan^{-1}\frac{z_1}{z_0}\right) &= q\pi \end{aligned} \tag{4.6-4}$$

and, by subtraction.

$$(k_1 - k_2)d = [(l + m + 1)_1 - (l + m + 1)_2]\left(\tan^{-1}\frac{z_2}{z_0} - \tan^{-1}\frac{z_1}{z_0}\right) \tag{4.6-5}$$

and

$$\Delta\nu = \frac{c}{2\pi nd}\Delta(l + m)\left(\tan^{-1}\frac{z_2}{z_0} - \tan^{-1}\frac{z_1}{z_0}\right) \tag{4.6-6}$$

for the change $\Delta\nu$ in the resonance frequency caused by a change $\Delta(l + m)$ in the sum $l + m$. As an example, in the case of a Symmetrical confocal resonator ($R = d$) we have, according to Equation (4.3-6), $z_2 = -z_1 = z_0$; therefore $\tan^{-1}(z_2/z_0) = -\tan^{-1}(z_1/z_0) = \pi/4$, and Equation (4.6-6) becomes

$$\Delta\nu_{\text{conf}} = \frac{1}{2}[\Delta(l + m)]\frac{c}{2nd} \tag{4.6-7}$$

Comparing Equation (4.6-7) to (4.6-4), we find that in the confocal resonator the resonance frequencies of the transverse modes, resulting from changing l and m, either coincide or fall halfway between those resulting from a change of the longitudinal mode index q. This situation is depicted in Figure 4.17.

To see what happens to the transverse resonance frequencies (i.e., those due to a variation of l and m) in a non-confocal resonator, we may consider the nearly planar resonator in which $|z_1|$ and z_2 are small compared to z_0 (i.e., $d \ll |R_1|$ and R_2). In this case, Equation (4.6-6) becomes

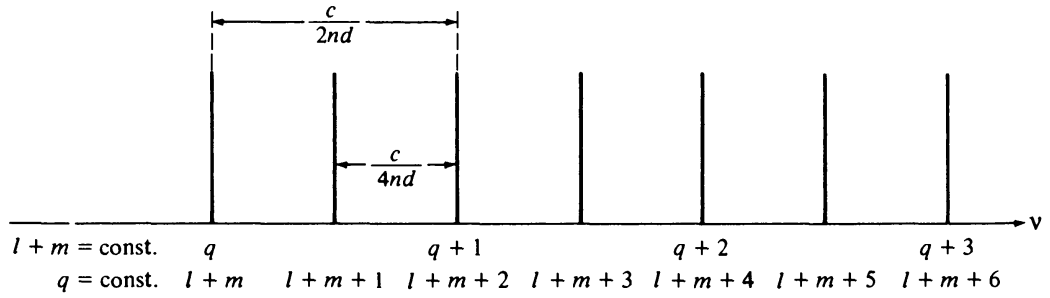


Figure 4.17 Position of resonance frequencies of a confocal ($d = R$) optical resonator as a function of the mode indices l, m, and q.

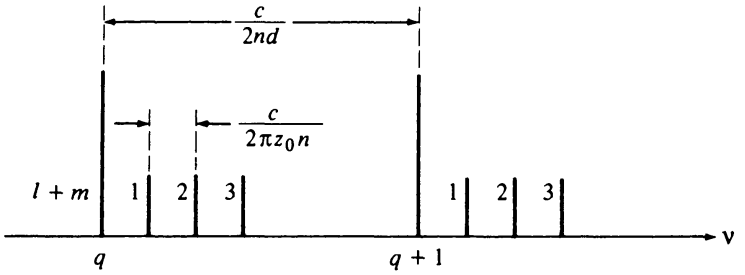


Figure 4.18 Resonant frequencies of a near-planar ($R \gg d$) optical resonator as a function of the mode indices l, m, and q.

$$\Delta\nu \simeq \frac{c}{2\pi n z_0}\Delta(l+m) \tag{4.6-8}$$

The mode grouping for this case is illustrated in Figure 4.18.

In the general case where $|z_1|$ and z_2 are comparable to z_0, the approximation used to derive Equation (4.6-8) does not hold. In this case, it is possible to show using a lengthy, but straightforward, algebra that

$$\tan^{-1}\frac{z_2}{z_0} - \tan^{-1}\frac{z_1}{z_0} = \cos^{-1}\left[\pm\sqrt{\left(1-\frac{d}{R_1}\right)\left(1-\frac{d}{R_2}\right)}\right] \tag{4.6-9}$$

The plus (+) sign is used when both $(1 - d/R)$ factors are positive while the minus (−) sign applies when both are negative. (The other options correspond to unstable resonators.) We can then solve either of Equations (4.6-4) for

$$\nu_{q,l,m} = \frac{c}{2nd}\left[q + (l+m+1)\frac{\cos^{-1}\left[\pm\sqrt{\left(1-\frac{d}{R_2}\right)\left(1-\frac{d}{R_1}\right)}\right]}{\pi}\right] \tag{4.6-10}$$

the resonant frequency of mode q, l, m.

The situation depicted in Figure 4.18 is highly objectionable if the resonator is to be used as a scanning interferometer. The reason is that in reconstructing the spectral profile of the unknown signal, an ambiguity is caused by the simultaneous transmission of more than one frequency. This ambiguity is resolved by using a confocal etalon whose mode spacing is as shown in Figure 4.17 and by choosing d to be small enough that the intermode spacing $c/4nd$ exceeds the width of the spectral region that is scanned.

4.7 LOSSES IN OPTICAL RESONATORS

An understanding of the mechanisms by which electromagnetic energy is dissipated in optical resonators and the ability to control them are of major importance in understanding and operating a variety of optical devices. For historical reasons as well as for reasons of convenience, these losses are often characterized by a number of different parameters. This book uses, in different places, the concepts of loss per pass, photon lifetime, and quality factor Q to describe losses in resonators. Let us see how these quantities are related to each other.

The decay lifetime (photon lifetime) t_c of a cavity mode is defined by means of the equation

$$\frac{d\mathscr{E}}{dt} = -\frac{\mathscr{E}}{t_c} \tag{4.7-1}$$

where $\mathscr{E}$ is the energy stored in the mode so that in a passive resonator $\mathscr{E}(t) = \mathscr{E}(0)\exp(-t/t_c) = \mathscr{E}(0)\exp(-\omega t/Q)$. If the fractional (intensity) loss per pass is L and the length of the resonator is l, then the fractional loss per unit time is cL/nl; therefore

$$\frac{d\mathscr{E}}{dt} = -\frac{cL}{nl}\mathscr{E}$$

and, from Equation (4.7-1),

$$t_c = \frac{nl}{cL} \tag{4.7-2}$$

for the case of a resonator with mirrors' reflectivities R_1 and R_2 and an average distributed loss constant α, the average loss per pass is for small losses $L = \alpha l - \ln\sqrt{R_1R_2}$ so that

$$t_c = \frac{n}{c[\alpha - (1/l)\ln\sqrt{R_1R_2}]} \approx \frac{nl}{c[\alpha l + (1 - \sqrt{R_1R_2})]} \tag{4.7-3}$$

where the approximate equality applies when $R_1R_2 \approx 1$.

The quality factor of the resonator is defined universally as

$$Q = \frac{\omega\mathscr{E}}{P} = -\frac{\omega\mathscr{E}}{d\mathscr{E}/dt} \tag{4.7-4}$$

where $\mathscr{E}$ is the stored energy, ω is the resonant frequency, and $P = -d\mathscr{E}/dt$ is the power dissipated. By comparing Equations (4.7-4) and (4.7-1) we obtain

$$Q = \omega t_c \tag{4.7-5}$$

The Q factor is related to the full width $\Delta\nu_{1/2}$ (at the half-power points) of the resonator's Lorentzian response curve as

$$\Delta\nu_{1/2} = \frac{\nu}{Q} = \frac{1}{2\pi t_c} \tag{4.7-6}$$

([4] and Section 5.5) so that, according to (4.7-3)

$$\Delta\nu_{1/2} = \frac{c[\alpha - (1/l)\ln\sqrt{R_1R_2}]}{2\pi n} \tag{4.7-7}$$

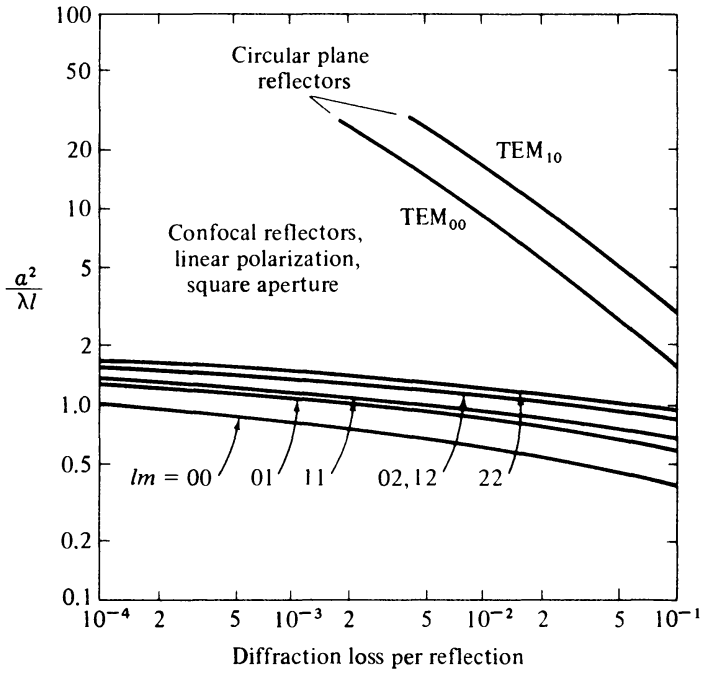


Figure 4.19 Diffraction losses for a plane-parallel resonator and several low-order confocal resonators; a is the mirror radius and l is their spacing. The pairs of numbers under the arrows refer to the transverse-mode indices l, m. (After Reference [7].)

The most common loss mechanisms in optical resonators are the following.

1. *Loss resulting from nonperfect reflection.* Reflection loss is unavoidable, since without some transmission no power output is possible. In addition, no mirror is ideal; and even when mirrors are made to yield the highest possible reflectivities, some residual absorption and scattering reduce the reflectivity to somewhat less than 100%.
2. *Absorption and scattering in the laser medium.* Transitions from some of the atomic levels, which are populated in the process of pumping, to higher-lying levels constitute a loss mechanism in optical resonators when they are used as laser oscillators. Scattering from inhomogeneities and imperfections is especially serious in solid-state laser media.
3. *Diffraction losses.* From Equation (4.3-1) or from Figure 2.7, we find that the energy of propagating-beam modes extends to considerable distances from the axis. When a resonator is formed by "trapping" a propagating beam between two reflectors, it is clear that for finite-dimension reflectors some of the beam energy will not be intercepted by the mirrors and will therefore be lost. For a given set of mirrors this loss will be greater, the higher the transverse mode indices l, m, since in this case the energy extends farther. This fact is used to prevent the oscillation of higher-order modes by inserting apertures into the laser resonator whose opening is large enough to allow most of the fundamental (0, 0, q) mode energy through, but small enough to increase substantially the losses of the higher-order modes. Figure 4.19 shows the diffraction losses of a number of low-order confocal resonators. Of special interest is the dramatic decrease of the diffraction losses that results from the use of spherical reflectors instead of the plane-parallel ones.

4.8 RING RESONATORS

Etalons, resonators, or interferometers can also be made with a ring configuration. These resonators (or interferometers) are employed in ring laser gyros, microresonators, and microdisk

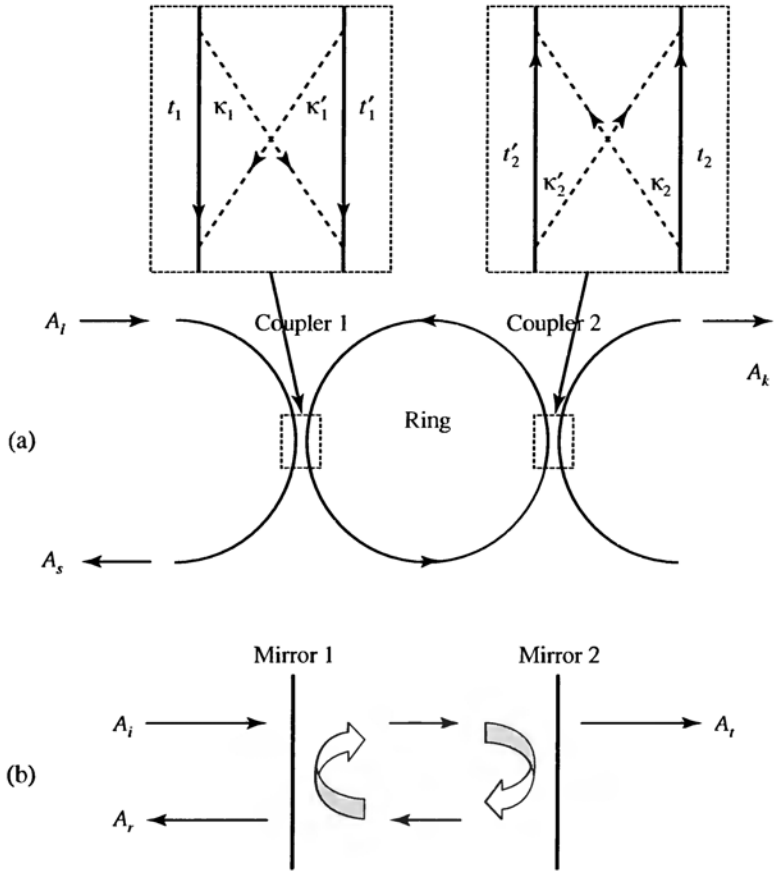


Figure 4.20 Schematic drawings of (a) a ring resonator consisting of a ring of fiber with two couplers and (b) a Fabry–Perot etalon made of two partial reflecting mirrors.

lasers [10–12]. Under appropriate conditions, optical electronic devices can be designed to provide the possibility of switching or modulations. In free space, a ring resonator can be formed by using three mirrors oriented in such a way that a beam of light can circulate along the triangle defined by the three mirrors. In guided waves, a ring resonator can easily be obtained by bending a fiber into a ring, or fabricating a channel waveguide with a circular path. For the guided wave versions, couplers are employed to couple the ring with external waveguides. Referring to Figure 4.20a, we consider a schematic drawing of a ring resonator that is made of a ring of fiber with two couplers. The couplers allow the possibility of energy exchange between the two waveguides (or fibers) with the ring resonator at the location of the couplers. We consider an input beam of amplitude A_i. At the first coupler, a fraction of the input beam will be propagating straight through, while a fraction of the input beam will be coupled into the ring. As the beam circulates inside the ring, multiple couplings occur at the couplers. The situation is similar to the case of multiple reflections in a Fabry–Perot etalon, as shown in Figure 4.20b.

We define the cross-coupling coefficients for the couplers as κ_1, κ_1', κ_2, and κ_2' and the straight-through coupling coefficients as t_1, t_1', t_2, and t_2' (see Figure 4.20a). The multiple circulations inside the ring resonator are physically identical to the multiple reflections inside a Fabry–Perot etalon. Thus we may view Figure 4.20b as an equivalent optical circuit of the ring resonator. Using a similar analysis leading to Equations (4.1-2) and (4.1-3), we obtain the straight-through mode amplitude A_s and the crossover mode amplitude A_k:

$$A_s = \{t_1 + \kappa_1\kappa_1' t_2' e^{-i\delta}[1 + t_1' t_2' e^{-i\delta} + (t_1' t_2' e^{-i\delta})^2 + \cdots]\}A_i \tag{4.8-1}$$

$$A_k = \kappa_1\kappa_2' e^{-i\delta/2}[1 + t_1' t_2' e^{-i\delta} + (t_1' t_2' e^{-i\delta})^2 + \cdots]A_i \tag{4.8-2}$$

with

$$\delta = 2\pi nd/\lambda \tag{4.8-3}$$

where d is the circumference of the ring and n is the effective index of the mode of propagation in the fiber. We note that these two expressions are identical to those of the Fabry–Perot etalon, Equations (4.1-28) and (4.1-29), provided we associate the straight-through coupling coefficients (t_1, t_1', t_2, t_2') with the amplitude reflection coefficients of the mirrors, and the cross-coupling coefficients (κ_1, κ_1', κ_2, κ_2') with the amplitude transmission coefficients of the mirrors. Here, without loss of generality, we assume that the couplers are situated symmetrically in the ring. Physically, δ is the round-trip propagation phase shift inside the resonator.

By summing up the infinite geometric series, we obtain the straight-through (transmission through the fiber) coefficient and crossover (coupled through the ring and finally to the other fiber) coefficient, respectively:

$$\sigma \equiv \frac{A_s}{A_i} = t_1 + \frac{\kappa_1\kappa_1' t_2' e^{-i\delta}}{1 - t_1' t_2' e^{-i\delta}} = \frac{t_1 + (\kappa_1\kappa_1' - t_1 t_1')t_2' e^{-i\delta}}{1 - t_1' t_2' e^{-i\delta}} \tag{4.8-4}$$

$$\chi \equiv \frac{A_k}{A_i} = \frac{\kappa_1\kappa_2' e^{-i\delta/2}}{1 - t_1' t_2' e^{-i\delta}} \tag{4.8-5}$$

Again, these two expressions are formally identical to those of the Fabry–Perot etalons, Equations (4.1-31) and (4.1-32), provided we associate the straight-through coupling coefficients (t_1, t_1', t_2, t_2') with the amplitude reflection coefficients of the mirrors, and the cross-coupling coefficients (κ_1, κ_1', κ_2, κ_2') with the amplitude transmission coefficients of the mirrors. These eight coupling coefficients for the couplers are not independent. They are related by the fundamental principles of reciprocity, conservation of energy, and the time-reversal symmetry. For simplicity, we assume that the couplers are lossless, and the coupling is limited to waves traveling in one sense; that is, no reflection takes place inside the waveguides (or fibers). In this case, these coupling coefficients reduce to two independent constants [13]. The coupling can conveniently be described in terms of a unitary coupling matrix. Before discussing the property of the ring resonators, we must have a clear understanding of the couplers.

Referring to Figure 4.21, we consider the coupling between two single-mode waveguides (or fibers). Let A_1 and A_2 be the mode amplitudes of the input, and B_1 and B_2 be the mode

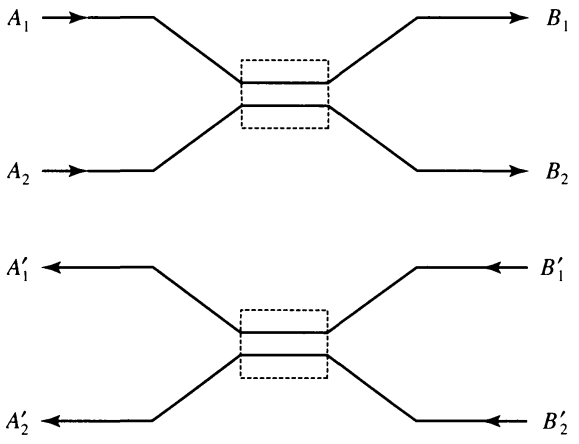


Figure 4.21 Schematic drawings of the couplers.

amplitudes of the output. The modes of propagation in the waveguides are normalized so that the absolute squares of the mode amplitudes represent the modal power flow. The linear input–output relationship of the coupler can be written

$$\begin{aligned} B_1 &= A_1\kappa_{11} + A_2\kappa_{21} \\ B_2 &= A_1\kappa_{12} + A_2\kappa_{22} \end{aligned} \tag{4.8-6}$$

or in matrix form

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} \kappa_{11} & \kappa_{21} \\ \kappa_{12} & \kappa_{22} \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \equiv X \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \tag{4.8-7}$$

where κ_{12} and κ_{21} are the cross-coupling coefficients, and κ_{11} and κ_{22} are the straight-through coupling coefficients. If we reverse the direction of propagation (see Figure 4.21), we can define another matrix X' such that

$$\begin{bmatrix} A_1' \\ A_2' \end{bmatrix} = \begin{bmatrix} \kappa_{11}' & \kappa_{21}' \\ \kappa_{12}' & \kappa_{22}' \end{bmatrix} \begin{bmatrix} B_1' \\ B_2' \end{bmatrix} \equiv X' \begin{bmatrix} B_1' \\ B_2' \end{bmatrix} \tag{4.8-8}$$

By virtue of time-reversal symmetry, these two matrices must satisfy the following relationship:

$$X'X^* = 1 \tag{4.8-9}$$

provided the coupler is lossless. Furthermore, the principle of reciprocity requires that

$$\kappa_{11}' = \kappa_{11}, \quad \kappa_{12}' = \kappa_{21}, \quad \kappa_{22}' = \kappa_{22}, \quad \kappa_{12}' = \kappa_{21} \tag{4.8-10}$$

or equivalently,

$$X' = \tilde{X} \tag{4.8-11}$$

In other words, X' is the transpose of X. As a result, according to Equations (4.8-9)–(4.8-11), we recognize that the matrix X is unitary; that is,

$$X^\dagger X = 1 \tag{4.8-12}$$

where the dagger † stands for Hermitian conjugate. By virtue of the unitary property, the matrix elements of X satisfy the following relations, according to Equation (4.8-12):

$$\begin{aligned} &\kappa_{22} = |X|\kappa_{11}^*, \qquad \kappa_{21} = -|X|\kappa_{12}^* \\ &|\kappa_{11}|^2 + |\kappa_{12}|^2 = |\kappa_{22}|^2 + |\kappa_{21}|^2 = 1 \\ &\kappa_{11}^*\kappa_{21} + \kappa_{12}^*\kappa_{22} = \kappa_{21}^*\kappa_{11} + \kappa_{22}^*\kappa_{12} = 0 \end{aligned} \tag{4.8-13}$$

where $|X|$ is the determinant of the matrix X,

$$|X| = \kappa_{11}\kappa_{22} - \kappa_{12}\kappa_{21} \tag{4.8-14}$$

Equation (4.8-13) is consistent with the conservation of energy.

It can easily be shown that the determinant $|X|$ is, in general, a complex number of a unit modulus (i.e., $|X| = \exp(i\theta)$, with θ being a real number). Without loss of generality, we can take

$$|X| = \kappa_{11}\kappa_{22} - \kappa_{12}\kappa_{21} = -1 \tag{4.8-15}$$

The choice leads to a convenience in the matrix analysis of coupled ring resonators. With this choice of the determinant ($|X| = -1$), the linear input–output relationship of the coupler can be written

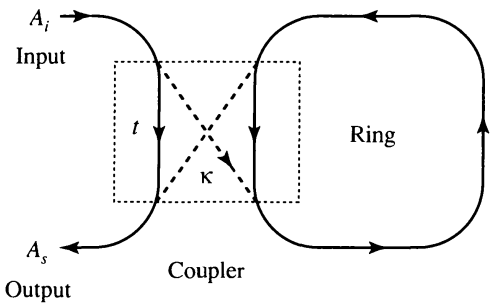


Figure 4.22 Schematic drawing of a ring resonator coupled with a waveguide. (t, κ) are the coupling coefficients of the coupler.

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = X \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} t & \kappa^* \\ \kappa & -t^* \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \tag{4.8-16}$$

where $t = \kappa_{11}$ is the straight-through coupling coefficient and $\kappa = \kappa_{12}$ is the cross-coupling coefficient of the coupler. Note that the coupling matrix X is Hermitian ($X^\dagger = X$), as a result of the choice of $|X| = -1$. These two coupling coefficients satisfy the following relationship, according to Equation (4.8-13):

$$|t|^2 + |\kappa|^2 = 1 \tag{4.8-17}$$

Matrix formulation can conveniently be employed to investigate the optical properties of coupled resonators when several ring resonators are involved. Using the unitary conditions (4.8-13) and (4.8-15), the straight-through transmission coefficient of the ring resonator of Figure 4.20a can be written, according to Equation (4.8-4),

$$\sigma \equiv \frac{A_s}{A_i} = \frac{t_1 + t_2' a e^{-i\delta}}{1 - t_1' t_2' a e^{-i\delta}} = \frac{t_1 + t_2' e^{-i\delta} e^{-\alpha d}}{1 - t_1' t_2' e^{-i\delta} e^{-\alpha d}} \tag{4.8-18}$$

where $a = \exp(-\alpha d)$, with α being an amplitude attenuation coefficient to account for loss due to bending or scattering. We now consider a special case of a ring resonator where $\kappa_2 = 0$ and $t_2 = -1$. This resonator is illustrated in Figure 4.22. For this case, the straight-through transmission coefficient becomes

$$\sigma \equiv \frac{A_s}{A_i} = \frac{t_1 - a e^{-i\delta}}{1 + t_1' a e^{-i\delta}} = \frac{t - a e^{-i\delta}}{1 - t^* a e^{-i\delta}} \tag{4.8-19}$$

where we recall δ is the round-trip phase shift inside the ring resonator, and $t_1' = -t_1^* \equiv -t^*$. Here, we drop the subscript 1, as there is only one coupler. By taking the absolute square of Equation (4.8-19), we obtain the intensity transmission (straight-through)

$$|\sigma|^2 \equiv \left| \frac{A_s}{A_i} \right|^2 = \frac{a^2 + t^2 - 2at \cos\delta}{1 + a^2 t^2 - 2at \cos\delta} \tag{4.8-20}$$

where, without loss of generality, we have taken t as a real number. This is legitimate, as the constant phase of t can always be lumped with the phase δ.

At resonance $\delta = m2\pi$ (m = integer), the straight-through intensity transmission becomes

$$|\sigma|^2 \equiv \left| \frac{A_s}{A_i} \right|^2 = \frac{(a - |t|)^2}{(1 - a|t|)^2} = \frac{(a - t)^2}{(1 - at)^2} \tag{4.8-21}$$

This simple relation, plotted in Figure 4.23 as a function of the cross-coupling efficiency $|\kappa|^2$, has two important features that are the key for some potential applications. (1) The straight-through transmission $|\sigma|^2$ is zero at a value of coupling $a = t = \sqrt{1 - |\kappa|^2}$, known as the

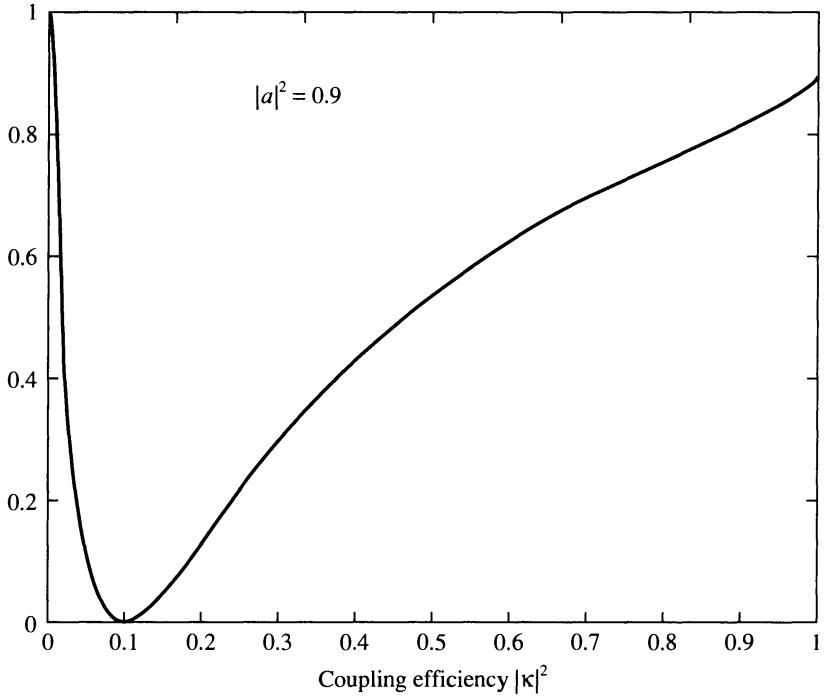


Figure 4.23 Straight-through transmission of a ring resonator is plotted as a function of the coupling efficiency $|\kappa|^2$. Critical coupling occurs when the coupling efficiency equals the ring resonator internal loss $(1 - |a|^2)$.

"critical coupling." (2) For high-finesse ring resonators with a low internal loss (round-trip loss = $1 - |a|^2$, a is near unity), the portion of the curve to the left of the critical coupling point is extremely steep. "Small changes" in a for a given coupling efficiency $|\kappa|^2$, or vice versa, can control the transmission between zero and unity. Such a feature is desirable in the design of switching of modulation devices with a high modulation depth [13]. Electroabsorption or optical amplifiers can be employed to control the loss [14].

Coupled Ring Resonators

We next consider the situation when several ring resonators are coupled serially. The optical property of such coupled ring resonators can conveniently be described by using the unitary matrix relationship, Equation (4.8-15), discussed earlier. Referring to Figure 4.24, we consider serially coupled ring resonators. Fiber 1 is for the input and straight-through output, fibers 2,

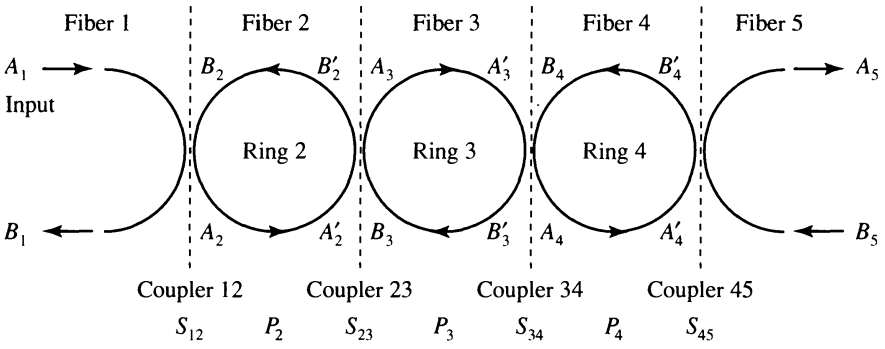


Figure 4.24 Schematic drawing of a system of serially coupled ring resonators, consisting of three ring resonators. These three ring resonators are labeled as ring 2, ring 3, and ring 4.

3, and 4 are ring resonators, and fiber 5 is for the cross output. For the first coupler (12) between fibers 1 and 2 in Figure 4.24, the input–output matrix relationship can be written, according to Equation (4.8-15),

$$\begin{bmatrix} B_1 \\ A_2 \end{bmatrix} = X_{12} \begin{bmatrix} A_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} \kappa_{11} & \kappa_{21} \\ \kappa_{12} & \kappa_{22} \end{bmatrix} \begin{bmatrix} A_1 \\ B_2 \end{bmatrix} \tag{4.8-22}$$

where κ_{11}, κ_{22}, κ_{12}, and κ_{21} are the coupling coefficients of the coupler. These coupling coefficients are related by the unitary conditions, according to Equations (4.8-13) and (4.8-15),

$$\kappa_{22} = -\kappa_{11}^*, \qquad \kappa_{21} = \kappa_{12}^* \tag{4.8-23}$$

Note that for this coupler (12), A_1 and B_2 are the input mode amplitudes, whereas B_1 and A_2 are the output mode amplitudes. For the purpose of analyzing the serially coupled ring resonators, we rewrite Equation (4.8-22) as

$$\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} = \frac{1}{\kappa_{12}^2} \begin{bmatrix} 1 & \kappa_{11}^* \\ \kappa_{11} & 1 \end{bmatrix} \begin{bmatrix} A_2 \\ B_2 \end{bmatrix} \equiv S_{12} \begin{bmatrix} A_2 \\ B_2 \end{bmatrix} \tag{4.8-24}$$

where S_{12} is defined as the scattering matrix of the coupler, A_1 is the incident mode amplitude, B_1 is the straight-through mode amplitude in the same fiber, and A_2 and B_2 are mode amplitudes inside the first ring resonator (see Figure 4.24). It is important to distinguish the difference between matrices X and S. Coupling matrix X is a linear relationship between the two input mode amplitudes and the two output mode amplitudes to a coupler, whereas matrix S is a linear relationship between the two mode amplitudes of one waveguide to the two mode amplitudes of the other waveguides. The determinant of the scattering matrix is given by, according to Equations (4.8-13), (4.8-23), and (4.8-24),

$$|S_{12}| = \frac{1}{\kappa_{12}^2}(1 - |\kappa_{11}|^2) = \frac{\kappa_{21}}{\kappa_{12}} = \frac{\kappa_{12}^*}{\kappa_{12}} \tag{4.8-25}$$

We note that the determinant is, in general, a complex number with a unit modulus. It becomes 1 when the cross-coupling constant κ_{12} is real. In this case, the scattering matrix S is Hermitian and unimodular.

We define A_2' and B_2' as the mode amplitudes in the first ring resonator (fiber 2, or ring 2) at the location of the next coupler (23). Similarly, (A_m, B_m) and (A_m', B_m') are the mode amplitudes in ring resonator m (fiber m). The amplitudes in the resonators are related by the propagation in the ring of fibers. Without loss of generality, we assume that the couplers are situated symmetrically. The relationship can be written

$$\begin{bmatrix} A_m \\ B_m \end{bmatrix} = \begin{bmatrix} e^{i\phi_m/2} & 0 \\ 0 & e^{-i\phi_m/2} \end{bmatrix} \begin{bmatrix} A_m' \\ B_m' \end{bmatrix} \equiv P_m \begin{bmatrix} A_m' \\ B_m' \end{bmatrix} \tag{4.8-26}$$

where ϕ_m is the round-trip phase shift in ring resonator m, and P_m is defined as the propagation matrix. A complex phase shift ϕ_m should be employed to account for propagation or bending loss. At the coupler between the mth fiber and the $(m + 1)$st fiber, the mode amplitudes are related by, according to Equation (4.8-24),

$$\begin{bmatrix} A_m' \\ B_m' \end{bmatrix} = \frac{1}{\kappa_{m,m+1}} \begin{bmatrix} 1 & \kappa_{mm}^* \\ \kappa_{mm} & 1 \end{bmatrix} \begin{bmatrix} A_{m+1} \\ B_{m+1} \end{bmatrix} \equiv S_{m,m+1} \begin{bmatrix} A_{m+1} \\ B_{m+1} \end{bmatrix}, \quad m \geq 1 \tag{4.8-27}$$

where $S_{m,m+1}$ is the scattering matrix for the coupler $(m, m + 1)$, for $m > 0$.

Using Equations (4.8-24), (4.8-26), and (4.8-27), the mode amplitudes of fiber 1 and fiber 5 can be written

$$\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} = S_{12}\ P_2\ S_{23}\ P_3\ S_{34}\ P_4\ S_{45} \begin{bmatrix} A_5 \\ B_5 \end{bmatrix} \tag{4.8-28}$$

where S_{12}, S_{23}, S_{34}, and S_{45} are the scattering matrices of the couplers, and P_2, P_3, and P_4 are the propagation matrices of the rings. The same equation can easily be extended for the case of N serially coupled ring resonators (with $N > 3$),

$$\begin{bmatrix} A_0 \\ B_0 \end{bmatrix} = S_{01}\ P_1\ S_{12}\ P_2\ S_{23} \cdots P_N\ S_{N,N+1} \begin{bmatrix} A_{N+1} \\ B_{N+1} \end{bmatrix} \equiv M \begin{bmatrix} A_{N+1} \\ B_{N+1} \end{bmatrix} \equiv \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} A_{N+1} \\ B_{N+1} \end{bmatrix} \tag{4.8-29}$$

where A_0 and B_0 are the mode amplitudes in fiber 0, and A_{N+1} and B_{N+1} are the mode amplitudes in fiber $N + 1$; fibers 1 through N are ring resonators. The matrix M is the product of all the matrices in sequence, similar to Equation (4.8-28). The straight-through transmission coefficient in fiber 0 is thus given by, according to Equation (4.8-29),

$$\sigma \equiv \left(\frac{B_0}{A_0} \right)_{B_{N+1}=0} = \frac{M_{21}}{M_{11}} \tag{4.8-30}$$

The crossover transmission coefficient from fiber 0 to fiber $N + 1$ is given by

$$\chi \equiv \left(\frac{A_{N+1}}{A_0} \right)_{B_{N+1}=0} = \frac{1}{M_{11}} \tag{4.8-31}$$

The matrix elements M_{11}, M_{12}, M_{21}, and M_{22} satisfy some relationship similar to that of S_{12} (see Equation (4.8-24)). Based on the conservation of power flow in a lossless system, the mode amplitudes must satisfy the following condition:

$$|A_0|^2 - |B_0|^2 = |A_{N+1}|^2 - |B_{N+1}|^2 \tag{4.8-32}$$

Using Equations (4.8-29) and (4.8-32), we obtain

$$\begin{aligned} |M_{11}|^2 - |M_{21}|^2 = |M_{22}|^2 - |M_{12}|^2 = 1 \\ M_{11}M_{12}^* - M_{21}M_{22}^* = 0 \end{aligned} \tag{4.8-33}$$

These equations are consistent with time-reversal symmetry and the principle of reciprocity. According to Equation (4.8-33), we have

$$|\sigma|^2 + |\chi|^2 = 1 \tag{4.8-34}$$

which is exactly conservation of energy. It follows directly from Equation (4.8-33) that the matrix elements satisfy the following relationships:

$$|M_{22}| = |M_{11}|, \quad |M_{21}| = |M_{12}| \tag{4.8-35}$$

For the case when all cross-coupling coefficients are real, the relationships become

$$M_{22} = M_{11}^*, \quad M_{21} = M_{12}^* \tag{4.8-36}$$

In the event when all couplers are identical, and all rings are identical, the matrix equation (4.8-29) can be written

$$\begin{bmatrix} A_0 \\ B_0 \end{bmatrix} = [SP]^N\ S \begin{bmatrix} A_{N+1} \\ B_{N+1} \end{bmatrix} \tag{4.8-37}$$

where S is the scattering matrix of the couplers, and P is the propagation matrix for the rings. Identical microrings can be fabricated using lithography replication techniques [15, 16]. Equations (4.8-31) and (4.8-32) can be employed to calculate the straight-through or cross-over transmission coefficients. Both intensity as well as the phase shift can then be obtained. The frequency derivative of the phase shift will yield the group delay. For the case of N identical ring resonators, the straight-through transmission efficiency can be written (see Problem 4.19)

$$|\sigma|^2 \equiv \left|\frac{B_0}{A_0}\right|^2_{B_{N+1}=0} = \left|\frac{M_{21}}{M_{11}}\right|^2 = \frac{(1-|\kappa|^2)}{(1-|\kappa|^2)+|\kappa|^2\dfrac{\sin^2\theta}{\sin^2(N+1)\theta}} \tag{4.8-38}$$

where θ is given by

$$\cos\theta = \frac{1}{|\kappa|}\cos(\phi/2) \tag{4.8-39}$$

where $\phi = 2\pi n_{\text{eff}} d/\lambda$ is the round-trip phase shift due to propagation in the ring resonators. In the spectral regimes where $|\cos\theta|$ is greater than unity, θ is a complex number. The second term in the denominator of Equation (4.8-38) decays exponentially as a function of N. These regimes are the so-called photonic bandgaps. The straight-through transmission (or "reflection" from the rings) approaches unity. Figure 4.25 is a plot of the straight-through transmission $|\sigma|^2$ of a series of six rings ($N = 6$) as a function of the phase shift ϕ. We notice that $|\sigma|^2$ is a periodic function of ϕ. The spectral response consists of a periodic array of stopbands, separated by $(N - 1)$ sidelobes. Further discussion of photonic bandgaps will be presented later in this book. Figure 4.26 is a photograph of a portion of a linear array of identical ring resonators [15, 16].

For an infinite chain of identical ring resonators with identical coupling constants, the fields are periodic at twice the lattice constant Λ. Here we define the lattice constant as a

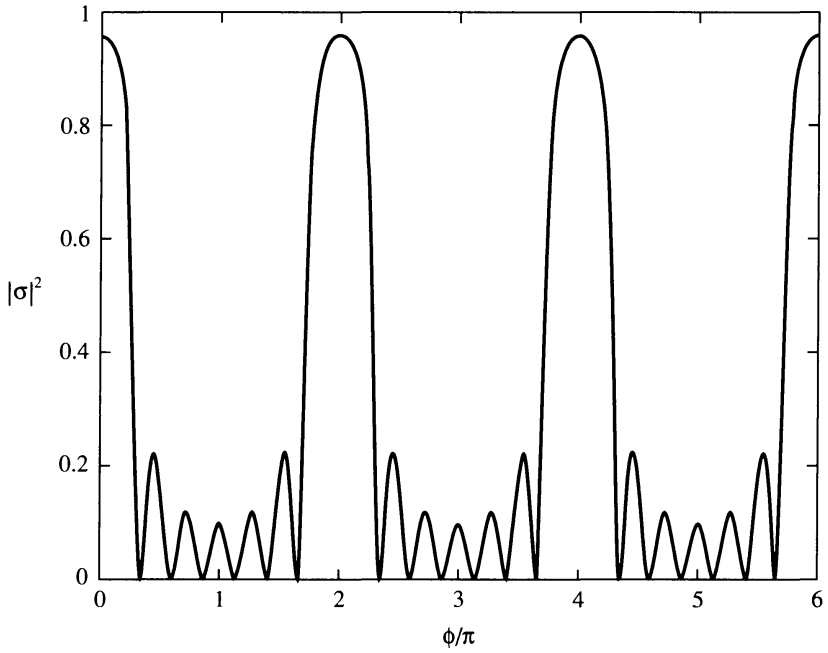


Figure 4.25 Straight-through transmission efficiency $|\sigma|^2$ as a function of phase shift ϕ. The parameters are $N = 6$ and $|\kappa| = 0.95$.

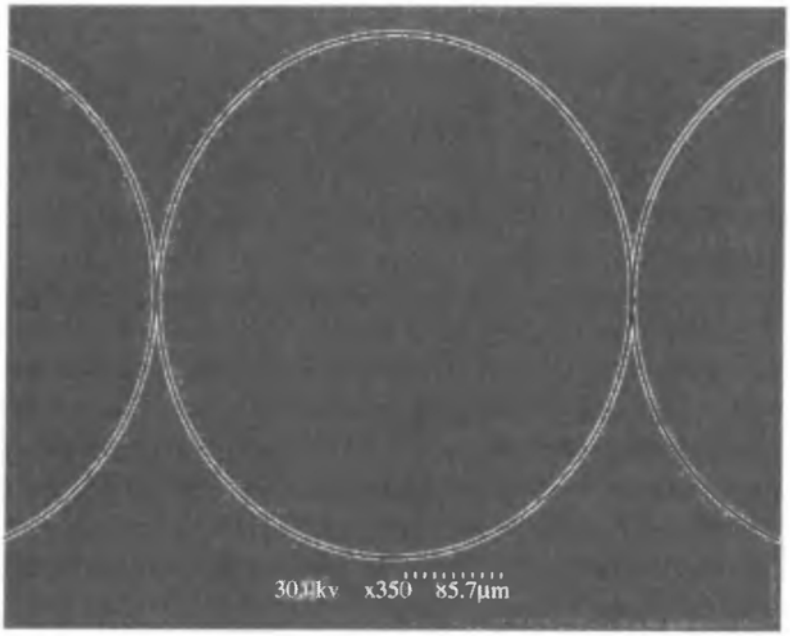


Figure 4.26 Photograph of a portion of a linear array of coupled ring resonators. The ring resonators have a radius of 50 μm, with a circumference of about 314 μm. The waveguides are made of a polymer (SU-8) with a refractive index of 1.57, on top of a silica substrate with a refractive index of 1.44 at $\lambda =$ 1550 nm. The waveguide dimensions are 1.6 μm thick and 2 μm wide. At the coupling region, the rings are separated by approximately 0.5 μm. (Courtesy of George T. Paloczi. References [15] and [16].)

characteristic length associated with a single ring. Using Bloch's theorem, the field amplitudes at the n-th ring are related to those of the (n + 2)-th ring by the following relationship,

$$\begin{pmatrix} A_n \\ B_n \end{pmatrix} = P_n\ S_{n,n+1}\ P_{n+1}\ S_{n+1,n+2} \begin{pmatrix} A_{n+2} \\ B_{n+2} \end{pmatrix} = (PS)^2 \begin{pmatrix} A_{n+2} \\ B_{n+2} \end{pmatrix} = e^{i2K\Lambda} \begin{pmatrix} A_{n+2} \\ B_{n+2} \end{pmatrix} \tag{4.8-40}$$

where P is the propagation matrix of a single ring ($P_n = P$ for all n), S is the scattering matrix of a single coupler ($S_{n,n+1} = S$ for all n), K is the Bloch wavenumber. The amplitudes of the normal modes of propagation thus satisfy the following eigenvalue equation

$$(PS)^2 \begin{pmatrix} A_0 \\ B_0 \end{pmatrix} = e^{i2K\Lambda} \begin{pmatrix} A_0 \\ B_0 \end{pmatrix} \tag{4.8-41}$$

and

$$\begin{pmatrix} A_{2n} \\ B_{2n} \end{pmatrix} = e^{-i2nK\Lambda} \begin{pmatrix} A_0 \\ B_0 \end{pmatrix} \tag{4.8-42}$$

The amplitudes A_0 and B_0 of the normal modes as well as the Block wavenumber can be obtained by solving Equation (4.8-41). The dispersion relation (K as a function of ω) is thus given by

$$\mathrm{Det}|(PS)^2 - e^{i2K\Lambda} I| = 0 \tag{4.8-43}$$

where I is the unit matrix. Using Equations (4.8-26) and (4.8-27) for P and S, we obtain

$$\cos K\Lambda = \frac{1}{|\kappa_{12}|} \cos \frac{1}{2}\phi \tag{4.8-44}$$

where we recall that ϕ is the roundtrip phase shift in a single ring resonator. For ring resonators made of single mode fibers, this phase shift can be written

$$\phi = \frac{\omega}{c} n_{eff} d = \frac{2\pi}{\lambda} n_{eff} d \tag{4.8-45}$$

where d is the circumference of the ring resonator and n_{eff} is the effective index of the mode of propagation in the fiber.

According to Equation (4.8-13), the magnitude of the coupling constant $|\kappa_{12}|$ is always less than (or equal to) unity ($0 \leq |\kappa_{12}| \leq 1$). As a result, the Bloch wavenumber K can be complex in some spectral regimes according to Equation (4.8-44). A Bloch wave with a complex wavenumber cannot propagate in the periodic medium. According to Equation (4.8-42), the field amplitudes decay exponentially in the spectral regimes where the Bloch wavenumber is complex. These spectral regimes are known as the photonic bandgaps. If we plot the dispersion relationship (ω vs K) using Equation (4.8-44), we find that photonic bandgaps occur in the spectral regime where the roundtrip phase shift ϕ is an integral number of 2π. The centers of the photonic bandgaps occur at exactly $\phi = 2m\pi$ ($m = 1, 2, 3 \ldots$). The Bloch wavenumbers at the centers of the photonic bandgaps are given by

$$K\Lambda = m\pi \pm i \cosh^{-1}\left(\frac{1}{|\kappa_{12}|}\right) \tag{4.8-46}$$

We note that the magnitude of the imaginary part of the Bloch wavenumber depends on the cross-coupling constant of the waveguide coupler. In Chapter 12, we will discuss more periodic structures that are also employed in photonics and optical electronics.

4.9 MULTICAVITY ETALONS

In the previous section, we described the transmission properties of single ring resonators as well as coupled ring resonators. Resonance occurs when the round-trip propagation phase shift inside the resonator is an integral multiple of 2π. For ring resonators made of single-mode fibers or other single-mode waveguides, propagation attenuation due to bending loss can be significant, particularly for ring resonators with circumferences in the range of micrometers. In this section, we discuss coupled etalons made of thin films or multilayer structures. Propagation attenuation is negligible for thin films with thicknesses in the range of micrometers. There are situations in optical electronics when multicavity etalons are needed for spectral filtering or dispersion compensation applications. The transmission and reflection properties of these multicavity etalons can be analyzed by using the results obtained in Sections 4.1 and 4.8. Particularly, the amplitude transmission and reflection coefficients of Equations (4.1-31) and (4.1-32), or equivalently Equations (4.8-4) and (4.8-5) can be very useful. Referring to Figure 4.27, we consider the transmission of light through an etalon consisting of two mirrors separated by a distance d. The transmission and reflection coefficients are written in a more general form as, according to Equations (4.1-31) and (4.1-32),

$$r = r_{12} + \frac{t_{12}t_{21}r_{23}e^{-i2\phi}}{1 - r_{21}r_{23}e^{-i2\phi}} = \frac{r_{12} + (t_{12}t_{21} - r_{12}r_{21})r_{23}e^{-i2\phi}}{1 - r_{21}r_{23}e^{-i2\phi}} \tag{4.9-1}$$

$$t = \frac{t_{12}t_{23}e^{-i\phi}}{1 - r_{21}r_{23}e^{-i2\phi}} \tag{4.9-2}$$

with

$$2\phi = 2k_x d = 2nd\cos\theta(\omega/c) \tag{4.9-3}$$

where d is the thickness of the cavity medium, n is the refractive index of the cavity medium, and k_x is the component of the wavevector along the normal of the mirror surface. The transmission and the reflection coefficients of the mirrors are defined as:

t_{12} = transmission coefficient of mirror 1 with incidence from medium 1
t_{21} = transmission coefficient of mirror 1 with incidence from medium 2
t_{23} = transmission coefficient of mirror 2 with incidence from medium 2
t_{32} = transmission coefficient of mirror 2 with incidence from medium 3
r_{12} = reflection coefficient of mirror 1 with incidence from medium 1
r_{21} = reflection coefficient of mirror 1 with incidence from medium 2
r_{23} = reflection coefficient of mirror 2 with incidence from medium 2
r_{32} = reflection coefficient of mirror 2 with incidence from medium 3

The above definitions are also illustrated in Figure 4.27.

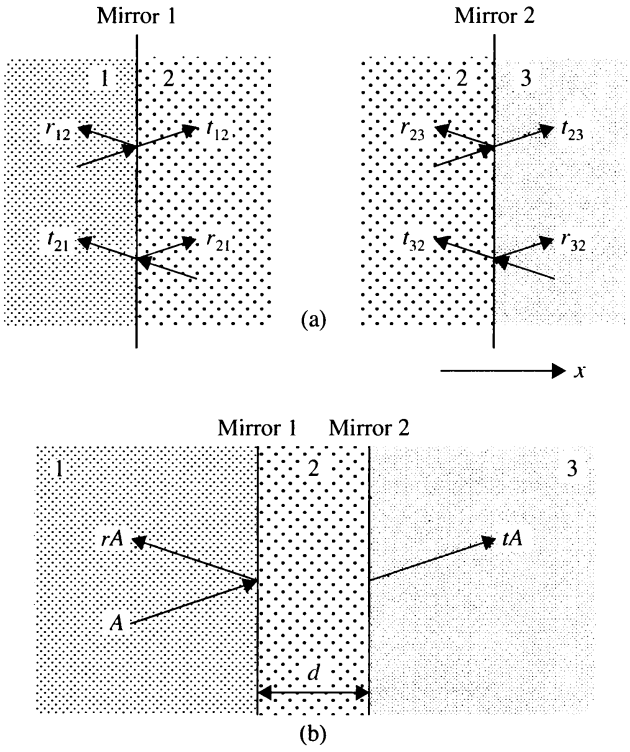


Figure 4.27 Definition of transmission and reflection coefficients. (a) Definition of the mirror transmission and reflection coefficients. (b) Transmission and reflection coefficients of the etalon formed by the two mirrors.

The four coefficients associated with each mirror are in general complex numbers. As discussed earlier, they are related by fundamental principles of physics, including reciprocity and time-reversal symmetry. Specifically, we have (see Problem 4.16)

$$\begin{aligned} t_{12}t_{21}^* + r_{21}r_{21}^* &= 1, \quad t_{23}t_{32}^* + r_{32}r_{32}^* = 1 \\ t_{12}r_{12}^* + r_{21}t_{12}^* &= 0, \quad t_{23}r_{23}^* + r_{32}t_{23}^* = 0 \end{aligned} \tag{4.9-4}$$

These equations are known as the Stokes relationships [6].

It is important to note that t_{12} and t_{21} are identical only when the refractive indices of media 1 and 2 are the same. For lossless media with different refractive indices, these two sets of transmission coefficients are related by

$$\frac{t_{12}}{k_{1x}} = \frac{t_{21}}{k_{2x}}, \quad \frac{t_{23}}{k_{2x}} = \frac{t_{32}}{k_{3x}} \tag{4.9-5}$$

where k_{1x}, k_{2x}, and k_{3x} are x components of the wavevectors in media 1, 2, and 3, respectively,

$$\begin{aligned} k_{1x} &= n_1 \cos\theta_1(\omega/c) \\ k_{2x} &= n_2 \cos\theta_2(\omega/c) \\ k_{3x} &= n_3 \cos\theta_3(\omega/c) \end{aligned} \tag{4.9-6}$$

where ω is the frequency of the incident beam of light, n_1, n_2, and n_3 are the refractive indices, and θ_1, θ_2, and θ_3 are the ray angles in the media, respectively. Equation (4.9-5) is consistent with the reciprocity of transmission of energy. The factors k_{1x}, k_{2x}, and k_{3x} are essential to account for the different speeds of light in the media.

To illustrate the use of the transmission and reflection formulas, we consider the following example.

EXAMPLE: THIN FILM SANDWICHED BETWEEN TWO DIFFERENT MEDIA

Referring to Figure 4.28, we consider a thin film of thickness d_2 and refractive index n_2. The film is sandwiched between two media with refractive indices n_1 and n_3.

The interface reflection and transmission coefficients are simply the Fresnel reflection and transmission coefficients. They are given by [6]

$$r_{12} = \begin{cases} \dfrac{k_{1x} - k_{2x}}{k_{1x} + k_{2x}} \\ \dfrac{n_1^2 k_{2x} - n_2^2 k_{1x}}{n_1^2 k_{2x} + n_2^2 k_{1x}} \end{cases} = \begin{cases} \dfrac{n_1 \cos\theta_1 - n_2 \cos\theta_2}{n_1 \cos\theta_1 + n_2 \cos\theta_2} & (s\text{-wave}) \\ \dfrac{n_1 \cos\theta_2 - n_2 \cos\theta_1}{n_1 \cos\theta_2 + n_2 \cos\theta_1} & (p\text{-wave}) \end{cases} \tag{4.9-7}$$

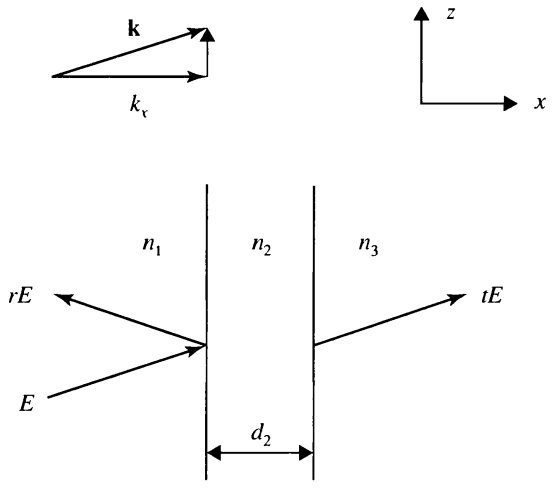


Figure 4.28 The transmission and reflection of light through and from a thin film. The x axis is perpendicular to the film surface.

$$t_{12} = \begin{cases} \dfrac{2k_{1x}}{k_{1x} + k_{2x}} \\ \dfrac{2n_1 n_2 k_{1x}}{n_1^2 k_{2x} + n_2^2 k_{1x}} \end{cases} = \begin{cases} \dfrac{2n_1 \cos\theta_1}{n_1 \cos\theta_1 + n_2 \cos\theta_2} & (s\text{-wave}) \\ \dfrac{2n_1 \cos\theta_1}{n_1 \cos\theta_2 + n_2 \cos\theta_1} & (p\text{-wave}) \end{cases} \tag{4.9-8}$$

where k_{1x} and k_{2x} are the x components of the wavevectors in media 1 and 2, respectively. θ_1 and θ_2 are ray angles in media 1 and 2, respectively. The ray angles are related to the x components of the wavevectors by the relationships (4.9-6). For an s-wave, the electric field vector is perpendicular to the plane of incidence (xz plane), whereas for a p-wave, the electric field vector is parallel to the plane of incidence. A similar set of transmission and reflection coefficients for the interface between media 2 and 3 can be obtained from the above by replacing $2 \rightarrow 3$ and $1 \rightarrow 2$. We note that the Fresnel reflection and transmission coefficients satisfy the following relationships:

$$\begin{aligned} t_{12}t_{21} - r_{12}r_{21} &= 1 \\ r_{21} &= -r_{12} \end{aligned} \tag{4.9-9}$$

For the purpose of illustrating the concept, we consider the special case of normal incidence. At normal incidence, $\theta_1 = \theta_2 = 0$, and the s-wave and p-wave are identical. The Fresnel transmission and reflection become

$$\begin{aligned} r_{12} &= \frac{n_1 - n_2}{n_1 + n_2}, \quad r_{21} = -r_{12} \\ r_{23} &= \frac{n_2 - n_3}{n_2 + n_3}, \quad r_{32} = -r_{23} \\ t_{12} &= \frac{2n_1}{n_1 + n_2}, \quad t_{23} = \frac{2n_2}{n_2 + n_3} \\ t_{21} &= \frac{2n_2}{n_1 + n_2}, \quad t_{32} = \frac{2n_3}{n_2 + n_3} \end{aligned} \tag{4.9-10}$$

For lossless media, these coefficients are all real. The reflection and transmission coefficients for the thin film shown in Figure 4.28 can now be written, according to Equations (4.9-1), (4.9-2), and (4.9-10),

$$r = \frac{r_{12} + r_{23}e^{-i2\phi}}{1 - r_{21}r_{23}e^{-i2\phi}} = \frac{\left(\dfrac{n_1 - n_2}{n_1 + n_2}\right) + \left(\dfrac{n_2 - n_3}{n_2 + n_3}\right)e^{-i2\phi}}{1 + \left(\dfrac{n_1 - n_2}{n_1 + n_2}\right)\left(\dfrac{n_2 - n_3}{n_2 + n_3}\right)e^{-i2\phi}} \tag{4.9-11}$$

$$t = \frac{t_{12}t_{23}e^{-i\phi}}{1 - r_{21}r_{23}e^{-i2\phi}} = \frac{\left(\dfrac{2n_1}{n_1 + n_2}\right)\left(\dfrac{2n_2}{n_2 + n_3}\right)e^{-i\phi}}{1 + \left(\dfrac{n_1 - n_2}{n_1 + n_2}\right)\left(\dfrac{n_2 - n_3}{n_2 + n_3}\right)e^{-i2\phi}} \tag{4.9-12}$$

where $\phi = k_2 d_2 = n_2 d_2(\omega/c) = 2\pi n_2 d_2/\lambda$.

EXAMPLE: ANTIREFLECTION (AR) COATING

We consider an example of a thin film coating to eliminate the Fresnel reflection at $\lambda = 1.5$ μm. Let medium 1 be air with a refractive index of $n_1 = 1$, and medium 3 be a semiconductor with a refractive index of $n_3 = 3.5$. Equation (4.9-11) can be employed to investigate the reflection as a function of the film thickness and its index of refraction. According to Equation (4.9-11), the reflection coefficient vanishes when

$$r_{12} + r_{23}e^{-i2\phi} = \left(\frac{n_1 - n_2}{n_1 + n_2}\right) + \left(\frac{n_2 - n_3}{n_2 + n_3}\right)e^{-i2\phi} = 0 \tag{4.9-13}$$

The above equation can easily be satisfied by choosing

$$n_2 = \sqrt{n_1 n_3} \quad \text{and} \quad \phi = \pi/2, 3\pi/2, 5\pi/2, 7\pi/2, \ldots \tag{4.9-14}$$

In other words, a thin film with a quarter-wave thickness and a refractive index equal to the geometric average of the bounding media can lead to zero reflectance. Since $\phi = k_2 d_2 = n_2 d_2(\omega/c) = 2\pi n_2 d_2/\lambda$, the quarter thickness is

$$d_2 = \frac{\lambda}{4n_2} = \frac{\lambda}{4\sqrt{n_1 n_3}} \tag{4.9-15}$$

Using $n_1 = 1$ and $n_3 = 3.5$, the refractive index of the thin film must be

$$n_2 = \sqrt{n_1 n_3} = 1.87$$

and the thickness can be

$$d_2 = \frac{\lambda}{4n_2} = \frac{\lambda}{4\sqrt{n_1 n_3}} = 0.2\ \mu\text{m}$$

or an odd integral multiple of it. It is important to note that the antireflection coating described above is only limited to the wavelength at $\lambda = 1.5$ μm. Equation (4.9-9) can be employed to investigate the reflectivity at wavelength around $\lambda_0 = 1.5$ μm. Figure 4.29 shows the reflectance ($|r|^2$) as a function of the phase ϕ for various indices of refraction of the thin film coating. The reflectance is a periodic function of the phase with minima occurring at $\phi = \pi/2, 3\pi/2, 5\pi/2, 7\pi/2, \ldots$ for $n_2 < n_3$. When the film index is greater than n_3, the reflectance is higher and the minima occur at $\phi = \pi, 2\pi, 3\pi, \ldots$

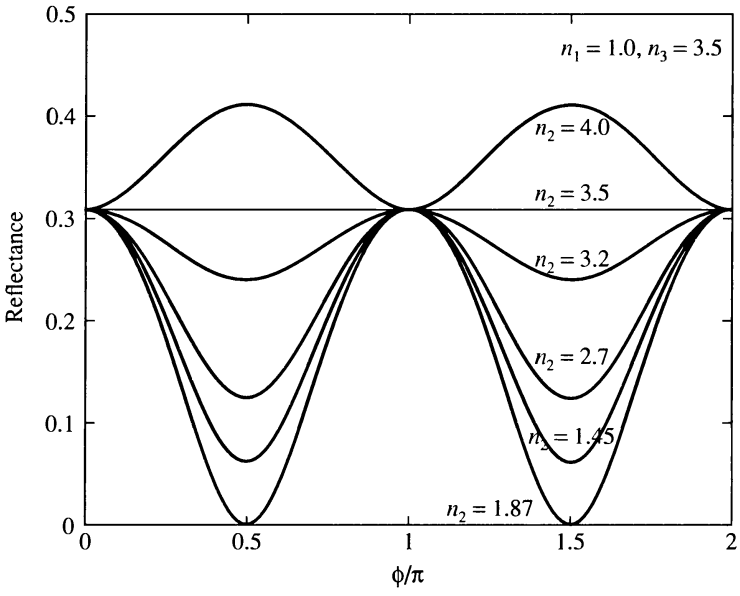


Figure 4.29 Reflectance versus ϕ for $n_1 = 1$, $n_3 = 3.5$, and various values of n_2.

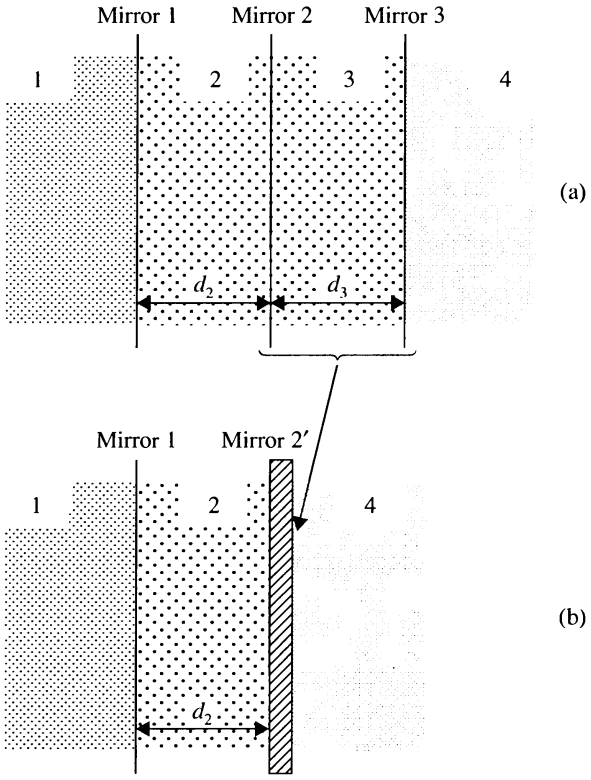


Figure 4.30 Transmission and reflection problem of a two-cavity etalon. The second etalon consisting of mirrors 2 and 3 separated by a distance d_3 is equivalent to a mirror (mirror 2′). This reduces the two-cavity etalon problem to a one-cavity etalon problem.

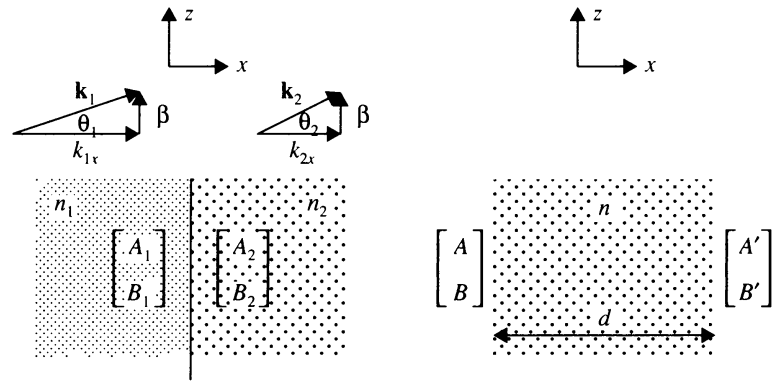
The formulas for transmission and reflection coefficients of the etalon, Equations (4.9-1) and (4.9-2), can be generalized to treat multicavity etalons. To illustrate this, we consider a two-cavity etalon as shown in Figure 4.30. To find the transmission and reflection coefficients, we first find the transmission and reflection coefficients of the second etalon formed by medium 3 sandwiched between mirrors 2 and 3. This is done by removing mirror 1 and medium 1. The formula in Equations (4.9-1) and (4.9-2) can be employed for this purpose. We then consider a one-cavity problem formed by mirror 1 and an equivalent mirror (mirror 2′). The equivalent mirror (mirror 2′) consists of the second etalon formed by mirrors 2 and 3, sandwiched between media 2 and 4. The equivalent one-cavity etalon is illustrated in Figure 4.30b. The transmission and reflection coefficients can then be obtained by employing again the general formulas of Equations (4.9-1) and (4.9-2). The method described here can be generalized to treat the transmission and reflection problem of multicavity etalons.

2 × 2 Matrix Method

In addition to the use of an equivalent mirror, the problem of transmission and reflection properties of multicavity etalons can be treated by using a 2 × 2 matrix method. The matrix method is so general that it can be applied to treat the transmission and reflection properties of a general multilayer structure. To introduce the matrix method, we consider a general case of plane wave propagation on both sides of a dielectric interface (see Figure 4.31). We can write the electric field as

$$E = \begin{cases} (A_1 e^{-ik_{1x}x} + B_1 e^{+ik_{1x}x}) e^{-i\beta z} & \text{medium 1} \\ (A_2 e^{-ik_{2x}x} + B_2 e^{+ik_{2x}x}) e^{-i\beta z} & \text{medium 2} \end{cases} \tag{4.9-16}$$

where A_1 and A_2 are the amplitudes of the right-traveling wave (along $+x$ direction), B_1 and B_2 are the amplitudes of the left-traveling wave (along $-x$ direction), β is the z component of



$$\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} = \frac{1}{t_{12}} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \begin{bmatrix} A_2 \\ B_2 \end{bmatrix} \equiv T_{12} \begin{bmatrix} A_2 \\ B_2 \end{bmatrix} \qquad \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} e^{i\Phi} & 0 \\ 0 & e^{-i\Phi} \end{bmatrix} \begin{bmatrix} A' \\ B' \end{bmatrix} \equiv P \begin{bmatrix} A' \\ B' \end{bmatrix}$$

(a) (b)

Figure 4.31 (a) Transition matrix for a dielectric interface between two media. (b) Propagation matrix for a homogeneous medium.

the wavevector, and k_{1x} and k_{2x} are x components of the wavevector. Note that β is the same in both media as the structure is homogeneous in the z direction. These components of the wavevectors are related by the following equations:

$$\begin{aligned} k_{1x}^2 + \beta^2 &= (n_1\omega/c)^2 \\ k_{2x}^2 + \beta^2 &= (n_2\omega/c)^2 \end{aligned} \tag{4.9-17}$$

Based on the definition of the transmission and reflection coefficients associated with the interface, the amplitudes are related by the following equations:

$$\begin{aligned} B_1 &= A_1 r_{12} + B_2 t_{21} \\ A_2 &= A_1 t_{12} + B_2 r_{21} \end{aligned} \tag{4.9-18}$$

where r_{12}, r_{21}, t_{12}, and t_{21} are the Fresnel reflection and transmission coefficients discussed earlier. Solving for A_1 and B_1 in terms of A_2 and B_2, we obtain

$$\begin{aligned} A_1 &= \frac{1}{t_{12}} A_2 - \frac{r_{21}}{t_{12}} B_2 \\ B_1 &= \frac{r_{12}}{t_{12}} A_2 + \left(t_{21} - \frac{r_{12} r_{21}}{t_{12}} \right) B_2 \end{aligned} \tag{4.9-19}$$

Using Equation (4.10-9), the above can be rewritten

$$\begin{aligned} A_1 &= \frac{1}{t_{12}} A_2 + \frac{r_{12}}{t_{12}} B_2 \\ B_1 &= \frac{r_{12}}{t_{12}} A_2 + \frac{1}{t_{12}} B_2 \end{aligned} \tag{4.9-20}$$

or, equivalently, in a matrix form

$$\begin{bmatrix} A_1 \\ B_1 \end{bmatrix} = \frac{1}{t_{12}} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \begin{bmatrix} A_2 \\ B_2 \end{bmatrix} \equiv T_{12} \begin{bmatrix} A_2 \\ B_2 \end{bmatrix} \tag{4.9-21}$$

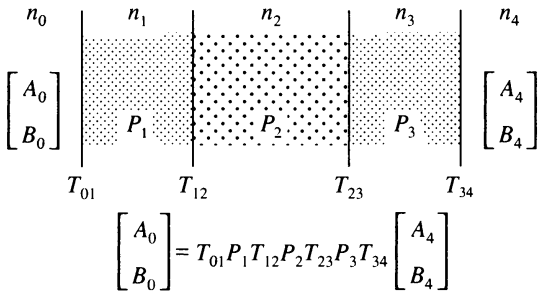


Figure 4.32 Schematic drawing showing the matrix method of treating an example of multilayer structures (with $N = 3$).

where T_{12} is defined as the transition matrix between layers 1 and 2. We notice the strong resemblance between the transition matrix T_{12} and the scattering matrix S_{12} of Equation (4.8-24) for ring resonators. Physically, t_{12} is equivalent to the cross-coupling coefficient κ_{12}, and r_{12} is equivalent to the straight-through coupling coefficient κ_{11}. It is important to note that the reflection and transmission coefficients of the interface, r_{12} and t_{12}, depend on the state of polarization (either s or p). Thus the problem of transmission and reflection is often treated separately for each of the polarization states. We note the transition matrix T_{12} is a symmetric matrix.

The transition matrix relates the field column vector on both sides of a dielectric interface. We now need a matrix for propagation from one end of a homogeneous layer to the other end of the layer. Referring to Figure 4.31b, we consider propagation of plane waves in a homogeneous medium with thickness d and refractive index n. Using the column vector for the field representation, we obtain

$$\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} e^{i\phi} & 0 \\ 0 & e^{-i\phi} \end{bmatrix} \begin{bmatrix} A' \\ B' \end{bmatrix} \equiv P \begin{bmatrix} A' \\ B' \end{bmatrix} \tag{4.9-22}$$

where A and B are the field amplitudes on the left end of the medium, whereas A' and B' are the field amplitudes on the right end of the medium. The phase shift is given by

$$\phi = k_x d = nd \cos \theta(\omega/c) \tag{4.9-23}$$

where k_x is the x component of the wavevector in the medium and θ is the ray angle (angle between the wavevector and the x axis). The diagonal matrix P is defined as the propagation matrix. For lossless media, the propagation matrix is unitary. The transition matrix T and the propagation matrix P can now be employed to treat a general multilayer structure. This is illustrated in Figure 4.32 for the case of a stack of three layers ($N = 3$) sandwiched between semi-infinite media of refractive indices n_0 and n_4. The column vector of the plane wave amplitudes (A_0, B_0) in medium 0 is related to the column vector of the plane wave amplitudes (A_{N+1}, B_{N+1}) by the following relationship:

$$\begin{bmatrix} A_0 \\ B_0 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} A_{N+1} \\ B_{N+1} \end{bmatrix} \equiv M \begin{bmatrix} A_{N+1} \\ B_{N+1} \end{bmatrix} \tag{4.9-24}$$

where M is the 2×2 matrix that can be obtained by multiplying the transition matrices and the propagation in sequence:

$$M = T_{01}P_1T_{12}P_2T_{23}P_3 \ldots P_{N-1}T_{N-1,N}P_NT_{N,N+1} \tag{4.9-25}$$

We note the propagation matrix is independent of the state of polarization, whereas the transition matrix depends on the state of polarization. Using the symmetric property of the transition matrix T_{12} and the unitary property of the propagation matrix, it can easily be shown that

$$M_{21} = M_{12}^*, \qquad M_{22} = M_{11}^* \tag{4.9-26}$$

for lossless multilayer structures. Using Equations (4.9-21) and (4.9-22) as well as the explicit forms of the Fresnel reflection and transmission coefficients (4.9-7) and (4.9-8), it can be shown that the determinant is given by

$$|M| = \frac{k_{N+1x}}{k_{0x}} \tag{4.9-27}$$

where k_{N+1x} and k_{0x} are the x components of the wavevectors in media 0 and $N + 1$, respectively. Equation (4.9-27) is valid even if loss is present in the layered structure. If the two media are identical (i.e., $n_0 = n_{N+1}$), then the matrix M is unimodular.

The transition matrices for both the s-wave and p-wave are given as, according to Equations (4.9-7), (4.9-8), and (4.9-21),

$$T_{12} = \begin{cases} \dfrac{1}{2k_{1x}} \begin{bmatrix} k_{1x} + k_{2x} & k_{1x} - k_{2x} \\ k_{1x} - k_{2x} & k_{1x} + k_{2x} \end{bmatrix} & (s\text{-wave}) \\ \dfrac{1}{2n_1 n_2 k_{1x}} \begin{bmatrix} n_1^2 k_{2x} + n_2^2 k_{1x} & n_1^2 k_{2x} - n_2^2 k_{1x} \\ n_1^2 k_{2x} - n_2^2 k_{1x} & n_1^2 k_{2x} + n_2^2 k_{1x} \end{bmatrix} & (p\text{-wave}) \end{cases} \tag{4.9-28}$$

or equivalently, in terms of ray angles

$$T_{12} = \begin{cases} \dfrac{1}{2n_1 \cos\theta_1} \begin{bmatrix} n_1 \cos\theta_1 + n_2 \cos\theta_2 & n_1 \cos\theta_1 - n_2 \cos\theta_2 \\ n_1 \cos\theta_1 - n_2 \cos\theta_2 & n_1 \cos\theta_1 + n_2 \cos\theta_2 \end{bmatrix} & (s\text{-wave}) \\ \dfrac{1}{2n_1 \cos\theta_1} \begin{bmatrix} n_1 \cos\theta_2 + n_2 \cos\theta_1 & n_1 \cos\theta_2 - n_2 \cos\theta_1 \\ n_1 \cos\theta_2 - n_2 \cos\theta_1 & n_1 \cos\theta_2 + n_2 \cos\theta_1 \end{bmatrix} & (p\text{-wave}) \end{cases} \tag{4.9-29}$$

The propagation matrix is given by

$$P = \begin{bmatrix} e^{ik_x d} & 0 \\ 0 & e^{-ik_x d} \end{bmatrix} \tag{4.9-30}$$

where d is the thickness of the medium, and k_x is the x component of the wavevector:

$$k_x = \sqrt{(n\omega/c)^2 - \beta^2} = (n\omega/c)\cos\theta \tag{4.9-31}$$

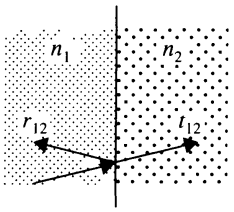
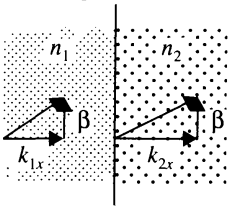
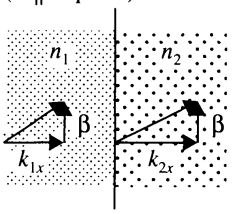
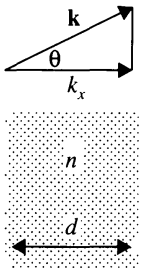
Table 4.1 is a summary of the 2×2 matrices for both the s-wave and p-wave. At normal incidence, the s-wave and p-wave are identical; the transition matrix becomes

$$T_{12} = \frac{1}{2n_1} \begin{bmatrix} n_1 + n_2 & n_1 - n_2 \\ n_1 - n_2 & n_1 + n_2 \end{bmatrix} \quad \text{(normal incidence)} \tag{4.9-32}$$

Once the matrix is obtained, the transmission and reflection coefficients of the multilayer structure can be calculated as follows. Let A_0 be the amplitude of the incident beam, B_0 be that of the reflected beam, and A_{N+1} be that of the transmitted beam. We set $B_{N+1} = 0$, as the beam is launched from the left side. These amplitudes are related by, according to Equation (4.9-24),

$$\begin{aligned} A_0 &= M_{11} A_{N+1} \\ B_0 &= M_{21} A_{N+1} \end{aligned} \tag{4.9-33}$$

TABLE 4.1 2 × 2 Matrices for Dielectric Interface and Homogeneous Layer

Dielectric interface between two media with refractive indices n_1 and n_2		$\frac{1}{t_{12}}\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}$
Dielectric interface between two media with refractive indices n_1 and n_2	s-polarization ($\mathbf{E} \perp xz$ plane) 	$\frac{1}{2k_{1x}}\begin{bmatrix} k_{1x}+k_{2x} & k_{1x}-k_{2x} \\ k_{1x}-k_{2x} & k_{1x}+k_{2x} \end{bmatrix}$
Dielectric interface between two media with refractive indices n_1 and n_2	p-polarization ($\mathbf{E} \parallel xz$ plane) 	$\frac{1}{2n_1n_2k_{1x}}\begin{bmatrix} n_1^2k_{2x}+n_2^2k_{1x} & n_1^2k_{2x}-n_2^2k_{1x} \\ n_1^2k_{2x}-n_2^2k_{1x} & n_1^2k_{2x}+n_2^2k_{1x} \end{bmatrix}$
Homogeneous medium of thickness d		$\begin{bmatrix} e^{ik_xd} & 0 \\ 0 & e^{-ik_xd} \end{bmatrix}$

The transmission and reflection coefficients are thus given by

$$t = \left(\frac{A_{N+1}}{A_0}\right)_{B_{N+1}=0} = \frac{1}{M_{11}}$$
$$r = \left(\frac{B_0}{A_0}\right)_{B_{N+1}=0} = \frac{M_{21}}{M_{11}} \tag{4.9-34}$$

When the incident beam is launched from the right side, the transmission and reflection coefficients are given by

$$t' = \left(\frac{B_0}{B_{N+1}}\right)_{A_0=0} = \frac{|M|}{M_{11}}$$
$$r' = \left(\frac{A_{N+1}}{B_{N+1}}\right)_{A_0=0} = -\frac{M_{12}}{M_{11}} \tag{4.9-35}$$

where $|M|$ is the determinant of the matrix. Let T and T' be the transmittances of the layered structure when light is incident from the left medium (medium 0) and right medium (medium $N + 1$), respectively. These two transmittances are given by

$$T = \frac{k_{N+1x}}{k_{0x}}|t|^2, \quad T' = \frac{k_{0x}}{k_{N+1x}}|t'|^2 \tag{4.9-36}$$

respectively. Using Equations (4.9-34) and (4.9-35) and the expression for $|M|$ in Equation (4.9-27), we obtain

$$T' = T \tag{4.9-37}$$

and

$$t' = |M|t = \frac{k_{N+1x}}{k_{0x}}t \tag{4.9-38}$$

Based on Equation (4.9-37), we note that the transmittances T' and T of a layered structure are identical, regardless of whether the beam is incident from the left side or the right side. This is true even if loss is present in the layered structure. In the event when media 0 and $N + 1$ are identical (i.e., $n_0 = n_{N+1}$), we have

$$t' = t \tag{4.9-39}$$

This is consistent with the principle of reciprocity.

Using Equations (4.9-34) and (4.9-35) for the general expressions of the transmission and reflection coefficients, we obtain

$$tt'^* + rr^* = 1, \quad tr'^* + rt^* = 0 \tag{4.9-40}$$

which are identical to the results obtained from the symmetry argument (time-reversal symmetry) [6]. They are known as the Stokes relationships. It can also be shown that, for a lossless layered structure,

$$R + T = 1 \tag{4.9-41}$$

which is consistent with the conservation of energy.

4.10 MODE MATCHING AND COUPLING LOSS

A basic problem of both theoretical and practical interest is how to couple efficiently an incident beam of light with complex amplitude $E_{\text{in}}(x, y)$ to a given mode of an optical resonator or an optical fiber, and also derive a measure of the residual, undesirable, excitation of other resonator modes. In the case of an optical fiber, this process is known as end-fire coupling. The problem arises frequently in optical communications when the light is coupled in and out of a single-mode fiber for wavelength selection, power amplification, or dispersion management purposes, and when the light is coupled in and out of a laser resonator (or amplifier). In optical fibers, a poor mode matching will lead to a significant insertion loss.

Referring to Figure 4.33, we designate the electric field of an incident beam of light at the "input" plane z_1 as $E_{\text{in}}(x, y)$ and the wavefunction of modes of the fiber or resonator as $E_{mn}(x, y)$, where m, n are the transverse mode integers of the Gaussian beam of an optical resonator, or the LP mode of an optical fiber. The set $E_{mn}(x, y)$ of modes of a fiber or a resonator constitutes a complete orthogonal set of wavefunctions. They satisfy

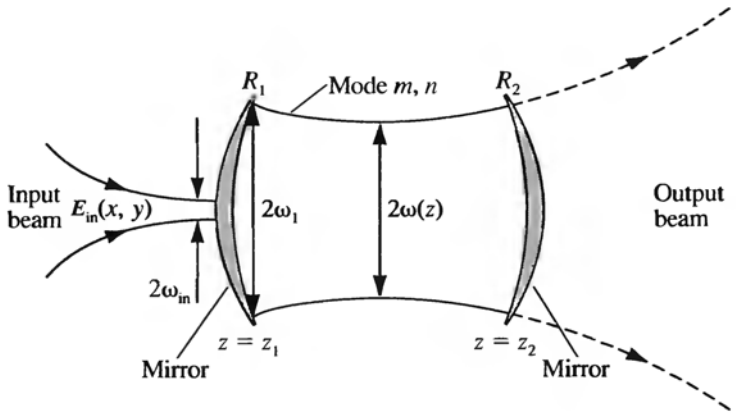


Figure 4.33 Excitation of a transverse mode m, n of a resonator (or an optical fiber) by an incident monochromatic beam of light.

$$\iint E_{mn}(x, y)E^*_{m'n'}(x, y)\,dx\,dy = 0, \quad \text{unless } m = m' \text{ and } n = n' \tag{4.10-1}$$

By taking advantage of the completeness of the set E_{mn} and expanding the incident field as a linear superposition of the modes, we can write

$$E_{in}(x, y) = \sum_{mn} a_{mn}E_{mn}(x, y) \tag{4.10-2}$$

where a_{mn} are constants. Multiplying both sides by E^*_{mn}, integrating over the entire xy plane, and making use of the orthogonal relationships (4.10-1), we obtain

$$a_{mn} = \frac{\iint E_{in}(x, y)E^*_{mn}(x, y)\,dx\,dy}{\iint E_{mn}(x, y)E^*_{mn}(x, y)\,dx\,dy} \tag{4.10-3}$$

The coupling efficiency of the incident field into a given spatial mode, say, E_{mn}, is defined as

$$\eta_{mn} = \frac{\text{Power coupled into mode } mn}{\text{Total incident power}} = \frac{\iint |a_{mn}E_{mn}(x, y)|^2\,dx\,dy}{\iint |E_{in}(x, y)|^2\,dx\,dy} \tag{4.10-4}$$

Substituting Equation (4.10-3) for the constant a_{mn} in Equation (4.10-4), we obtain

$$\eta_{mn} = \frac{\left|\iint E_{in}(x, y)\,E^*_{mn}(x, y)\,dx\,dy\right|^2}{\iint |E_{in}(x, y)|^2\,dx\,dy \cdot \iint |E_{mn}(x, y)|^2\,dx\,dy} \tag{4.10-5}$$

This is the main result of this section. We note η_{mn} is always less than or equal to unity, according to Equation (4.10-5) and the Schwarz inequality. This is in agreement with the conservation of energy.

It follows from Equation (4.10-5) that if the input beam at the input plane has the *same* spatial dependence as that of the mode to be excited, that is, when

$$E_{in}(x, y) \propto E_{mn}(x, y) \tag{4.10-6}$$

then $\eta_{mn} = 1$, and all other $\eta_{m'n'}$ are zero, and all the incident power goes into exciting the single E_{mn} mode. In practice, mode matching requires that, at the input facet of the fiber or the

resonator mirror, the incident beam possesses the same transverse mode numbers m and n as the mode to be excited, as well as possessing the same spot size and radius of curvature. In arriving at the result (4.10-5), we assume that no waves are reflected at the input facet. An additional condition for the case of an optical resonator is that the mode m, n thus excited fulfill the longitudinal Fabry–Perot resonance condition of the resonator. Otherwise, most of the incident beam power will be reflected. In the situation depicted in Figure 4.33, where $\omega_{in} \ll \omega_1$, a large number of modes will be excited. It will be left as an exercise to show that the number of modes that are excited "substantially" is $\sim(\omega_1/\omega_{in})^2$. By "mode matching" we can avoid the excitation of higher-order modes when we use a scanning Fabry–Perot etalon (made of spherical mirrors) to analyze the spectrum of an incident fundamental Gaussian beam.

EXAMPLE: INSERTION LOSS

An obvious condition of an efficient coupling requires that the incident beam be propagating along the axis (z axis) of the resonator or optical fiber. Any lateral shift (x, y) of the incident beam can lead to a reduced coupling efficiency into a resonator or a single-mode fiber. The insertion loss can easily be calculated as follows. We consider the lateral shift only along the x axis. The y dependence can be ignored in the analysis. Let the wavefunction of the resonator or the fiber possess a Gaussian transverse field profile of the form $E_{00} = e^{-x^2/\omega^2}$, where ω is the spot size of the Gaussian beam at the input facet. Let the wavefunction of the laterally shifted beam be written $E_{in} = e^{-(x-a)^2/\omega^2}$, where a is the lateral displacement along the x axis. Substituting these wavefunctions into Equation (4.10-5) and carrying out the integration, we obtain

$$\eta = e^{-a^2/\omega^2}$$

For a Gaussian beam of spot size $\omega = 10$ μm, and a lateral beam displacement of $a = 2$ μm, the above formula yields $\eta = e^{-1/25} \approx 0.96$, which corresponds to a 4% insertion loss (−0.13 dB). Similarly, a displacement of $a = 10$ μm will lead to a coupling efficiency of $\eta = e^{-1} \approx 0.368$, which corresponds to a 63.2% insertion loss (−4.4 dB).

PROBLEMS

4.1 Plot I_r/I_i versus δ of a Fabry–Perot etalon with $R = 0.9$.

4.2 Show that if a Fabry–Perot etalon has an intensity gain per pass of G, its peak transmission is given as $(1 - R)^2G/(1 - RG)^2$.

4.3 Starting with the definition

$$F \equiv \frac{\nu_{m+1} - \nu_m}{\Delta\nu_{1/2}}$$

for the finesse of a Fabry–Perot etalon and using semiquantitative arguments, show why in the case where the root-mean-square surface deviation from perfect flatness is approximately λ/N, the finesse cannot exceed $F \cong N/2$. [Hint: Consider the spreading of the transmission peak due to a small number of etalons of nearly equal length transmitting in parallel.]

4.4 Show that the angular spread of a beam that is incident normally on a plane-parallel Fabry–Perot etalon must not exceed

$$\theta_{1/2} = \sqrt{\frac{2n\lambda}{lF}}$$

if its peak transmission is not to deviate substantially from unity.

4.5 Complete the derivation of Equations (4.1-4), (4.1-5), (4.1-6), and (4.1-7).

4.6 Consider a diverging monochromatic beam that is incident on a plane-parallel Fabry–Perot etalon.

(a) Obtain an expression for the various angles along which the output energy is propagating. [Hint: These correspond to the different values of θ in Equation (4.1-8) that result from changing m.]
(b) Let the output beam in (a) be incident on a lens with a focal length f. Show that the energy distribution in the focal plane consists of a series of circles, each corresponding to a different value of m. Obtain an expression for the radii of the circles.
(c) Consider the effect in (b) of having simultaneously two frequencies ν_1 and ν_2 present in the input beam. Derive an expression for the separation of the respective circles in the focal plane. Show that the smallest separation $\nu_1 - \nu_2$ that can be resolved by this technique is given by $(\Delta\nu) \sim c/2nlF$.

4.7 Calculate the fraction of the power of a fundamental ($l = m = 0$) Gaussian beam that passes through an aperture with a radius equal to the beam spot size.

4.8 Show that in the case of a conventional two-reflector resonator the stability condition, Equation (4.5-5), reduces to Equation (4.4-2).

4.9 Consider a spherical mirror with a radius of curvature R whose reflectivity varies as

$$\rho(r) = \rho_0 \exp(-r^2/a^2)$$

where r is the radial distance from the center. Show that the (A, B, C, D) matrix of this mirror is given by

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\dfrac{2}{R} - i\dfrac{\lambda}{\pi a^2} & 1 \end{bmatrix}$$

4.10 Consider an optical resonator with values as shown in Figure P4.10.

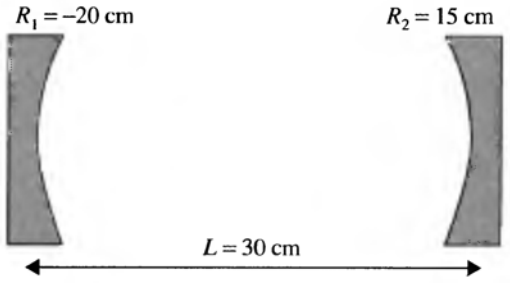


Figure P4.10

(a) Calculate the position of the waist of the mode at $\lambda = 1\ \mu$m.
(b) Calculate the diameter of the waist.

4.11 Derive the following expression for the group delay of the Gires–Tournois etalon:

$$\tau = \frac{\sigma}{1 + (\sigma^2 - 1)\sin^2\phi}\tau_0$$

where $\tau_0 = 2d\phi/d\omega$ is the group delay of the etalon when the front mirror reflectivity is zero. Show that the group delay (time delay) reaches a maximum of $\sigma\tau_0$ at resonsnce ($\phi = m\pi$) and a minimum of τ_0/σ occurs at $\phi = m\pi + \pi/2$, m = integer.

4.12 Show by simple arguments that in the situation depicted in Figure 4.33 the number of Hermite–Gaussian resonator modes that are excited substantially (say, within an order of magnitude of the maximal value) by a fundamental (incident) Gaussian beam is $\sim(\omega_1/\omega_{in})^2$. Ignore longitudinal resonance and consider only transverse (x, y) modes. [Hint: Consider the integral in the numerator of expression (4.10-5) for η_{mn}. Reason why the main contribution to the integral comes from a circle of radius ω_{in} or ω_1/l (l, m are the Hermite polynomial integers) whichever is smaller.]

4.13 Obtain an expression for the coupling efficiency of an incident Gaussian fundamental beam into an optical resonator whose mirrors have a radius of curvature R and a spot radius at the mirrors of ω. The incident beam has a radius R_b and spot size radius ω_b at the mirror.

4.14 The free spectral range (FSR = $\Delta\nu = c/2nd$)) of a Fabry–Perot etalon is determined by the resonance condition: $2kd = 2(2\pi\nu/c)nd = 2m\pi$, where n is the index of refraction of the cavity medium and d is the cavity spacing.

(a) Show that if material dispersion is present, the free spectral range becomes

$$\Delta\nu = \frac{v_g}{2d} = \frac{c}{2n_g d}$$

where v_g is the group velocity and n_g is the group index defined as $n_g = c/v_g$.
(b) In the above, we ignore the phase shifts upon reflection from the mirrors. These phase shifts can be a function of frequency. The resonance condition can be written $2kd + \phi_1 + \phi_2 = 2m\pi$. Derive an expression for the free spectral range by taking into account the frequency dependence of the phase shifts.

4.15 Consider a symmetric Fabry–Perot etalon with mirror reflectivity R.

(a) Derive an analytic expression for the phase delay of the transmitted beam (phase of E_t/E_i).

(b) Derive an expression for the group delay (derivative of phase with respect to angular frequency). Find the maximum and minimum of the group delay, in units of the cavity transit time. (Assume $n = 1$.)

4.16 As indicated in this book, the reflection and transmission coefficients r, r', t, and t' are related by the fundamental reciprocity and time-reversal symmetry. These coefficients are defined in Figure P4.16. When time is reversed, $(tA)^*$ and $(rA)^*$ should recombine and produce A^*. Show that

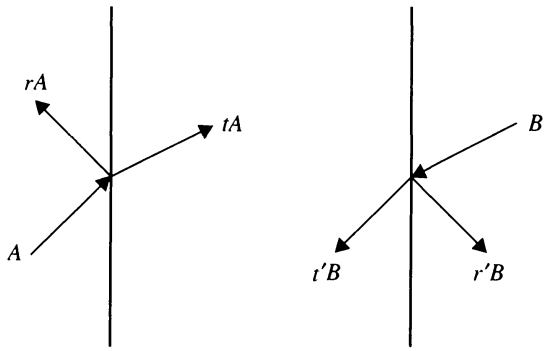


Figure P4.16

$$tt'^* + rr^* = 1$$

$$tr'^* + rt^* = 0$$

These are the Stokes relationships.

4.17 Optical interleavers are passive devices that can separate the spectral content of an incoming beam into two interleaving sets. In other words, if the incoming beam contains frequencies $(\nu_1, \nu_2, \nu_3, \nu_4, \nu_5, \nu_6, \nu_7, \nu_8)$, the first set of output consists of $(\nu_1, \nu_3, \nu_5, \nu_7)$ and the second set of output consists of $(\nu_2, \nu_4, \nu_6, \nu_8)$. Interferometers like the Michelson (or Mach–Zehnder) can be employed for this purpose (see Figure P4.17). Let $\Delta L = L_2 - L_1$ be the path difference between the two arms of interferance; the intensity at one of the output ports is given by

$$I = \frac{1}{2}\left[1 + \cos\left(\frac{2\pi}{c}\nu 2\,\Delta L + \phi_2 - \phi_1\right)\right]$$

where ϕ_1 and ϕ_2 are phase shifts of the mirrors. Note that the round-trip path difference is $2\,\Delta L$.

(a) Assuming $\phi_1 = \phi_2 = 0$, find the frequencies where the output intensity is zero, and the frequencies where the intensity is maximum. Show that the maxima (or minima) are separated by

$$\Delta\nu = \frac{c}{2\Delta L}$$

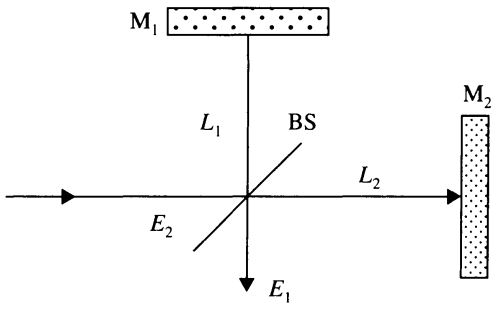


Figure P4.17

(b) A Gires–Tournois etalon can function like a mirror with a strong phase dispersion (i.e., phase shift depends on frequency). Several etalons can be used as mirrors in the interleavers, provided the period of the phase dispersion matches the period of the intensity output. Find the cavity spacing of the etalons.

(c) If the front mirror reflectivities of the etalons are properly chosen, a flat-top passband can be obtained. This is desirable in engineering application. Plot the output intensity for the case of $R_1 = 0\%$, $R_2 = 20\%$.

(d) Repeat (c) for the case when $R_1 = 3\%$, $R_2 = 40\%$.

4.18 Let the reflection and transmission coefficients for the beam splitter (BS) in Problem 4.17 be r, r', t, and t'. Assuming a lossless beam splitter, the conservation of energy requires that

$$|t|^2 + |r|^2 = 1 \quad \text{and} \quad |t'|^2 + |r'|^2 = 1$$

The principle of reciprocity requires that $t' = t$, assuming that the media on both sides of the beam splitter are identical.

(a) Show that $|r'| = |r|$.

(b) Show that the output fields (E_1, E_2) can be written

$$E_1 = rtAe^{-2ikL_1} + tr'Ae^{-2ikL_2} \quad \text{and}$$

$$E_2 = r^2Ae^{-2ikL_1} + tt'Ae^{-2ikL_2}$$

(c) Using the conservation of energy $|E_1|^2 + |E_2|^2 = |A|^2$, show that

$$tt'^* + rr^* = 1 \quad \text{and} \quad tr'^* + rt^* = 0$$

4.19 Consider the case of N identical rings with $(N + 1)$ identical couplers. The mode amplitudes are given by Equation (4.8-37).

(a) Show that Equation (4.8-37) can be rewritten

$$\begin{bmatrix} A_0 \\ B_0 \end{bmatrix} = [SP]^{N+1} P^{-1} \begin{bmatrix} A_{N+1} \\ B_{N+1} \end{bmatrix} = [SP]^{N+1} \begin{bmatrix} A_{N+1} \exp(-i\phi/2) \\ B_{N+1} \exp(+i\phi/2) \end{bmatrix}$$

(b) Assuming a real cross-coupling coefficient κ, show that

$$[SP] = \frac{1}{\kappa} \begin{bmatrix} \exp(i\phi/2) & t^* \exp(-i\phi/2) \\ t \exp(i\phi/2) & \exp(-i\phi/2) \end{bmatrix} \equiv \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

and

$$AD - BC = 1$$

(c) Using Chebyshev's identity (2.1-11), show that

$$|\chi|^2 \equiv \left| \frac{A_{N+1}}{A_0} \right|^2_{B_{N+1}=0} = \left| \frac{1}{M_{11}} \right|^2 = \frac{1}{1 + |C|^2 \dfrac{\sin^2(N+1)\theta}{\sin^2\theta}}$$

(d) Derive Equation (4.8-38).

REFERENCES

1. Ramo, S., J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*. Wiley, New York, 1965.
2. Yariv, A., *Quantum Electronics*, 3rd ed. Wiley, New York, 1989, p. 99.
3. Fabry, C., and A. Perot, Theorie et applications d'une nouvelle methode de spectroscopie interferentielle. *Ann. Chim. Phys.* **16**:115 (1899).
4. Born, M., and E. Wolf, *Principles of Optics*, 3rd ed. Pergamon, New York, 1965, Chapter 7.
5. Peterson, D. G., and A. Yariv, Interferometry and laser control with Fabry–Perot etalons. *Appl. Opt.* **5**:985 (1966).
6. See, for example, Pochi Yeh, *Optical Waves in Layered Media*. Wiley, New York, 1988.
7. Boyd, G. D., and J. P. Gordon, Confocal multimode resonator for millimeter through optical wavelength masers. *Bell Syst. Tech. J.* **40**:489 (1961).
8. Boyd, G. D., and H. Kogelnik, Generalized confocal resonator theory. *Bell Syst. Tech. J.* **41**:1347 (1962).
9. Casperson, L., Gaussian light beams in inhomogeneous media. *Appl. Opt.* **12**:2434 (1973).
10. Levi, A. F., R. E. Slusher, S. L. McCall, J. L. Glass, S. J. Pearson, and Ra. A. Logan, Directional light coupling from micro-disk lasers. *Appl. Phys. Lett.*, **62**:561 (1993).
11. Little, B. E., Ultra compact Si–SiO_2 micro-ring resonator optical channel dropping filters. *Opt. Lett.* **23**:1570 (1998).
12. Cai, M., O. Painter, and K. J. Vahala, Observation of critical coupling in a fiber taper to a silica-microsphere whispering gallery mode system. *Phys. Rev. Lett.* **85**:74 (2000).
13. Yariv, A., Universal relations for coupling of optical power between micro-resonators and dielectric waveguides. *Electron. Lett.* **36**:32 (2000).
14. Choi, J. M., R. K. Lee, and A. Yariv, Ring fiber resonators based on fused silica-fiber grating add-drop filters: application to resonator coupling, *Opt. Lett.* **27**:1598 (2002).
15. Huang, Y., G. T. Paloczi, J. Scheuer, and A. Yariv, Soft lithography replication of polymetric microring optical resonators. *Optics Express* **11**:2452 (2003).
16. Paloczi, G., Y. Huang, and A. Yariv, Polymeric Mach–Zehnder interferometer using serially coupled microring resonators. *Optics Express* **11**:2666 (2003).

ADDITIONAL READING AND REFERENCES

J. K. S. Poon, J. Scheuer, S. Mookherjea, G. T. Paloczi, Y. Huang, and A. Yariv, "Matrix analysis of microring coupled-resonator optical waveguides," Optics Express, 12(1), 90–103, 2004.

J. K. S. Poon, J. Scheuer, Y. Xu, and A. Yariv, "Designing coupled-resonator optical waveguide delay lines," J. Opt. Soc. Am., B, 21(9), 1665–1673, 2004.

J. E. Heebner and R.W. Boyd, "Slow and fast light in resonator-coupled waveguides," J. Modern Optics, 49, 2629–2636, 2002.

Y. Xu, Y. Li, R. K. Lee, and A. Yariv, "Scattering-theory analysis of waveguide-resonator coupling," Physical Review E. 62(5), 7389–7404, 2000.

CHAPTER 5

INTERACTION OF RADIATION AND ATOMIC SYSTEMS

5.0 INTRODUCTION

In optical electronics, especially optical communications, the index of refraction of a homogeneous medium is the key physical parameter that determines the transmission property of an optical beam. In the previous chapter we defined the index of refraction of a medium in terms of its dielectric constant. In this chapter we investigate the physical origins of the index of refraction, dispersion, and absorption. Specifically, we consider what happens to an electromagnetic wave propagating in an atomic medium (e.g., atomic vapors). Transitions between quantum states of an atomic system may involve the absorption or emission of photons. As a result of the interaction, the atomic system may behave like an oscillating electric dipole. A collection of these individual dipole moments leads to a macroscopic electric polarization of the medium. This is the origin of the electric susceptibility, dielectric constant, and index of refraction. We are also particularly concerned with the possibility of growth (or attenuation) of the radiation resulting from its interaction with the atoms, as well as the changes in the velocity of propagation of light due to such interaction. The concepts derived in this chapter will be used later in the text in treating the laser oscillator. We first discuss the connection between atomic transitions and the absorption and emission of electromagnetic radiation. We then discuss the classical electron model and derive the atomic polarizability. The index of refraction and its dependence on frequency (dispersion) of an atomic system can then be obtained. We then discuss induced transition, absorption, and amplification of an atomic system. Finally, we take up the subject of gain saturation in optical amplifiers.

5.1 ATOMIC TRANSITIONS AND ELECTROMAGNETIC WAVES

One of the fundamental results of the theory of quantum mechanics is that each physical system can be found, upon measurement, in only one of a predetermined set of energy states—the *eigenstates* of the system. With each of these states we associate an energy that corresponds to the total energy of the system when occupying the state. Some of the simpler systems, which are treated in any basic text on quantum mechanics, include the free electron, the hydrogen atom, and the harmonic oscillator. Examples of more complicated systems

include the hydrogen molecule and the semiconducting crystal. With each state, the state m of the electron in a hydrogen atom, say, we associate an eigenfunction

$$\psi_m(\mathbf{r}, t) = u_m(\mathbf{r}) \exp(-iE_m t/\hbar), \quad m = 1, 2, 3, \ldots \tag{5.1-1}$$

where $|u_m(\mathbf{r})|^2\, dx\, dy\, dz$ gives the probability of finding the electron, once it is known to be in the state m, within the volume element $dx\, dy\, dz$, which is centered on the point $\mathbf{r}$. E_m is the state energy described above and $\hbar \equiv h/2\pi$, where $h = 6.626 \times 10^{-34}$ joule-second is Planck's constant.

One of the main tasks of quantum mechanics is the determination of the eigenfunctions $u_m(\mathbf{r})$ and the corresponding energy E_m of various physical systems. In this book, however, we will accept the existence of these states, their energy levels, as well as a number of other related results whose justification is provided by the experimentally proved formalism of quantum mechanics. Some of these results are discussed in the following.

Radiative Process and Atomic Transitions

There are, in general, many eigenstates in most atomic systems. Let us concentrate on two of these levels—1 and 2, for example. In the case of a hydrogen atom, these can be the 1s and 2p state of the electron. Such a two-level system greatly simplifies the discussion and introduction of the basic concepts. Figure 5.1 shows the schematic diagrams of the energy levels and the possible radiative processes.

If an atomic system is initially in its ground state (E_1 in Figure 5.1), and the atomic system is subject to the radiation of an incident photon, the photon may be absorbed by the atomic system provided the photon energy satisfies the follow condition:

$$h\nu = E_2 - E_1 \tag{5.1-2}$$

and certain selection rules (e.g., conservation of angular momentum) are also satisfied. As a result of the absorption of the incident photon, the atomic system is elevated to the upper state (with energy E_2). If the atomic system is known to be in state 2 at $t = 0$, there is a finite probability per unit time that it will undergo a transition to state 1, emitting in the process a photon of energy $h\nu = E_2 - E_1$. This process, occurring as it does without the inducement of a radiation field, is referred to as *spontaneous emission*. If the atomic system is initially in state 2 while at the same time under the action of an incident photon, the external action may also induce emission of a photon in addition to the spontaneous emission. This is called stimulated emission and is very important in many cases, including lasers and optical amplifiers.

The interaction of an atomic system with external electromagnetic radiation can be treated with the use of quantum mechanics, or a simple classical electron model. During the irradiation, the atomic system may behave as an oscillating electric dipole. In the next section,

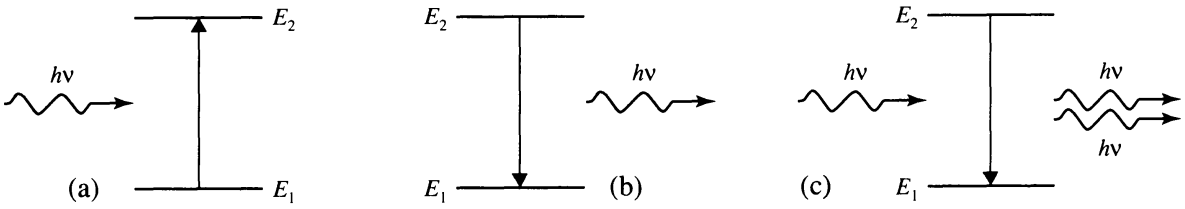


Figure 5.1 Radiative processes: (a) absorption—an incident photon is absorbed while the atomic state is elevated from eigenstate 1 to eigenstate 2; (b) spontaneous emission—a photon is emitted while the atomic system descends from eigenstate 2 to eigenstate 1; and (c) stimulated emission—an additional photon is emitted when an atomic system is under the action of an incident photon.

we will discuss the induced dipole moment of the atomic system due to the presence of an electromagnetic field.

5.2 ATOMIC POLARIZABILITY AND DIELECTRIC CONSTANT

Consider the transmission of a beam of light through a transparent isotropic medium. Under the action of the electric field of the optical beam, the elementary charge particles (mostly electrons) inside each atom are displaced from their equilibrium positions. In most dielectrics, this charge separation is directly proportional to and in the same direction as the electric field of the light beam. Thus the induced dipole moment $\mathbf{p}$ can be written

$$\mathbf{p} = \alpha \mathbf{E} \tag{5.2-1}$$

where the constant α is known as the atomic polarizability (or molecular polarizability) and $\mathbf{E}$ is the electric field. Generally, the molecular polarizability is a tensor. For atomic systems with spherical symmetry, the polarizability reduces to a scalar. This atomic polarizability α can be derived by using a simple classical electron model (see next section). The dielectric constant of a medium will depend on the manner in which the atoms are assembled. Consider a simple case of a gas medium. Let N be the number of atoms per unit volume. Then the polarization can be written approximately as

$$\mathbf{P} = N\,\mathbf{p} = N\alpha\mathbf{E} \equiv \varepsilon_0 \chi \mathbf{E} \tag{5.2-2}$$

where χ is the electric susceptibility of the medium. In Equation (5.2-2), we ignore the difference between the local field and the macroscopic field. This is legitimate for gas media provided $N\alpha/\varepsilon_0 \ll 1$. The dielectric constant ε of the medium is thus given by, according to Equations (1.1-5), and (5.2-2)

$$\varepsilon = \varepsilon_0(1 + \chi) = \varepsilon_0(1 + N\alpha/\varepsilon_0) \tag{5.2-3}$$

If the medium is nonmagnetic, the index of refraction of the medium is written, according to Equation (1.4-15),

$$n^2 = 1 + \chi = 1 + N\alpha/\varepsilon_0 \tag{5.2-4}$$

It is important to note that the index of refraction as well as the dielectric constant are dependent on the frequency of light. We mentioned this for silica in Chapter 3. This is known as the dispersion, which will be discussed further in the next section when we describe the classical electron model.

In the case of solids and liquids, the density N is about 1000 times higher. Thus the electric susceptibility is significantly higher. Under the circumstances, the electric field experienced by an individual atom (or molecule) is not the average electric field $\mathbf{E}$ in the medium. The field experienced by the atom (or molecule) is known as the local field. To find the local field, we consider a uniform medium in the presence of a mean electric field $\mathbf{E}$, which is uniform in the medium (on macroscopic length scales). Consider one of the atoms (or molecules) that constitutes the dielectric. We draw a sphere of radius r about this particular atom. This is intended to represent the boundary between the microscopic and the macroscopic range of phenomena affecting the atom. We shall treat the dielectric outside the sphere as a continuous medium and the dielectric inside the sphere as a collection of polarized atoms. As a result of uniform polarization, there are surface charges on the inner surface of the sphere. It can easily be shown that the field at the atom due to the surface charges on the sphere is

$$\mathbf{E}_{\text{surface charge}} = \frac{\mathbf{P}}{3\varepsilon_0} \tag{5.2-5}$$

Thus the net electric field (the local field) seen by an individual molecule is

$$\mathbf{E}_{\text{local}} = \mathbf{E} + \frac{\mathbf{P}}{3\varepsilon_0} \tag{5.2-6}$$

This is *larger* than the average electric field in the dielectric. The above analysis indicates that the local field is due to the long-range interactions of the atom (or molecule) with the other atoms (or molecules) in the medium. Substituting Equation (5.2-6) as the electric field in (5.2-2), we obtain

$$\mathbf{P} = N\alpha\mathbf{E}_{\text{local}} = N\alpha\left(\mathbf{E} + \frac{\mathbf{P}}{3\varepsilon_0}\right) = \varepsilon_0\chi\mathbf{E} \tag{5.2-7}$$

By eliminating $\mathbf{P}$, we obtain

$$\chi = \frac{\dfrac{N\alpha}{\varepsilon_0}}{1 - \dfrac{N\alpha}{3\varepsilon_0}} \tag{5.2-8}$$

This is the dielectric susceptibility constant, by taking into account of the local field effect. The above equation is often written, according to Equation (5.2-3),

$$\frac{\varepsilon - \varepsilon_0}{\varepsilon + 2\varepsilon_0} = \frac{N\alpha}{3\varepsilon_0} \tag{5.2-9}$$

This is known as the *Clausius–Mossotti* relation. We note that the dielectric susceptibility reduces to $\chi = N\alpha/\varepsilon_0$ in the case of dilute systems where $\chi = N\alpha/\varepsilon_0 \ll 1$.

5.3 CLASSICAL ELECTRON MODEL

In order to find the atomic polarizability, we need to solve for the dipole moment of an atom induced by the electric field of an incident optical wave. Since the atoms are so small compared to the wavelength of light, we can assume that the electric field is uniform in each atom. Let the electric field of the optical wave in the atom be

$$E = E_0 \exp(i\omega t) \tag{5.3-1}$$

where E_0 is a constant amplitude and ω is the frequency. Each of the charged particles in the atom will feel this electric field and will be driven up and down (or back and forth) by the time-varying electric force. Because of the great difference in their masses, we expect that electrons contribute dominantly to the induced dipole moment. In classical physics, we further assume that the electrons are "fastened" elastically to the atoms and obey the following equation of motion:

$$m\frac{d^2}{dt^2}X + m\gamma\frac{d}{dt}X + m\omega_0^2 X = -eE \tag{5.3-2}$$

where X is the position of the electron relative to the atom, m is the electron mass, and $-e$ is the electron charge. The parameter ω_0 is the resonant frequency of the electron motion, and γ is the damping coefficient. For the case of a two-level system described in Figure 5.1,

$\omega_0 = (E_2 - E_1)/\hbar$. For a harmonic driving field, this equation has the following steady-state solution:

$$X = \frac{-eE_0}{m(\omega_0^2 - \omega^2 + i\gamma\omega)} \exp(i\omega t) \tag{5.3-3}$$

The induced dipole moment is thus given by

$$p = -eX = \frac{e^2}{m(\omega_0^2 - \omega^2 + i\gamma\omega)} E_0 \exp(i\omega t) \tag{5.3-4}$$

Using definition (5.2-2), we obtain the following expression for the atomic polarizability:

$$\alpha = \frac{e^2}{m(\omega_0^2 - \omega^2 + i\gamma\omega)} \tag{5.3-5}$$

If there are N atoms per unit volume, the index of refraction of such a medium is given by, according to Equations (5.3-5) and (5.2-4),

$$n^2 = 1 + \chi = 1 + \frac{Ne^2}{m\varepsilon_0(\omega_0^2 - \omega^2 + i\gamma\omega)} \tag{5.3-6}$$

where the electric susceptibility χ is often written in terms of its real and imaginary parts as

$$\chi = \chi' - i\chi'' = \frac{Ne^2}{m\varepsilon_0(\omega_0^2 - \omega^2 + i\gamma\omega)} \tag{5.3-7}$$

We note that the dielectric susceptibility χ is maximum when the driving frequency ω is close to the resonant frequency ω_0 of the atoms. At resonance, $\omega = \omega_0$, the susceptibility χ is a pure imaginary number. If the second term in Equation (5.3-6) is small compared to 1, the index of refraction can be written approximately as

$$n = 1 + \tfrac{1}{2}\chi = 1 + \frac{Ne^2}{2m\varepsilon_0(\omega_0^2 - \omega^2 + i\gamma\omega)} \quad (\text{when } \chi \ll 1) \tag{5.3-8}$$

In the spectral regime near resonance ($\omega \approx \omega_0$), these expressions can be written

$$n^2 = 1 + \chi = 1 + \frac{Ne^2}{2\omega_0 m\varepsilon_0(\omega_0 - \omega + i\gamma/2)} \tag{5.3-9}$$

$$\chi = \chi' - i\chi'' = \frac{Ne^2}{2\omega_0 m\varepsilon_0(\omega_0 - \omega + i\gamma/2)} \tag{5.3-10}$$

$$n = 1 + \tfrac{1}{2}\chi = 1 + \frac{Ne^2}{4\omega_0 m\varepsilon_0(\omega_0 - \omega + i\gamma/2)} \quad (\text{when } \chi \ll 1) \tag{5.3-11}$$

Equation (5.3-6) not only gives the index of refraction in terms of the basic atomic quantities, but also shows the variation with the frequency ω of light. However, it is not very practical to compute the index of refraction because both N and ω_0 vary from material to material. In addition, information about the nature of the resonance frequency ω_0 requires a calculation using quantum mechanics. Therefore we cannot expect to get a simple formula for the index of refraction that applies to all substances. However, Equations (5.3-9) and (5.3-11) are still very useful for many practical purposes. First, consider gas media such as air, helium, and hydrogen. The natural resonance frequencies of the electron oscillations correspond to ultraviolet light. These frequencies are higher than the frequencies of visible light; that is,

ω_0 is much larger than the ω of visible light. To a first approximation, we can neglect ω^2 and $i\gamma\omega$ in the denominator in comparison with ω_0^2. Then we find that the index of refraction is nearly constant. If we take $N = 2.69 \times 10^{25}$ m^{-3} for air at standard temperature (0 °C) and pressure (760 mmHg) and take the absorption band at $\lambda_0 = 80$ nm, Equation (5.3-7) yields $n = 1.0002$ for air. For most transparent substances such as glass and water, ω_0 is also in the ultraviolet region. However, for solids and liquid, N is three orders of magnitude larger than that of air; thus the electric susceptibility is significantly higher. Equations (5.3-7) and (5.3-10) are also useful when we study the gain or loss in laser media. Equations (5.3-6)–(5.3-11) are also useful for describing the gain profiles of the lasing transitions in gas lasers (e.g., He–Ne, argon lasers).

A similar expression can be obtained by using the time-dependent perturbation theory in quantum mechanics. Assuming the atomic system is initially in the lower state (state 1 in Figure 5.1a), the atomic polarizability is given by [1]

$$\alpha = \frac{f_{21}e^2}{2\omega_0 m(\omega_0 - \omega + i\gamma/2)} \tag{5.3-12}$$

where γ is the linewidth of the transition, f_{21} is the oscillator strength given by

$$f_{21} = \frac{2m\omega_0}{\hbar}|\langle u_2|x|u_1\rangle|^2 = \frac{2m\omega_0}{\hbar}\iiint u_2^* x u_1 \, dx \, dy \, dz \tag{5.3-13}$$

where u_1 and u_2 are the wavefunctions of the eigenstates, and we assume that the electric field is polarized along the x direction. The oscillator strength is a dimensionless parameter. It's a measure of the relative strength of the various transitions in the atomic system. It can be shown that

$$\sum_{m=2}^{\infty} f_{m1} = 1 \tag{5.3-14}$$

where the summation is over all states, except the initial state (state 1). Equation (5.3-14) is known as the *sum rule*. We note that the classical electron model provides an accurate result as long as the two levels are the "dominant levels" of the atomic system.

5.4 DISPERSION AND COMPLEX REFRACTIVE INDEX

If we examine closely the expression for the index of refraction, Equation (5.3-11), we notice that as ω rises and becomes closer to ω_0 the index of refraction also rises. So n rises slowly with ω_0. This is true for almost all transparent materials. Thus the index of refraction is higher for blue light than for red light. This is the reason a prism bends the light more in the blue than in the red. The phenomenon that the refractive index depends on frequency is called *chromatic dispersion*. Equation (5.3-7) is called the dispersion equation. As we will see later, dispersion affects the propagation of signals in optical communications.

So far, we have been ignoring the imaginary term $i\gamma\omega$ in the denominator of Equation (5.3-6). This term accounts for the damping of electron motion (for $\gamma > 0$) and gives rise to the phenomenon of *optical absorption*. Because of this damping term, $i\gamma\omega$, the index of refraction is a complex number according to Equation (5.3-7). The imaginary part of the index of refraction is significant only when ω is close to ω_0. By working out the real and imaginary parts of Equation (5.3-8), the complex refractive index can be written

$$n' = n - i\kappa = 1 + \frac{Ne^2(\omega_0^2 - \omega^2)}{2m\varepsilon_0[(\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2]} - i\frac{Ne^2\gamma\omega}{2m\varepsilon_0[(\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2]} \tag{5.4-1}$$

where we use n' to denote the complex refractive index and $-\kappa$ is the imaginary part of the refractive index. From now on, we use n to represent the real part of the index of refraction. The constant κ is referred to as the *extinction coefficient* because it represents an absorption (or attenuation) of the electromagnetic wave. To see the effect of κ on electromagnetic radiation, we consider the propagation of a monochromatic plane wave in a medium with a complex refractive index $(n - i\kappa)$. According to the discussion in Section 1.4, such a wave can be written

$$E = A\exp[i(\omega t - k'z)] \tag{5.4-2}$$

where k' denotes the wavenumber in the complex medium and is given by

$$k' = \omega\sqrt{\mu\varepsilon} = \frac{\omega}{c}n' = \frac{2\pi}{\lambda}n' \tag{5.4-3}$$

Using the complex refractive index (5.4-1), the wavenumber k' becomes complex and is

$$k' = \frac{2\pi}{\lambda}(n - i\kappa) \tag{5.4-4}$$

where n is now the real part of the complex refractive index and κ is the extinction coefficient. Substituting Equation (5.4-4) for k' in Equation (5.4-2), the electric field of the wave becomes

$$E = A\exp\left[i\left(\omega t - \frac{2\pi}{\lambda}nz\right)\right]\exp\left(-\frac{2\pi}{\lambda}\kappa z\right) \tag{5.4-5}$$

We notice that the imaginary part of the complex refractive index leads to an attenuation of electromagnetic radiation along its direction of propagation. The *attenuation coefficient* is usually defined as

$$\alpha \equiv \frac{1}{I}\frac{dI}{dz} \tag{5.4-6}$$

where I is the intensity of electromagnetic radiation and, aside from a constant factor, is proportional to E^*E. According to Equation (5.4-6), the intensity of an attenuating beam can be written

$$I(z) = I(0)\exp(-\alpha z) \tag{5.4-7}$$

where $I(0)$ is the intensity at $z = 0$. The attenuation coefficient α defined in Equation (5.4-6) is related to the extinction coefficient κ by

$$\alpha = \frac{4\pi}{\lambda}\kappa \tag{5.4-8}$$

according to Equations (5.4-7) and (5.4-5). The attenuation coefficient is proportional to the imaginary part of the complex refractive index. According to Equation (5.4-8), even a small κ will lead to a strong attenuation for visible light. For example, with $\kappa = 0.0001$ and $\lambda = 0.5\ \mu m$, Equation (5.4-8) yields an attenuation coefficient of $\alpha = 25\ cm^{-1}$.

Both the real part n and the imaginary part κ of the complex index of refraction are functions of ω. The dependence on ω is significant, especially when ω is close to ω_0. To examine the dispersion at frequencies near resonance, we put $\omega \cong \omega_0$ in Equation (5.4-1) and obtain

$$n - i\kappa = 1 + \frac{Ne^2(\omega_0 - \omega)}{4m\omega_0\varepsilon_0[(\omega_0 - \omega)^2 + (\gamma/2)^2]} - i\frac{Ne^2(\gamma/2)}{4m\omega_0\varepsilon_0[(\omega_0 - \omega)^2 + (\gamma/2)^2]} \tag{5.4-9}$$

Equivalently, the electric susceptibility can be written

$$\chi = \chi' - i\chi'' = \frac{Ne^2(\omega_0 - \omega)}{2m\omega_0\varepsilon_0[(\omega_0 - \omega)^2 + (\gamma/2)^2]} - i\frac{Ne^2(\gamma/2)}{2m\omega_0\varepsilon_0[(\omega_0 - \omega)^2 + (\gamma/2)^2]} \tag{5.4-10a}$$

The electric susceptibility is often written in terms of the frequency ν as

$$\begin{aligned}\chi' &= \frac{Ne^2(\nu_0 - \nu)}{4\pi m\omega_0\varepsilon_0[(\nu_0 - \nu)^2 + (\Delta\nu/2)^2]} \\ \chi'' &= \frac{Ne^2(\Delta\nu/2)}{4\pi m\omega_0\varepsilon_0[(\nu_0 - \nu)^2 + (\Delta\nu/2)^2]}\end{aligned} \tag{5.4-10b}$$

where $\Delta\nu$ is the FWHM bandwidth in units of Hz. Figure 5.2 shows a normalized plot of $n - 1$ and κ as functions of ω. Note that the extinction coefficient κ is maximum at $\omega = \omega_0$ and decreases like $|\omega - \omega_0|^{-2}$ as $|\omega - \omega_0|$ increases. In fact, $\kappa(\omega)$ in Equation (5.4-9) has a Lorentzian lineshape. On the other hand, the real part of the index of refraction approaches 1, like $|\omega - \omega_0|^{-1}$ as $|\omega - \omega_0|$ increases. This explains the fact that dispersion exists in most transparent materials at frequencies where κ is negligible. According to Equation (5.4-9), the index of refraction is greater than unity for low frequencies ($\omega < \omega_0$) and increases with ω as the resonant frequency ω_0 is approached. Such an increasing behavior is the case of *normal dispersion* and occurs in most transparent media. Very near the resonance frequency, there is a small spectral region where the slope is negative. Such a negative slope is often referred to as *anomalous dispersion.* As we can see from Figure 5.2, anomalous dispersion always occurs when absorption is significant.

In the classical electron model discussed in Section 5.3, we only assumed a single resonance frequency ω_0, which is for a two-level system and is somewhat simpler than in nature.

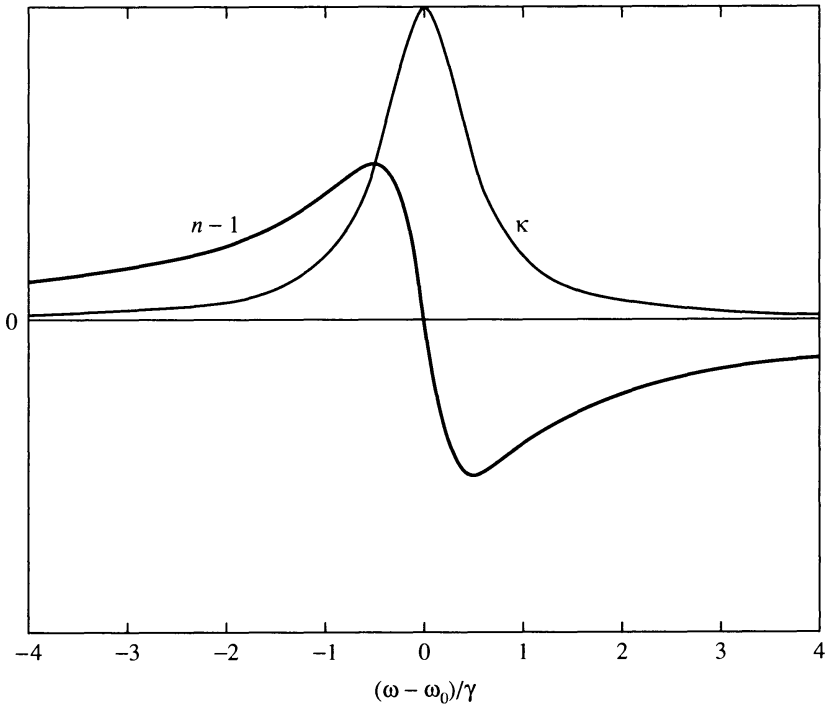


Figure 5.2 Normalized plot of n and κ versus ω.

In fact, there are several resonance frequencies for each atom. We can now modify the dispersion equation by assuming that not all the electrons are identically bound and that all the electrons oscillate independently. Let f_j be the fraction of electrons whose natural resonance frequency is ω_j and whose damping factor is γ_j. Then by adding the contribution from all the electron oscillators, we obtain

$$n^2 = 1 + \frac{Ne^2}{m\varepsilon_0} \sum_j \frac{f_j}{(\omega_j^2 - \omega^2 + i\gamma_j\omega)} \tag{5.4-11}$$

The fractions f_j are known as oscillator strengths, defined in Equation (5.3-13), and are dimensionless. By definition, the oscillator strengths obey the following sum rule:

$$\sum_j f_j = Z \tag{5.4-12}$$

where Z is the number of electrons per atom. Theoretical evaluation of oscillator strengths requires knowledge of the wavefunction of the atomic system. To illustrate the magnitude of the oscillator strengths and the rule (5.4-12), we note that the oscillator strength corresponding to the 1s $\rightarrow$ 2p transition ($\omega_{1s\text{-}2p} = 1.55 \times 10^{16}$ s^{-1}) in the hydrogen atom is 0.4162. The sum of the oscillator strengths for transitions from the ground state 1s to all other states different from 2p is 0.5838. In the preceding discussion, we assumed that there is a damping coefficient γ_j associated with each resonance frequency. These coefficients are all positive, and electron damping leads to attenuation of electromagnetic radiation.

If some electrons are in their excited states, electromagnetic radiation will drive these electrons and induce the emission of radiation at the resonance frequency. For such an excited system (or inverted system), Equation (5.4-11) is still valid, provided we replace the parameter N by $N_g - N_j$, where N_g is the number of electrons in the ground state per unit volume and N_j is the number of electrons in the jth excited state per unit volume. In an inverted system, N_j is greater than N_g; thus we see that the complex refractive index of an excited system has a positive imaginary part in a small frequency range near the resonance frequency ω_j for which $N_g - N_j$ is negative. According to Equation (5.4-5), the propagation of electromagnetic radiation in a medium with $\kappa < 0$ will lead to amplification of the energy. Such a medium is called a *gain medium* and is described by a complex index of refraction with a positive imaginary part (i.e., $\kappa < 0$).

Figure 5.3 shows the dispersion of the index of refraction of several transparent optical materials in the spectral range $\lambda = 0.3$ to 30 μm. Traditionally, the dispersion of n is usually expressed in terms of wavelength λ.

Complex Refractive Index of Electron Gas (Metals)

In a metal, there are free electrons that do not oscillate around atoms but are free to move under the action of the applied electric field. For these electrons, there is no restoring force. The equation of motion (5.3-2) is still valid, provided we set $\omega_0 = 0$. Thus all the results obtained in Section 5.3 can still be used for metals. Putting $\omega_0 = 0$ in Equation (5.3-6), we obtain the complex refractive index

$$n^2 = 1 - \frac{Ne^2}{m\varepsilon_0(\omega^2 - i\gamma\omega)} \tag{5.4-13}$$

where N is electron density. In fact, Equation (5.4-13) becomes a good approximation to the actual complex refractive index, provided we can neglect the contribution from the deeply

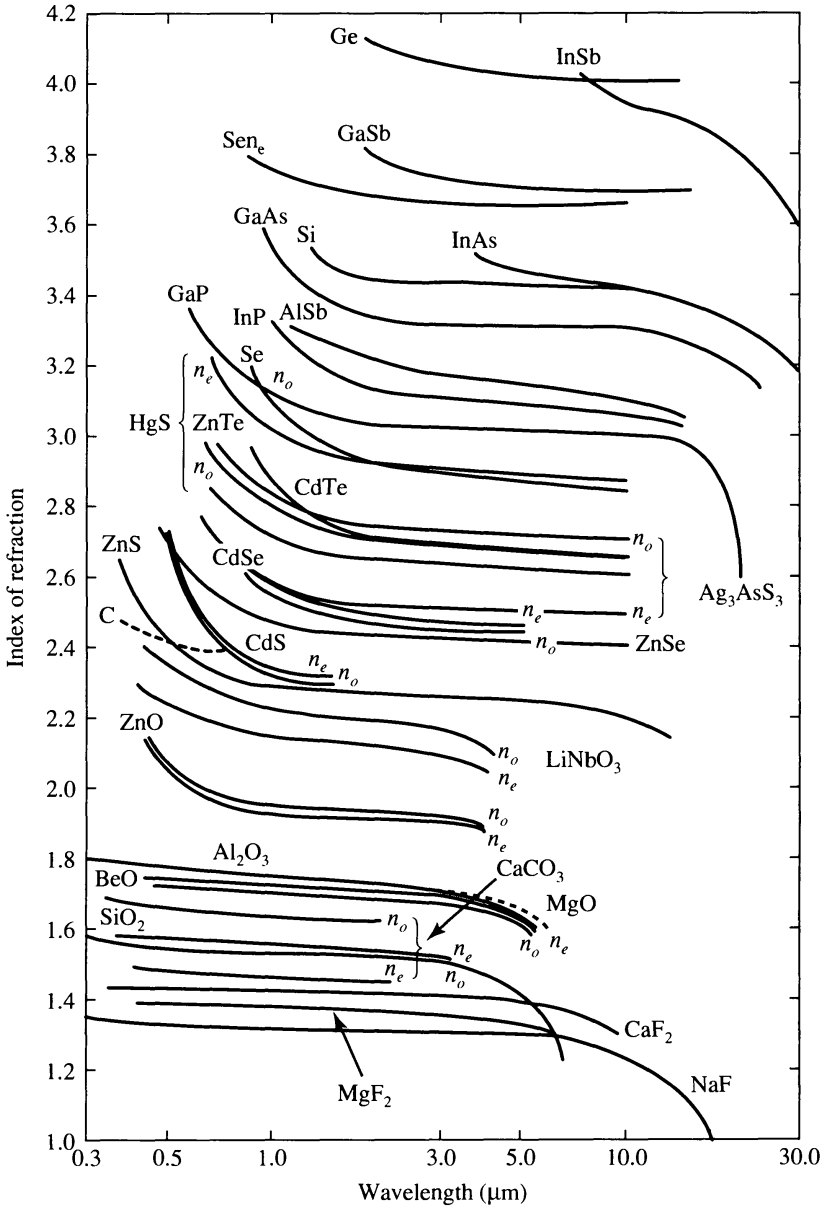


Figure 5.3 The index of refraction n versus wavelength λ. For birefringent crystals, n_o is the ordinary refractive index and n_e is the extraordinary refractive index. (Data adapted from Reference [2].)

bound electrons in metals and some semiconductors. The damping constant is some sort of average rate of collisions involving appreciable momentum transfer. If we further assume that $\gamma \ll \omega$, the index of refraction is given by

$$n^2 = 1 - \frac{Ne^2}{m\varepsilon_0\omega^2} = 1 - \frac{\omega_p^2}{\omega^2} \tag{5.4-14}$$

where ω_p is the *plasma angular frequency* given by

$$\omega_p^2 = \frac{Ne^2}{m\varepsilon_0} \tag{5.4-15}$$

For high-frequency radiation ($\omega > \omega_p$), the index of refraction is real and the waves propagate freely. For frequencies lower than the plasma frequency ω_p, n is purely imaginary. Thus, within the metal, the fields fall off exponentially with distance from the surface

(see also Problem 5.4). Consequently, such electromagnetic radiation incident on a high-conductivity metal will be reflected from the surface. For metals such as aluminum, copper, gold, and silver, the density of the free electrons is on the order of $N = 10^{23}$ cm^{-3}. This means that $\omega_p \approx 2 \times 10^{16}$ s^{-1}, so that for visible and infrared radiation with $\omega < \omega_p$, the index of refraction is imaginary according to Equation (5.4-14). The index of refraction, in general, is complex when γ is finite.

5.5 LINESHAPE FUNCTION—HOMOGENEOUS AND INHOMOGENEOUS BROADENING

According to Equations (5.4-8) and (5.4-9), the absorption coefficient of a two-level atomic system can be written

$$\alpha(\omega) = \frac{4\pi}{\lambda} \frac{Ne^2\gamma}{8m\omega_0\varepsilon_0[(\omega_0 - \omega)^2 + (\gamma/2)^2]} \tag{5.5-1}$$

where, we recall, ω is the frequency of light, and ω_0 is the natural resonance frequency of the atomic system. By examining the above expression, we note that the absorption is not strictly monochromatic (i.e., of one frequency). In fact, absorption peaks at ω_0 and drops off rapidly as the frequency of light ω is detuned from ω_0. The lineshape described by Equation (5.5-1) is known as a Lorentzian lineshape. In the classical electron model, the absorption spectrum is the same as the spontaneous emission spectrum. So, if one performs a spectral analysis of the radiation emitted by spontaneous $2 \rightarrow 1$ transitions, one finds that the radiation is not strictly monochromatic but occupies a finite frequency bandwidth. The normalized lineshape function describing the spectral distribution of emitted radiation of a two-level atomic system in the vicinity of ω_0 can thus be written, according to Equation (5.5-1),

$$f(\omega) = \frac{\gamma}{[(\omega_0 - \omega)^2 + (\gamma/2)^2]} \tag{5.5-2}$$

The above lineshape is exactly the Fourier spectrum of a damped oscillating dipole (see Problem 5.7). In optical electronics, the lineshape function $g(\nu)$ is often expressed in terms of the frequency ν and is normalized according to

$$\int_{-\infty}^{+\infty} g(\nu)\, d\nu = 1 \tag{5.5-3}$$

We can consequently view $g(\nu)\, d\nu$ as the a priori probability that a given spontaneous emission from level 2 to level 1 will result in a photon whose frequency is between ν and $\nu + d\nu$.

From the experimental point of view, $g(\nu)$ can be obtained by spectrally analyzing the emitted radiation from the atomic system, or by performing a transmission measurement of a beam of light through a sample containing the atomic system as a function of the frequency. The fact that both the emission and the absorption are described by the same lineshape function $g(\nu)$ can be verified experimentally and follows from basic quantum mechanical considerations.

Homogeneous and Inhomogeneous Broadening

Using the result obtained from the classical electron model, the lineshape function of the atomic transition is written

$$g(\nu) = \frac{1}{2\pi} \frac{\gamma/2\pi}{(\nu - \nu_0)^2 + (\gamma/4\pi)^2} \tag{5.5-4}$$

where $\nu_0 = \omega_0/2\pi$, and γ is the damping factor. The separation $\Delta\nu$ between the two frequencies at which the Lorentzian lineshape function is down to half its peak value is referred to as the linewidth and is given by

$$\Delta\nu = \gamma/2\pi \tag{5.5-5}$$

In terms of the linewidth $\Delta\nu$, the normalized lineshape function can be written

$$g(\nu) = \frac{\Delta\nu}{2\pi[(\nu - \nu_0)^2 + (\Delta\nu/2)^2]} \tag{5.5-6}$$

One of the possible causes for the frequency spread of spontaneous emission is the finite lifetime τ of the emitting state. According to Equation (5.3-2), the oscillation amplitude of the electron will decay when the driving field is turned off. If we solve Equation (5.3-2) in the absence of the driving field, we obtain

$$|X(t)|^2 \propto \exp(-\gamma t)$$

Thus the damping factor is related to the spontaneous lifetime as

$$\gamma = 1/\tau \tag{5.5-7}$$

Using Equation (5.5-5), the linewidth can be written

$$\Delta\nu = \frac{1}{2\pi\tau} \tag{5.5-8}$$

Here we have an expression of the linewidth broadening due to the finite lifetime of the upper state (τ_u). In many laser systems, the lower state (τ_l) may also have a finite lifetime. In this case, the total linewidth broadening is a sum of the two individual broadenings. Further broadening can be due to collisions. In the process of coherent radiation, the radiating field from an atom can experience an abrupt change of phase due to elastic collisions from other atoms (see Problem 5.8). Let τ_{co} be the collision time. We thus generalize Equation (5.5-8) to read

$$\Delta\nu = \frac{1}{2\pi}\left(\tau_u^{-1} + \tau_l^{-1} + \tau_{co}^{-1}\right) \tag{5.5-9}$$

The type of broadening (i.e., the finite width of the emitted spectrum) described above is called *homogeneous broadening*. It is characterized by the fact that the spread of the response over a band $\Delta\nu$ is characteristic of *each* atom in the sample. The function $g(\nu)$ thus describes the response of any of the atoms, which are indistinguishable.

As mentioned earlier, homogeneous broadening is due most often to the finite interaction lifetime of the emitting or absorbing atoms. Some of the most common mechanisms are:

1. The spontaneous lifetime of the excited state.
2. Collision of an atom embedded in a crystal with a phonon. This may involve the emission or absorption of acoustic energy. Such a collision does not terminate the lifetime of the atom in its absorbing or emitting state. It does, however, interrupt the relative phase between the atomic oscillation and that of the field, thus causing a broadening of the radiation.

3. Pressure broadening of atoms in a gas. At sufficiently high atomic densities, the collisions between atoms become frequent enough that lifetime termination and phase interruption as in the preceding mechanism dominate the broadening mechanism.

There are, however, many physical situations in which the individual atoms are distinguishable, each having a slightly different transition frequency ν_0. If one observes, in this case, the spectrum of the spontaneous emission, its spectral distribution will reflect the spread in the individual transition frequencies and not the broadening due to the finite lifetime of the excited state. Two typical situations give rise to this type of broadening, referred to as *inhomogeneous broadening*.

First, the energy levels, hence the transition frequencies, of ions present as impurities in a host crystal depend on the immediate crystalline surroundings. The ever present random strain, as well as other types of crystal imperfections, cause the crystal surroundings to vary from one ion to the next, thus effecting a spread in the transition frequencies. Second, the transition frequency ν of a gaseous atom (or molecule) is Doppler-shifted due to the finite velocity of the atom along the direction of observation according to

$$\nu = \nu_0 + \frac{v_x}{c}\nu_0 \tag{5.5-10}$$

where v_x is the component of the velocity along the direction connecting the observer with the moving atom, c is the velocity of light in the medium, and ν_0 is the frequency corresponding to a stationary atom. The Maxwell velocity distribution function of a gas with atomic mass M that is at equilibrium at temperature T is [3]

$$f(v_x, v_y, v_z) = \left(\frac{M}{2\pi kT}\right)^{3/2} \exp\left(-\frac{M}{2kT}(v_x^2 + v_y^2 + v_z^2)\right) \tag{5.5-11}$$

where $k = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant, and $f(v_x, v_y, v_z)\, dv_x\, dv_y\, dv_z$ is thus the fraction of all the atoms whose x component of velocity is contained in the interval v_x to $v_x + dv_x$ while, simultaneously, their y and z components lie between v_y and $v_y + dv_y$, v_z and $v_z + dv_z$, respectively. Alternatively, we may view $f(v_x, v_y, v_z)\, dv_x\, dv_y\, dv_z$ as the a priori probability that the velocity vector $\mathbf{v}$ of any given atom terminates within the differential volume $dv_x\, dv_y\, dv_z$ centered on $\mathbf{v}$ in velocity space so that

$$\iiint_{-\infty}^{\infty} f(v_x, v_y, v_z)\, dv_x\, dv_y\, dv_z = 1 \tag{5.5-12}$$

According to Equation (5.5-10) the probability $g(\nu)\, d\nu$ that the transition frequency is between ν and $\nu + d\nu$ is equal to the probability that v_x will be found between $v_x = (\nu - \nu_o)(c/\nu_o)$ and $(\nu + d\nu - \nu_o)(c/\nu_o)$ irrespective of the values of v_y and v_z (since if $v_x = (\nu - \nu_o)(c/\nu_o)$, the Doppler-shifted frequency will be equal to ν regardless of v_y and v_z). This probability is thus obtained by substituting $v_x = (\nu - \nu_o)(c/\nu_o)$ in $f(v_x, v_y, v_z)\, dv_x\, dv_y\, dv_z$ and then integrating over all values of v_y and v_z. The result is

$$g(\nu)\, d\nu = \left(\frac{Mc^2}{2\pi kT}\right)^{3/2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(M/2kT)(v_y^2+v_z^2)} dv_y\, dv_z \times e^{-(Mc^2/2kT)(\nu-\nu_0)^2/\nu_0^2} \left(\frac{1}{\nu_0}\right) d\nu \tag{5.5-13}$$

Using the definite integral

$$\int_{-\infty}^{\infty} e^{-(M/2kT)v_z^2}\, dv_z = \left(\frac{2\pi kT}{M}\right)^{1/2}$$

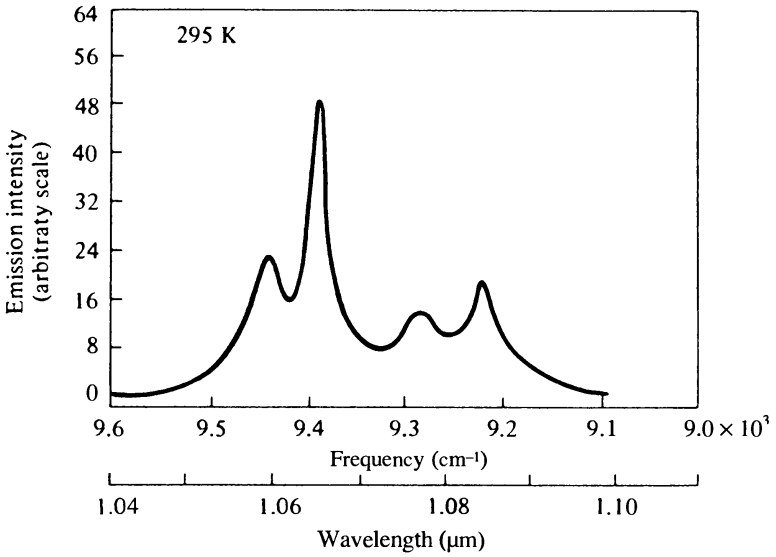


Figure 5.4 Emission spectrum of Nd^{3+}:$CaWO_4$ in the vicinity of the 1.06 μm laser transition. The main peak corresponds to the laser transition. (After Reference [4].) The unit 1 cm^{-1} corresponds to a frequency of $\nu = 30$ GHz, or a photon energy of $h\nu = 1.24 \times 10^{-4}$ eV.

we obtain, from Equation (5.5-13),

$$g(\nu)\,d\nu = \frac{1}{\nu_0}\left(\frac{Mc^2}{2\pi kT}\right)^{1/2} e^{-(Mc^2/2kT)(\nu-\nu_0)^2/\nu_0^2}\,d\nu \tag{5.5-14}$$

for the *normalized Doppler-broadened lineshape*. The functional dependence of $g(\nu)$ in Equation (5.5-14) is referred to as Gaussian. The width of $g(\nu)$ in this case is taken as the frequency separation between the points where $g(\nu)$ is down to half its peak value. It is obtained from Equation (5.5-14) as

$$\Delta\nu_D = 2\nu_0\sqrt{\frac{2kT}{Mc^2}\ln 2} \tag{5.5-15}$$

where the subscript D stands for *Doppler*. We can reexpress $g(\nu)$ in terms of $\Delta\nu_D$, obtaining

$$g(\nu) = \frac{2(\ln 2)^{1/2}}{\pi^{1/2}\Delta\nu_D} e^{-[4(\ln 2)(\nu-\nu_0)^2/\Delta\nu_D^2]} \tag{5.5-16}$$

In Figure 5.4 we show, as an example of a lineshape function, the spontaneous emission spectrum of Nd^{3+} when present as an impurity ion in a $CaWO_4$ lattice. The spectrum consists of a number of transitions, which are partially overlapping.

EXAMPLE: THE DOPPLER LINEWIDTH OF Ne

Consider the 6328 Å transition in Ne, which is used in the popular He–Ne lasers. Using atomic mass 20 for neon in Equation (5.5-15) and taking $T = 300$ K, we obtain

$$\Delta\nu_D \approx 1.5 \times 10^9 \text{ Hz}$$

for the Doppler linewidth. The 10.6 μm transition in the CO_2 laser has, according to Equation (5.5-15), a linewidth $\Delta\nu_D \approx 6 \times 10^7$ Hz.

5.6 INDUCED TRANSITIONS—ABSORPTION AND AMPLIFICATION

Consider a two-level atomic system as shown in Figure 5.1. In the presence of an electromagnetic field of frequency $\nu \sim (E_2 - E_1)/h$, an atom can undergo a transition from state 1 to 2, *absorbing* in the process a quantum of excitation (photon) with energy $h\nu$ from the field. If the atom happens to be in state 2 at the moment when it is first subjected to the electromagnetic field, it may make a downward transition to state 1, *emitting* a photon of energy $h\nu$. What distinguishes the process of induced transition from the spontaneous one described in Section 5.1 is the fact that the induced rate for $2\rightarrow1$ and $1\rightarrow2$ transitions is *equal*, whereas the spontaneous $1\rightarrow2$ (i.e., the one in which the atomic energy increases) transition rate is zero. Another fundamental difference—one that again follows from quantum mechanical considerations—is that the induced rate is *proportional* to the *intensity* of the electromagnetic field, whereas the spontaneous rate is independent of it. The relationship between the induced transition rate and the (inducing) field intensity is of fundamental importance in treating the interaction of atomic systems with electromagnetic fields. Its derivation follows.

Consider first the interaction of an assembly of identical atoms with a radiation field whose energy density is distributed uniformly in frequency in the vicinity of the transition frequency (within the atomic lineshape function). Let the energy density per unit frequency be $\rho(\nu)$. We assume that the induced transition rates per atom from $2\rightarrow1$ and $1\rightarrow2$ are both proportional to $\rho(\nu)$ and take them as

$$\begin{aligned}(W'_{21})_{\text{induced}} &= B_{21}\rho(\nu)\\ (W'_{12})_{\text{induced}} &= B_{12}\rho(\nu)\end{aligned} \tag{5.6-1}$$

where B_{21} and B_{12} are constants to be determined. The total downward ($2\rightarrow1$) transition rate is the sum of the induced and spontaneous contributions

$$W'_{21} = B_{21}\rho(\nu) + A_{21} \tag{5.6-2}$$

where A_{21} is the spontaneous transition rate. The total upward ($1\rightarrow2$) transition rate is

$$W'_{12} = (W'_{12})_{\text{induced}} = B_{12}\rho(\nu) \tag{5.6-3}$$

Our first task is to obtain an expression for B_{12} and B_{21}. Since the magnitudes of the coefficients B_{21} and B_{12} depend on the atoms and not on the radiation field, we consider, without loss of generality, the case where the atoms are in a thermal equilibrium with a blackbody (thermal) radiation field at temperature T. In this case the radiation density is given by [5]

$$\rho(\nu) = \frac{8\pi n^3 h\nu^3}{c^3}\frac{1}{e^{h\nu/kT} - 1} \tag{5.6-4}$$

Since at thermal equilibrium the average populations of levels 2 and 1 are constant with time, it follows that the number of $2\rightarrow1$ transitions in a given time interval is equal to the number of $1\rightarrow2$ transitions; that is,

$$N_2 W'_{21} = N_1 W'_{12} \tag{5.6-5}$$

where N_1 and N_2 are the equilibrium population densities of levels 1 and 2, respectively. Using Equations (5.6-2) and (5.6-3) in Equation (5.6-5), we obtain

$$N_2[B_{21}\rho(\nu) + A_{21}] = N_1 B_{12}\rho(\nu)$$

and, substituting for $\rho(\nu)$ from Equation (5.6-4),

$$N_2\left(B_{21}\frac{8\pi n^3 h\nu^3}{c^3(e^{h\nu/kT}-1)} + A_{21}\right) = N_1\left(B_{12}\frac{8\pi n^3 h\nu^3}{c^3(e^{h\nu/kT}-1)}\right) \tag{5.6-6}$$

Since the atoms are in thermal equilibrium, the ratio N_2/N_1 is given by the Boltzmann factor [5] as

$$\frac{N_2}{N_1} = e^{-h\nu/kT} \tag{5.6-7}$$

Equating N_2/N_1 as given by Equation (5.6-6) to that in Equation (5.6-7) gives

$$\frac{8\pi n^3 h\nu^3}{c^3(e^{h\nu/kT}-1)} = \frac{A_{21}}{B_{12}e^{h\nu/kT} - B_{21}} \tag{5.6-8}$$

The last equality can be satisfied only when

$$B_{12} = B_{21} \tag{5.6-9}$$

and simultaneously

$$\frac{A_{21}}{B_{21}} = \frac{8\pi n^3 h\nu^3}{c^3} \tag{5.6-10}$$

The last two equations were first derived by Einstein [6]. We can, using Equation (5.6-10), rewrite the induced transition rate (5.6-1) as

$$W_i' = \frac{A_{21}c^3}{8\pi n^3 h\nu^3}\rho(\nu) = \frac{c^3}{8\pi n^3 h\nu^3 t_{\text{spont}}}\rho(\nu) \tag{5.6-11}$$

where $t_{\text{spont}} \equiv 1/A_{21}$ is the spontaneous lifetime of the atom. By virtue of Equation (5.6-9), the distinction between 2→1 and 1→2 induced transition rates is superfluous.

Equation (5.6-11) gives the transition rate per atom due to a field with a uniform (white) spectrum with energy density per unit frequency $\rho(\nu)$. In the event when the energy density per unit frequency $\rho(\nu)$ is not exactly uniform within the atomic lineshape function $g(\nu)$, the induced transition rate should be written

$$W_i = \int \frac{c^3}{8\pi n^3 h\nu^3 t_{\text{spont}}}\rho(\nu)g(\nu)\,d\nu \tag{5.6-12}$$

In optical electronics our main concern is the transition rates induced by a monochromatic (i.e., single-frequency) field of frequency ν. Let us denote this transition rate as $W_i(\nu)$. For a monochromatic field at frequency ν, the energy density per unit frequency $\rho(\nu')$ is a delta function,

$$\rho(\nu') = U_0\delta(\nu' - \nu) \tag{5.6-13}$$

where U_0 is the energy density (joules per unit volume) of the monochromatic field. Substituting Equation (5.6-13) for the energy density per unit frequency in Equation (5.6-12), we obtain

$$W_i = \frac{c^3 U_0}{8\pi n^3 h\nu^3 t_{\text{spont}}}g(\nu) \tag{5.6-14}$$

where U_0 is the energy density (joules per cubic meter) of the electromagnetic field inducing the transitions. This is the induced transition rate due to the presence of a monochromatic field with an energy density U_0 and a frequency ν.

Returning to our central result, Equation (5.6-14), we can rewrite it in terms of the beam intensity $I_0 = cU_0/n$ (watts per square meter) of the optical wave as

$$W_i = \frac{A_{21}c^2 I_0}{8\pi n^2 h\nu^3} g(\nu) = \frac{\lambda^2 I_0}{8\pi n^2 h\nu t_{\text{spont}}} g(\nu) \tag{5.6-15}$$

where I_0 is the beam intensity at frequency ν, c is the velocity of propagation of light in vacuum, λ is the vacuum wavelength, and $t_{\text{spont}} \equiv 1/A_{21}$. We note that the induced transition rate is proportional to the intensity of the incoming optical beam.

Absorption and Amplification

Consider the case of a monochromatic plane wave of frequency ν and intensity I_0 propagating through an atomic medium with N_2 atoms per unit volume in level 2 and N_1 in level 1. According to Equation (5.6-15) there will occur $N_2 W_i$ induced transitions per unit time per unit volume from level 2 to level 1 and $N_1 W_i$ transitions from level 1 to level 2. The net power generated within a unit volume is thus

$$\frac{P}{\text{Volume}} = (N_2 - N_1) W_i h\nu \tag{5.6-16}$$

This radiation is added coherently (i.e., with a definite phase relationship) to that of the traveling wave so that it is equal, in the absence of any dissipation mechanisms, to the increase in the intensity per unit length. Consider the increase of power in a differential distance of dz along the direction of propagation; using Equation (5.6-15), we obtain

$$\frac{dI_0}{dz} = (N_2 - N_1) \frac{c^2 g(\nu)}{8\pi n^2 \nu^2 t_{\text{spont}}} I_0 \tag{5.6-17}$$

The solution of Equation (5.6-17) is

$$I_0(z) = I_0(0) e^{\gamma(\nu) z} \tag{5.6-18}$$

where $\gamma(\nu)$ is the gain coefficient

$$\gamma(\nu) = (N_2 - N_1) \frac{c^2}{8\pi n^2 \nu^2 t_{\text{spont}}} g(\nu) \tag{5.6-19}$$

We note that the intensity grows exponentially when the population is inverted ($N_2 > N_1$) or is attenuated when $N_2 < N_1$. The first case corresponds to laser-type amplification, whereas the second case is the one encountered in atomic systems at thermal equilibrium. The two situations are depicted in Figure 5.5. We recall that at thermal equilibrium

$$\frac{N_2}{N_1} = e^{-h\nu/kT} \tag{5.6-20}$$

so that systems at thermal equilibrium are always absorbing. The inversion condition $N_2 > N_1$ can still be represented by Equation (5.6-20), provided we take T as negative. As a matter of fact, the condition $N_2 > N_1$ is often referred to as one of "negative temperature"—the

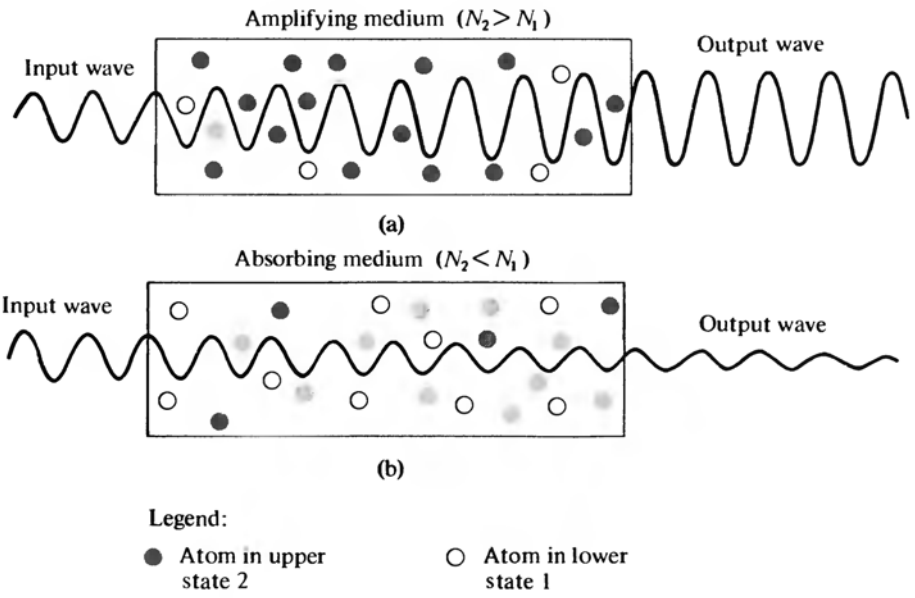


Figure 5.5 Amplification of a traveling electromagnetic wave in (a) an inverted population ($N_2 > N_1$) and (b) attenuation in an absorbing ($N_2 < N_1$) medium.

"temperature" in this case serving as an indicator of the population ratio, in accordance with Equation (5.6-20). The absorption, or amplification, of electromagnetic radiation by an atomic transition can be described not only by means of the exponential gain constant $\gamma(\nu)$, but also alternatively, in terms of the imaginary part of the electric susceptibility $\chi''(\nu)$ of the propagation medium. According to Equation (1.3-24), the density of absorbed power is

$$\frac{\overline{\text{Power}}}{\text{Volume}} = \frac{\omega\varepsilon_0\chi''(\nu)}{2}|E|^2 \tag{5.6-21}$$

where $\chi''(\nu)$ is the imaginary part of the electric susceptibilities. This last result must agree with a derivation using the concept of the induced transition rate $W_i(\nu)$ according to which

$$\frac{\overline{\text{Power}}}{\text{Volume}} = (N_1 - N_2)W_i(\nu)h\nu \tag{5.6-22}$$

Equating (5.6-21) and (5.6-22), substituting (5.6-15) for $W_i(\nu)$, and using the relation $I_0 = (c/n)\varepsilon|E|^2/2$ (see Equation (1.4-21)), we obtain

$$\chi''(\nu) = \frac{(N_1 - N_2)\lambda^3}{16\pi^2 n t_{\text{spont}}} g(\nu) \tag{5.6-23}$$

where $n^2 \equiv \varepsilon/\varepsilon_0$ and λ is the wavelength in vacuum. In the case of a Lorentzian lineshape function $g(\nu)$, the last result can be written

$$\chi''(\nu) = \frac{(N_1 - N_2)\lambda^3}{8\pi^3 n t_{\text{spont}}\,\Delta\nu} \frac{1}{1 + [4(\nu - \nu_0)^2]/(\Delta\nu)^2} \tag{5.6-24}$$

This is a key result and it will be used on numerous occasions.

EXAMPLE: EXPONENTIAL GAIN CONSTANT IN A RUBY LASER

Let us estimate the exponential gain constant at line center of a ruby (Al_2O_3 doped with Cr^{3+} ions) crystal having the following characteristics:

$$N_2 - N_1 = 5 \times 10^{17}/\text{cm}^3$$

$$\Delta\nu \cong \frac{1}{g(\nu_0)} = 2 \times 10^{11} \text{ Hz at 300 K}$$

$$t_{\text{spont}} = 3 \times 10^{-3} \text{ s}$$

$$\nu = 4.326 \times 10^{14} \text{ Hz}$$

$$\frac{c}{n}(\text{in ruby}) \cong 1.69 \times 10^{10} \text{ cm/s}$$

Using these values in Equation (5.6-19) gives

$$\gamma(\nu) = 5 \times 10^{-2} \text{ cm}^{-1}$$

Thus the intensity of a wave with a frequency corresponding to the center of the transition is amplified by approximately 5% per centimeter in its passage through a ruby rod with the foregoing characteristics.

By comparing Equations (5.6-19) and (5.6-23), we obtain the following relationship between the gain coefficient and χ'':

$$\gamma(\nu) = -\frac{2\pi}{\lambda n}\chi'' = -\frac{k}{n^2}\chi'' \tag{5.6-25}$$

In a passive medium with $\chi'' > 0$, the gain coefficient is negative. Many laser media consist of a host medium and an assembly of active atoms whose atomic transitions are responsible for the emission of laser radiation. For example, chromium atoms are the active ingredient in ruby lasers, whereas an Al_2O_3 crystal is the host medium. The complex index of refraction of the laser medium can be written

$$n'^2 = n^2 + \chi' - i\chi'' \tag{5.6-26}$$

where n is the index of refraction of the host medium, and $\chi' - i\chi''$ is the dielectric susceptibility of the active atoms. The complex wavenumber of propagation in such a medium can thus be written

$$k' = k_0 n' = nk_0\left(1 + \frac{\chi'}{n^2} - i\frac{\chi''}{n^2}\right)^{1/2} \cong nk_0\left(1 + \frac{\chi'}{2n^2} - i\frac{\chi''}{2n^2}\right) \tag{5.6-27}$$

where k_0 is the wavenumber of light in vacuum. If we write $k = nk_0$, then the complex wavenumber can be written

$$k' = k\left(1 + \frac{\chi'}{2n^2} - i\frac{\chi''}{2n^2}\right) = k + k\frac{\chi'}{2n^2} - ik\frac{\chi''}{2n^2} \tag{5.6-28}$$

As we know, the intensity of a plane wave with a complex propagation constant will grow exponentially with distance of propagation. Using Equation (5.6-28), we can obtain a gain coefficient that is consistent with Equation (5.6-25).

Combining Equations (5.6-24) and (5.6-25), we can write the exponential gain coefficient of an active medium with a Lorentzian lineshape as

$$\gamma(\nu) = \frac{(N_2 - N_1)\lambda^2}{4\pi^2 n^2 t_{\text{spont}}\,\Delta\nu} \frac{1}{1 + [4(\nu - \nu_0)^2]/(\Delta\nu)^2} \tag{5.6-29}$$

with a gain bandwidth (FWHM) of $\Delta\nu$. The gain at the line center is, according to Equation (5.6-29),

$$\gamma_0 = \gamma(\nu_0) = \frac{(N_2 - N_1)\lambda^2}{4\pi^2 n^2 t_{\text{spont}}\,\Delta\nu} \tag{5.6-30}$$

where we note that the peak gain coefficient is inversely proportional to the linewidth of the gain profile. We can also write, according to Equations (5.6-29) and (5.6-30),

$$\gamma(\nu) = \frac{\gamma_0}{1 + [4(\nu - \nu_0)^2]/(\Delta\nu)^2} \tag{5.6-31}$$

where $\gamma_0 = \gamma(\nu_0)$ is the peak gain coefficient at the line center.

5.7 GAIN SATURATION IN HOMOGENEOUS LASER MEDIA

In the last section we derived an expression (5.6-19) for the exponential gain constant due to a population inversion. It is given by

$$\gamma(\nu) = (N_2 - N_1)\frac{c^2}{8\pi n^2 \nu^2 t_{\text{spont}}} g(\nu) \tag{5.7-1}$$

where N_2 and N_1 are the population densities of the two atomic levels involved in the induced transition. There is nothing in Equation (5.7-1) to indicate what causes the inversion $(N_2 - N_1)$, and this quantity can be considered as a parameter of the system. In practice, the inversion is caused by a "pumping" agent, hereafter referred to as the pump, that can take various forms such as the electric current in injection lasers, the flashlamp light in pulsed ruby lasers, or the energetic electrons in plasma-discharge gas lasers.

Consider next the situation prevailing at some point *inside* a laser medium in the presence of an optical wave. The pump establishes a population inversion, which in the absence of any optical field has a value ΔN^0. The presence of the optical field induces $2 \rightarrow 1$ and $1 \rightarrow 2$ transitions. Since $N_2 > N_1$ and the induced rates for $2 \rightarrow 1$ and $1 \rightarrow 2$ transitions are equal, it follows that more atoms are induced to undergo a transition from level 2 to level 1 than in the opposite direction. This can lead to a net decrease of the population difference ΔN^0.

The reduction in the population inversion and hence of the gain constant brought about by the presence of an electromagnetic field is called *gain saturation*. Its understanding is of fundamental importance in optical electronics. As an example, which will be treated in the next chapter, we may point out that gain saturation is the mechanism that reduces the gain inside laser oscillators to a point where it just balances the losses so that steady oscillation can result. The same mechanism also reduces the gain inside optical amplifiers in optical networks.

In Figure 5.6 we show the ground state 0 as well as the two laser levels 2 and 1 of a four-level laser system. The density of atoms pumped per unit time into level 2 is taken as R_2, and that pumped into level 1 is R_1. Pumping into level 1 is, of course, undesirable since it leads to a reduction of the inversion. In many practical situations it cannot be avoided. The actual "decay" lifetime of atoms in level 2 in the absence of any radiation field is taken as t_2. This

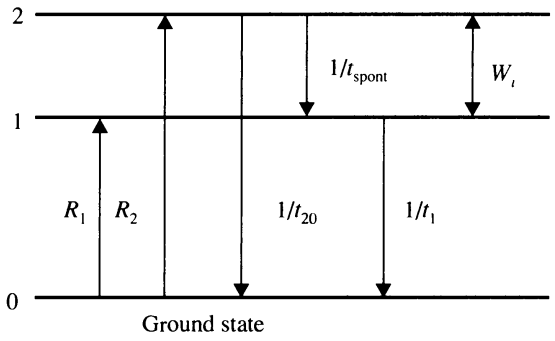


Figure 5.6 Energy levels and transition rates of a four-level laser system. (The fourth level, which is involved in the original excitation by the pump, is not shown and the pumping is shown as proceeding directly into levels 1 and 2.) The total lifetime of level 2 is t_2, where $1/t_2 = 1/t_{spont} + 1/t_{20}$, where $1/t_{20}$ is the decay rate to the ground state (0).

decay rate has a contribution t_{spont}^{-1} that is due to spontaneous (photon emitting) 2→1 transitions as well as to additional nonradiative relaxation from 2 to 1. The lifetime of atoms in level 1 is t_1. The induced rate for 2→1 and 1→2 transitions due to a radiation field at frequency ν is denoted by $W_i(\nu)$ and, according to Equation (5.6-15), is given by

$$W_i(\nu) = \frac{\lambda^2 g(\nu)}{8\pi n^2 h\nu t_{spont}} I_0 \tag{5.7-2}$$

where $g(\nu)$ is the normalized lineshape of the transition and I_0 is the intensity (watts per square meter) of the optical field.

The equations describing the populations of levels 2 and 1 in the combined presence of a radiation field at ν and a pump are

$$\frac{dN_2}{dt} = R_2 - \frac{N_2}{t_2} - (N_2 - N_1)W_i(\nu) \tag{5.7-3}$$

$$\frac{dN_1}{dt} = R_1 - \frac{N_1}{t_1} + \frac{N_2}{t_{spont}} + (N_2 - N_1)W_i(\nu) \tag{5.7-4}$$

where N_2 and N_1 are the population densities (m^{-3}) of levels 2 and 1, respectively. R_2 and R_1 are the pumping rates (m^{-3}-s^{-1}) into these levels. N_2/t_2 is the change per unit time in the population of level 2 due to decay out of level 2 to all levels. This includes spontaneous transitions to level 1 but *not* induced transitions. The rate for the latter is $N_2W_i(\nu)$ so that the net change in N_2 due to induced transitions is given by the last term of Equation (5.7-3). At steady state the populations are constant with time, so putting $d/dt = 0$ in the two preceding equations, we can solve for N_1 and N_2 and obtain

$$N_2 - N_1 = \frac{R_2 t_2 - (R_1 + \delta R_2)t_1}{1 + [t_2 + (1 - \delta)t_1]W_i(\nu)} \tag{5.7-5}$$

where $\delta = t_2/t_{spont}$. (Note: $0 < \delta < 1$.) Here we assume that levels 1 and 2 are high enough (in energy) so that the role of thermal processes in populating them can be neglected. If the optical field is absent, $W_i(\nu) = 0$, and the population inversion density is given by, according to Equation (5.7-5),

$$\Delta N^0 = R_2 t_2 - (R_1 + \delta R_2)t_1 \tag{5.7-6}$$

We can use Equation (5.7-6) to rewrite (5.7-5) as

$$N_2 - N_1 = \frac{\Delta N^0}{1 + \phi t_{spont} W_i(\nu)} \tag{5.7-7}$$

where the dimensionless parameter ϕ is defined by

$$\phi = \delta\left(1 + (1 - \delta)\frac{t_1}{t_2}\right)$$

We note that ϕ is always positive since $0 < \delta < 1$. We also note that in efficient laser systems $t_2 \sim t_{spont}$, so $\delta \approx 1$, and that $t_1 \ll t_2$, so $\phi \approx 1$. Substituting Equation (5.7-2) for $W_i(\nu)$, the last equation becomes

$$N_2 - N_1 = \frac{\Delta N^0}{1 + [\phi\lambda^2 g(\nu)/8\pi n^2 h\nu]I_0} = \frac{\Delta N^0}{1 + I_0/I_s(\nu)} \tag{5.7-8}$$

where $I_s(\nu)$, the saturation intensity, is given by

$$I_s(\nu) = \frac{8\pi n^2 h\nu}{\phi\lambda^2 g(\nu)} = \frac{8\pi n^2 h\nu}{(t_2/t_{spont})\lambda^2 g(\nu)} = \frac{8\pi n^2 h\nu\,\Delta\nu}{(t_2/t_{spont})\lambda^2} \tag{5.7-9}$$

and corresponds to the intensity level (watts per square meter) that causes the inversion to drop to one-half of its nonsaturated value (ΔN^0). In Equation (5.7-9), the last equality is obtained by putting $1/g(\nu) = \Delta\nu$, which is valid for frequencies near the peak of the lineshape function. By using Equation (5.7-8) in the gain expression (5.7-1), we obtain our final result:

$$\gamma(\nu) = \frac{1}{1 + I_0/I_s(\nu)}\left(\frac{\Delta N^0\,\lambda^2}{8\pi n^2 t_{spont}}\right)g(\nu) = \frac{\gamma_0(\nu)}{1 + I_0/I_s(\nu)} \tag{5.7-10}$$

where $\gamma_0(\nu)$ is the gain coefficient at low intensities ($I_0 \ll I_s$). Equation (5.7-10) shows the dependence of the gain constant on the optical intensity. In closing, we recall that Equation (5.7-10) applies to a homogeneous laser system. This is due to the fact that, in the rate equations (5.7-3) and (5.7-4), we considered all the atoms as equivalent and, consequently, experiencing the same transition rates. This assumption is no longer valid in inhomogeneous laser systems. This case is treated in the next section.

5.8 GAIN SATURATION IN INHOMOGENEOUS LASER MEDIA

In Section 5.7 we considered the reduction in optical gain—that is, saturation—due to the optical field in a homogeneous laser medium. In this section we treat the problem of gain saturation in inhomogeneous systems. According to the discussion of Section 5.5, in an inhomogeneous atomic system the individual atoms are distinguishable, with each atom having a unique transition frequency $(E_2 - E_1)/h$. We can thus imagine the inhomogeneous medium as made up of classes of atoms each designated by a continuous variable ξ. The variable ξ can, as an example, correspond to the center frequency of the lineshape function $g^{\xi}(\nu)$ of atoms in group ξ. Furthermore, we define a function $p(\xi)$ so that the a priori probability that an atom has its ξ parameter between ξ and $\xi + d\xi$ is $p(\xi)\,d\xi$. It follows that

$$\int_{-\infty}^{\infty} p(\xi)\,d\xi = 1 \tag{5.8-1}$$

since any atom has a unit probability of having its ξ value between $-\infty$ and ∞.

The atoms within a given class ξ are considered as homogeneously broadened, having a lineshape function $g^{\xi}(\nu)$ that is normalized so that

$$\int_{-\infty}^{\infty} g^{\xi}(\nu)\,d\nu = 1 \tag{5.8-2}$$

In Section 5.5 we defined the transition lineshape $g(\nu)$ by taking $g(\nu)\,d\nu$ to represent the a priori probability that a spontaneous emission will result in a photon whose frequency is between ν and $\nu + d\nu$. Using this definition, we obtain

$$g(\nu)\,d\nu = \left[\int_{-\infty}^{\infty} p(\xi) g^{\xi}(\nu)\,d\xi\right] d\nu \tag{5.8-3}$$

which is a statement of the fact that the probability of emitting a photon of frequency between ν and $\nu + d\nu$ is equal to the probability $g^{\xi}(\nu)\,d\nu$ of this occurrence, given that the atom belongs to class ξ, summed up over all the classes.

Next, we proceed to find the contribution to the inversion that is due to a single class ξ. The rate equations are

$$\begin{aligned} \frac{dN_2^{\xi}}{dt} &= R_2 p(\xi) - \frac{N_2^{\xi}}{t_2} - (N_2^{\xi} - N_1^{\xi}) W_i^{\xi}(\nu) \\ \frac{dN_1^{\xi}}{dt} &= R_1 p(\xi) - \frac{N_1^{\xi}}{t_1} + \frac{N_2^{\xi}}{t_{\text{spont}}} + (N_2^{\xi} - N_1^{\xi}) W_i^{\xi}(\nu) \end{aligned} \tag{5.8-4}$$

and are similar to Equations (5.7-3) and (5.7-4), except that N_2^{ξ} and N_1^{ξ} refer to the upper and lower level densities of atoms in class ξ only. The pumping rate (atoms/m^3-s) into levels 2 and 1 is taken to be proportional to the probability of finding an atom in class ξ and is given by $R_1 p(\xi)$ and $R_2 p(\xi)$ respectively. The total pumping rate into level 2 is, as in Section 5.7, R_2 since

$$\int_{-\infty}^{\infty} R_2 p(\xi)\,d\xi = R_2 \int_{-\infty}^{\infty} p(\xi)\,d\xi = R_2$$

where we made use of Equation (5.8-1). The induced transition rate N_2^{ξ} is given, according to Equation (5.6-15), by

$$W_i^{\xi}(\nu) = \frac{\lambda^2}{8\pi n^2 h\nu t_{\text{spont}}} g^{\xi}(\nu) I_0 \tag{5.8-5}$$

which is of a form identical to (5.7-2) except that $g^{\xi}(\nu)$ refers to the lineshape function of atoms in class ξ. The steady-state $d/dt = 0$ solution of Equation (5.8-4) yields

$$N_2^{\xi} - N_1^{\xi} = \frac{\Delta N^0 p(\xi)}{1 + \phi t_{\text{spont}} W_i^{\xi}(\nu)} \tag{5.8-6}$$

where ΔN^0 and ϕ have the same significance as in Section 5.7. The total power emitted by induced transitions per unit volume by atoms in class ξ is thus

$$\frac{P^{\xi}(\nu)}{V} = (N_2^{\xi} - N_1^{\xi}) h\nu W_i^{\xi}(\nu) = \frac{\Delta N^0 p(\xi) h\nu}{1/W_i^{\xi}(\nu) + \phi t_{\text{spont}}} \tag{5.8-7}$$

where the spontaneous lifetime is assumed the same for all the groups ξ.

Summing Equation (5.8-7) over all the classes, we obtain an expression for the total power at ν per unit volume emitted by the atoms

$$\frac{P(\nu)}{V} = \frac{\Delta N^0 h\nu}{t_{\text{spont}}} \int_{-\infty}^{\infty} \frac{p(\xi)\,d\xi}{1/(W_i^{\xi}(\nu) t_{\text{spont}}) + \phi} \tag{5.8-8}$$

which, by the use of Equation (5.8-5), can be rewritten

$$\frac{P(\nu)}{V} = \frac{\Delta N^0 h\nu}{t_{\text{spont}}} \int_{-\infty}^{\infty} \frac{p(\xi)\,d\xi}{8\pi n^2 h\nu/[\lambda^2 I_0 g^{\xi}(\nu)] + \phi} \tag{5.8-9}$$

The stimulated emission of power causes the intensity of the traveling optical wave to increase with distance z according to $I_0 = I_0(0)\exp[\gamma(\nu)z]$, where

$$\begin{aligned}\gamma(\nu) &= \frac{dI_0/dz}{I_0} = \frac{P(\nu)/V}{I_0} \\ &= \frac{\Delta N^0 \lambda^2}{8\pi n^2 t_{\text{spont}}} \int_{-\infty}^{\infty} \frac{p(\nu_\xi)\,d\nu_\xi}{[1/g^{\xi}(\nu)] + (\phi\lambda^2 I_0/8\pi n^2 h\nu)}\end{aligned} \tag{5.8-10}$$

where we replaced $p(\xi)\,d\xi$ by $p(\nu_\xi)\,d\nu_\xi$. Here we may view ν_ξ as the center frequency of the lineshape function $g^{\xi}(\nu)$ of atoms in group ξ. This is our basic result.

As a first check on Equation (5.8-10), we shall consider the case in which $I_0 \ll 8\pi n^2 h\nu/\phi\lambda^2 g^{\xi}(\nu)$ and therefore the effects of saturation can be ignored. Using Equation (5.8-3) in (5.8-10), we obtain

$$\gamma(\nu) = \frac{\Delta N^0 \lambda^2}{8\pi n^2 t_{\text{spont}}} g(\nu)$$

which is the same as Equation (5.6-19). This shows that in the absence of saturation the expressions for the gain of a homogeneous and an inhomogeneous atomic system are identical.

Our main interest in this treatment is in deriving the saturated gain constant for an inhomogeneously broadened atomic transition. If we assume that in each class ξ all the atoms are identical (homogeneous broadening), we can use Equation (5.5-4) for the lineshape function $g^{\xi}(\nu)$, and therefore

$$g^{\xi}(\nu) = \frac{\Delta\nu}{2\pi[(\nu - \nu_\xi)^2 + (\Delta\nu/2)^2]} \tag{5.8-11}$$

where $\Delta\nu$ is called the homogeneous linewidth of the inhomogeneous line. Atoms with transition frequencies that are clustered within $\Delta\nu$ from each other can be considered as indistinguishable. The term "homogeneous packet" is often used to describe them. Using Equation (5.8-11) in (5.8-10) leads to

$$\gamma(\nu) = \frac{\Delta N^0 \lambda^2\,\Delta\nu}{16\pi^2 n^2 t_{\text{spont}}} \int_{-\infty}^{\infty} \frac{p(\nu_\xi)\,d\nu_\xi}{(\nu - \nu_\xi)^2 + (\Delta\nu/2)^2 + (\phi\lambda^2 I_0\,\Delta\nu/16\pi^2 n^2 h\nu)} \tag{5.8-12}$$

In extreme inhomogeneous cases, the width of $p(\nu_\xi)$ is by definition very much larger than the remainder of the integrand in Equation (5.8-12) and thus it is essentially a constant over the region in which the integrand is appreciable. In this case, we can put $p(\nu_\xi)_{\nu_\xi=\nu} = p(\nu)$ outside the integral sign in Equation (5.8-12), obtaining

$$\gamma(\nu) = \frac{\Delta N^0 \lambda^2\,\Delta\nu}{16\pi^2 n^2 t_{\text{spont}}} p(\nu) \times \int_{-\infty}^{\infty} \frac{d\nu_\xi}{(\nu - \nu_\xi)^2 + (\Delta\nu/2)^2 + (\phi\lambda^2 I_0\,\Delta\nu/16\pi^2 n^2 h\nu)} \tag{5.8-13}$$

Using the definite integral

$$\int_{-\infty}^{\infty} \frac{dx}{x^2 + a^2} = \frac{\pi}{a}$$

to evaluate Equation (5.8-13), we obtain

$$\gamma(\nu) = \frac{\Delta N^0 \lambda^2 p(\nu)}{8\pi n^2 t_{\text{spont}}} \frac{1}{\sqrt{1 + \phi\lambda^2 I_0/4\pi n^2 h\nu\,\Delta\nu}} \tag{5.8-14}$$

$$= \gamma_0(\nu) \frac{1}{\sqrt{1 + I_0/I_s}} \tag{5.8-15}$$

where

$$I_s = \frac{4\pi^2 n^2 h\nu\,\Delta\nu}{\phi\lambda^2} \tag{5.8-16}$$

is the saturation intensity. In Equation (5.8-16), we recall that the parameter ϕ is defined in Section 5.7, and $\Delta\nu$ is the width of the lineshape function $g^{\xi}(\nu)$. A comparison of Equations (5.8-15) and (5.7-10) shows that, because of the square root, the saturation—that is, decrease in gain—sets in more slowly as the intensity I_0 is increased in the case of inhomogeneous broadening.

PROBLEMS

5.1 Use Equations (5.4-8) and (5.4-9) for the following.

(a) Show that the absorption coefficient can be written

$$\alpha(\omega) = \alpha_0 \frac{(\gamma/2)^2}{(\omega - \omega_0)^2 + (\gamma/2)^2} \quad \text{with} \quad \alpha_0 = \frac{Ne^2}{mc\varepsilon_0\gamma}$$

(b) Show that

$$I = \int_{-\infty}^{\infty} \alpha(\omega)\, d\omega = \pi\alpha_0\gamma/4$$

(c) Show that the integrated absorption is independent of γ (even when $\gamma \to 0$).

5.2 Consider a two-level atomic gas with an absorption line center at $\nu_0 = 2.0 \times 10^{15}$ Hz and a Lorentzian lineshape. The atomic gas has a refractive index of 1.0003 (real part) at $\nu = 2.0 \times 10^{14}$ Hz. Assume a Lorentzian linewidth of $\Delta\nu = 1$ GHz.

(a) Find the absorption coefficient in units of m^{-1} at the line center and at $\nu = 2.0 \times 10^{14}$ Hz.

(b) Find the index of refraction (real part) at the line center, and at a frequency detuning of 1 GHz from the line center.

5.3 Show that if $\chi(\nu)$ possesses a pole in the lower half of the complex plane, a step excitation $e(t)$ would lead to a response $p(t)$ that grows exponentially in time. [Hint: Use the Fourier integral relation

$$p(t) = \int_{-\infty}^{\infty} \tilde{p}(\nu) \exp(i2\pi\nu t)\, d\nu$$

where

$$\tilde{p}(\nu) = \varepsilon_0\chi(\nu)\tilde{e}(\nu)$$

5.4 Show that for $p(t)$ to be a real function in Problem 5.3

$$\chi(-\nu) = \chi^*(\nu) \tag{5.P-1}$$

or equivalently,

$$\begin{aligned} \chi'(\nu) &= \chi'(-\nu) \\ \chi''(\nu) &= -\chi''(-\nu) \end{aligned} \tag{5.P-2}$$

5.5

(a) Show that, according to Equation (5.3-7), the poles of χ occur at

$$\omega = i\frac{\gamma}{2} \pm \sqrt{\omega_0^2 - \gamma^2/4}$$

Note that these poles are all in the upper half of the complex plane.

(b) Show that the following integral is zero for $\tau < 0$:

$$F(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} [\varepsilon(\omega) - \varepsilon_0] e^{i\omega\tau}\, d\omega$$

According to Equation (C-9), the displacement field is given by

$$D(t) = \varepsilon_0 E(t) + \int_{-\infty}^{\infty} F(\tau)E(t-\tau)\, d\tau$$

$$= \varepsilon_0 E(t) + \int_{0}^{\infty} F(\tau)E(t-\tau)\, d\tau$$

The displacement field $D(t)$ at time t depends on the value of electric field E at all earlier times.

5.6 Show that, for a Lorenztian lineshape centered at ν_0, the peak gain can be written

$$\gamma(\nu_0) = (N_2 - N_1)\frac{c^2}{8\pi n^2 \nu^2 t_{spont}}\frac{2}{\pi\,\Delta\nu}$$

where $\Delta\nu$ is the full width at half-maximum of the lineshape function.

5.7

(a) Show that in the absence of a driving field, the solution of the classical electron model can be written

$$X(t) = A\exp(-\gamma t/2)\cos(\omega' t + \alpha)$$

where A and α are constants, and ω' is given by

$$\omega'^2 = \omega_0^2 - (\gamma/2)^2$$

(b) Given a damped oscillating dipole $X(t) = A\exp(-\gamma t/2)\cos(\omega_0 t)$, show that the Fourier transform is approximately given by

$$f(\omega) = \frac{A}{4\pi}\frac{i}{(\omega_0 - \omega) + i\gamma/2}$$

5.8 Consider two atoms (or molecules) approaching each other in an elastic collision. The quantum states, strictly speaking, depend on the distance between these two atoms as well as on their orientations, which are functions of time.

(a) If the atoms are radiating with a photon frequency of $\omega = (E_2 - E_1)/\hbar$, show that the net result of an elastic collision is to add a phase to the radiating field:

$$\phi = \int \frac{[\Delta E_2(t) - \Delta E_1(t)]}{\hbar}\,dt$$

If $\phi \gg 1$, one collision is enough to completely dephase the oscillation of the radiating field. For a sequence of collisions, we may treat ϕ as a random number.

(b) Using the model in (a) and examining the Fourier spectrum of the radiating field, show that the broadening of the linewidth is proportional to $1/\tau_{co}$.

REFERENCES

1. See, for example, A. Yariv, *Quantum Electronics*, 3rd ed. Wiley, New York, 1989.
2. Yeh, P., *Optical Waves in Layered Media*. Wiley, New York, 1988.
3. See, for example, R. Kubo, *Statistical Mechanics*. North Holland, Amsterdam, 1964, p. 31.
4. Johnson, L. F., Optically pumped pulsed crystal lasers other than ruby. In *Lasers*, Vol. I, A. K. Levine (ed.). Marcel Dekker, New York, 1966, p. 137.
5. Kittel, C., *Elementary Statistical Physics*. Wiley, New York, 1958, p. 197.
6. Einstein, A., Zur Quantentheorie der Strahlung, *Phys. Z.* **18**:121 (1917).

CHAPTER 6

THEORY OF LASER OSCILLATION AND SOME SPECIFIC LASER SYSTEMS

6.0 INTRODUCTION

In Chapter 5 we found that an atomic medium (e.g., a two-level atomic system) with an inverted population ($N_2 > N_1$) is capable of amplifying an electromagnetic wave in the spectral regime within the transition lineshape (or gain profile). In this chapter we consider the situation in which the inverted atomic medium (also called laser medium) is placed inside an optical resonator. As the electromagnetic wave bounces back and forth between the two reflectors, it passes through the inverted atomic medium and is amplified. If the amplification exceeds the losses caused by transmission (or absorption) loss in the mirrors and scattering loss in the laser medium, the field energy stored in the resonator will increase with time under the appropriate conditions. The increase in the field energy causes the gain coefficient of the gain medium to decrease as a result of gain saturation discussed in the last chapter. The field energy in the resonator will keep increasing until the saturated gain per pass just equals the losses. At this point in time, the net gain per pass is unity and no further increase in the radiation intensity is possible—that is, steady-state oscillation is reached. In this chapter we will derive the threshold population inversion of the atomic medium needed to sustain laser oscillation, beginning with the theory of the Fabry–Perot laser. We will also obtain an expression for the oscillation frequency of the laser oscillator and show how it is affected by the dispersion of the atomic medium. We then consider the problem of optimum output coupling. Multimode laser oscillation and mode locking will be addressed later in the chapter. Several laser systems will be described, including ruby lasers, Nd:YAG lasers, and Er:silica lasers. An important class of lasers, the semiconductor lasers, will be discussed in Chapters 15 and 16.

6.1 FABRY–PEROT LASER

Referring to Figure 6.1, we consider an optical resonator consisting of two mirrors. This is basically a Fabry–Perot etalon. Let the space between the mirrors be filled with an amplifying medium with an inverted atomic population. As a result of the presence of the amplifying medium, the complex index of refraction n' can be written

$$n'^2 = n^2 + \chi' - i\chi'' \tag{6.1-1}$$

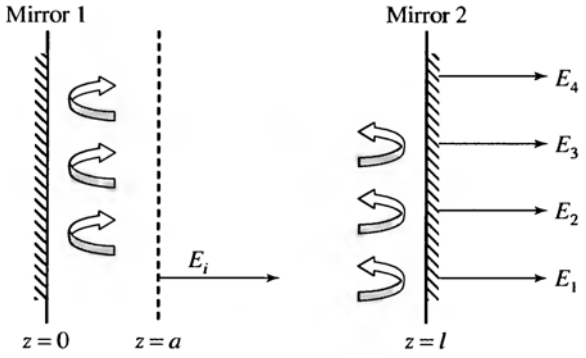


Figure 6.1 Schematic drawing of multiple reflections inside a Fabry–Perot cavity and the corresponding partially transmitted beams. E_1 is the first transmitted beam, E_2 is the transmitted beam after one round-trip inside the cavity, E_3 is the transmitted beam after two round-trips inside the cavity, and so on.

where n is the index of refraction of the host medium, and $\chi' - i\chi''$ is the complex dielectric susceptibility due to the presence of the amplifying materials (the active atoms that are responsible for the gain). In the case of ruby lasers, $n = 1.77$ is the index of refraction of the host material Al_2O_3, and $\chi' - i\chi''$ is due to the chromium atoms (actually Cr^{3+} ions). In most laser systems, $|\chi' - i\chi''| \ll n^2$. The complex propagation constant of plane waves in the gain medium can thus be written

$$k'(\omega) = k_0 n' = k_0\sqrt{n^2 + \chi' - i\chi''} - i\frac{\alpha}{2} \cong k + k\frac{\chi'(\omega)}{2n^2} - ik\frac{\chi''(\omega)}{2n^2} - i\frac{\alpha}{2} \tag{6.1-2}$$

where $k_0 = \omega/c$, $k = n\omega/c$, and α is the linear attenuation coefficient to account for a residual loss due to a variety of mechanisms, such as scattering at imperfections, absorption by excited atomic levels, and others. The attenuation resulting from all of these mechanisms is lumped into the distributed loss constant α. We may view $k - i\alpha/2$ as the propagation constant of the medium at frequencies well removed from that of the laser transition. Since α accounts for the distributed passive losses of the medium, the intensity loss factor per pass is $\exp(-\alpha l)$, where l is the length of the cavity.

Referring again to Figure 6.1, we consider now a white noise with a uniform spectral distribution inside the cavity. The noise can be originating from blackbody radiation or spontaneous emission. Let E_i be the amplitude of the plane wave component (due to the noise) propagating along the positive axis of the cavity at frequency ω.

The electromagnetic wave E_i will undergo multiple reflections inside the resonator, as shown in Figure 6.1. Let us examine the partially transmitted waves at mirror 2 located at $z = l$. We follow the same procedure employed in Chapter 4 to obtain the output wave at mirror 2. The output wave can be written as a sum of all partially transmitted waves,

$$E_{\text{out}} = E_1 + E_2 + E_3 + E_4 + \cdots \tag{6.1-3}$$

where

$$\begin{aligned} E_1 &= t_2 e^{-ik'(l-a)}E_i \\ E_2 &= (r_1 r_2 e^{-i2k'l})t_2 e^{-ik'(l-a)}E_i \\ E_3 &= (r_1 r_2 e^{-i2k'l})^2 t_2 e^{-ik'(l-a)}E_i \\ E_4 &= (r_1 r_2 e^{-i2k'l})^3 t_2 e^{-ik'(l-a)}E_i \end{aligned} \tag{6.1-4}$$

where a is the location of the noise source inside the resonator, t_2 is the amplitude transmission coefficient of the second mirror (at $z = l$), and r_1 and r_2 are the amplitude reflection coefficients (looking from inside the cavity) of the mirrors, respectively. We note that as a

result of the multiple reflections inside the resonator, the sum of the transmitted waves forms a geometric series. Using Equations (6.1-3) and (6.1-4), we obtain

$$\begin{aligned} E_{\text{out}} &= t_2 e^{-ik'(l-a)} E_i \{1 + (r_1 r_2 e^{-i2k'l})^1 + (r_1 r_2 e^{-i2k'l})^2 + (r_1 r_2 e^{-i2k'l})^3 + \cdots\} \\ &= \frac{t_2 e^{-ik'(l-a)}}{1 - r_1 r_2 e^{-i2k'l}} E_i \end{aligned} \tag{6.1-5}$$

where k' is the complex propagation constant given by Equation (6.1-1), and l is the length of the etalon. Since χ'' is related to the gain coefficient, we can rewrite the complex propagation constant as

$$k' = k + \Delta k + i(\gamma - \alpha)/2 \tag{6.1-6}$$

with

$$\Delta k = k \frac{\chi'(\omega)}{2n^2} \tag{6.1-7}$$

$$\gamma = -k \frac{\chi''(\omega)}{n^2} = (N_2 - N_1) \frac{\lambda^2}{8\pi n^2 t_{\text{spont}}} g(\nu) \tag{6.1-8}$$

Physically, Δk is a correction to the propagation constant due to the presence of the active atoms (e.g., Cr atoms in ruby lasers). γ is a gain coefficient due to the active atoms. The output wave can thus be written

$$E_{\text{out}} = \frac{t_2 e^{-i[(k+\Delta k)(l-a)]} e^{(\gamma-\alpha)(l-a)/2}}{1 - r_1 r_2 e^{-i2(k+\Delta k)l} e^{(\gamma-\alpha)l}} E_i \tag{6.1-9}$$

If the noise source is located outside the resonator (say, at $z < 0$), then E_i can be viewed as an incident wave from the mirror at $z = 0$ (mirror 1 in Figure 6.1). In this case, the output wave can be written

$$E_{\text{out}} = \frac{t_1 t_2 e^{-i(k+\Delta k)l} e^{(\gamma-\alpha)l/2}}{1 - r_1 r_2 e^{-i2(k+\Delta k)l} e^{(\gamma-\alpha)l}} (E_i)_{\text{outside}} \tag{6.1-10}$$

where t_1 is the amplitude transmission coefficient of mirror 1 and $(E_i)_{\text{outside}}$ is the amplitude of an incident wave arriving at mirror 1. We notice that the denominators in the above two equations are identical. If the atomic transition is inverted ($N_2 > N_1$), then $\gamma > 0$ and the denominator of Equation (6.1-9) or (6.1-10) can become very small. The output wave E_{out} can become much larger than the source wave E_i. The Fabry–Perot etalon (with the laser medium) in this case acts as an amplifier with a power gain $|E_{\text{out}}/E_i|^2$.

If the denominator of Equation (6.1-9) or (6.1-10) becomes zero, which happens when

$$r_1 r_2 e^{-i2(k+\Delta k)l} e^{(\gamma-\alpha)l} = 1 \tag{6.1-11}$$

then the ratio E_{out}/E_i becomes infinite. This corresponds to a finite output wave E_{out}, with an infinitesimal source wave ($E_i \cong 0$)—that is, to *oscillation*. Physically, condition (6.1-11) represents the case in which a wave making a complete round trip inside the resonator returns to the starting plane with the *same amplitude* and, except for some integral multiple of 2π, with the *same phase*. Separating the oscillation condition (6.1-11) into the amplitude and phase requirements gives

$$r_1 r_2 e^{[\gamma(\omega)-\alpha]l} = 1 \tag{6.1-12}$$

for the threshold gain constant $\gamma_t(\omega)$ and

$$2[k + \Delta k(\omega)]l = 2m\pi, \quad m = 1, 2, 3, \ldots \tag{6.1-13}$$

for the phase condition. Here, without loss of generality, we assume real reflection coefficients r_1 and r_2. The amplitude condition (6.1-12) can be written

$$\gamma_t(\omega) = \alpha - \frac{1}{l} \ln r_1 r_2 \tag{6.1-14}$$

where $\gamma_t(\omega)$ is known as the threshold gain coefficient. The threshold gain condition can be written in terms of the threshold population inversion, according to Equation (6.1-8),

$$N_t \equiv (N_2 - N_1)_t = \frac{8\pi n^2 t_{\text{spont}}}{g(\nu)\lambda^2}\left(\alpha - \frac{1}{l} \ln r_1 r_2\right) \tag{6.1-15}$$

This is the population inversion density at threshold. It was derived originally by Schawlow and Townes in their classic paper on the feasibility of lasers; see Reference [1]. The population inversion density threshold is often written in terms of the photon cavity lifetime, which is defined as follows. Consider the case in which the mirror losses and the distributed losses are all small, and therefore $r_1^2 \approx 1$, $r_2^2 \approx 1$, and $\exp(-\alpha l) \approx 1$. An electromagnetic wave starting with a unit intensity will return after one round trip with an intensity $R_1 R_2 \exp(-2\alpha l)$, where $R_1 \equiv r_1^2$ and $R_2 \equiv r_2^2$ are the mirrors' reflectivities. The fractional intensity loss per round trip is thus $1 - R_1 R_2 \exp(-2\alpha l)$. Since this loss occurs in a round-trip flight time of $2ln/c$. The intensity of the electromagnetic wave will decay exponentially with decay time constant t_c given by (see Problem 6.15)

$$\frac{1}{t_c} = \frac{(1 - R_1 R_2 e^{-2\alpha l})c}{2nl} \tag{6.1-16}$$

This is known as the photon cavity lifetime.

With this definition, the energy U stored in the passive resonator decays as $dU/dt = -U/t_c$. Since $R_1 R_2 e^{-2\alpha l} \cdot e^{2\gamma_t l} = 1$ at oscillation, the photon cavity lifetime can also be written

$$\frac{1}{t_c} = \frac{(1 - e^{-2\gamma_t l})c}{2nl} = \frac{c}{n}\gamma_t = \frac{c}{n}\left(\alpha - \frac{1}{l} \ln r_1 r_2\right) \tag{6.1-17}$$

where we assume a small threshold gain, so that $e^{-2\gamma_t l} \approx 1 - 2\gamma_t l$. Using this expression of the photon cavity lifetime, the threshold condition (6.1-14) becomes

$$N_t \equiv (N_2 - N_1)_t = \frac{8\pi n^3 \nu^2 t_{\text{spont}}}{c^3 t_c g(\nu)} \tag{6.1-18}$$

where $N \equiv N_2 - N_1$, and the subscript t signifies threshold.

EXAMPLE: POPULATION INVERSION

To get an order of magnitude estimate of the critical population inversion $(N_2 - N_1)$, we use data typical of a 6328 Å He–Ne laser. The appropriate constants are

$$\lambda = 6.328 \times 10^{-5} \text{ cm}$$

$$t_{\text{spont}} = 10^{-7} \text{ s}$$

$$l = 12 \text{ cm}$$

$$\frac{1}{g(\nu_0)} \approx \Delta\nu \approx 10^9 \text{ Hz}$$

The last figure is the Doppler-broadened width of the laser transition.

The cavity decay time t_c is calculated from Equation (6.1-17) assuming $\alpha = 0$ and $R_1 = R_2 = 0.98$. Since $R_1 = R_2 \sim 1$, we can use the approximation $\ln r_1 r_2 = \ln R \approx R - 1$ to write

$$t_c \approx \frac{nl}{c(1-R)} = 2 \times 10^{-8} \text{ s}$$

Using the foregoing data in Equation (6.1-18), we obtain

$$N_t \approx 10^9 \text{ cm}^{-3}$$

Figure 6.2 consists of a plot of the output wave intensity, at a given noise source, as a function of the round-trip phase shift. In Figure 6.2a, a uniform gain is assumed. Each curve in Figure 6.2a is for a different value of the distributed gain constant γ. We note that when $\exp[(\gamma - \alpha)l] > 1$—that is, when the net gain per pass exceeds unity—the output can exceed the input source and the Fabry–Perot etalon functions as an amplifier. Figure 6.2b shows the output intensity by assuming a Lorentzian gain profile. We notice that the output intensities of the modes are different. This is a result of the difference in the gain for each of the modes. As a result of the gain profile, only one mode can satisfy the oscillation condition (6.1-11). At oscillation, the gain exactly balances the losses, and a steady-state field energy is built up

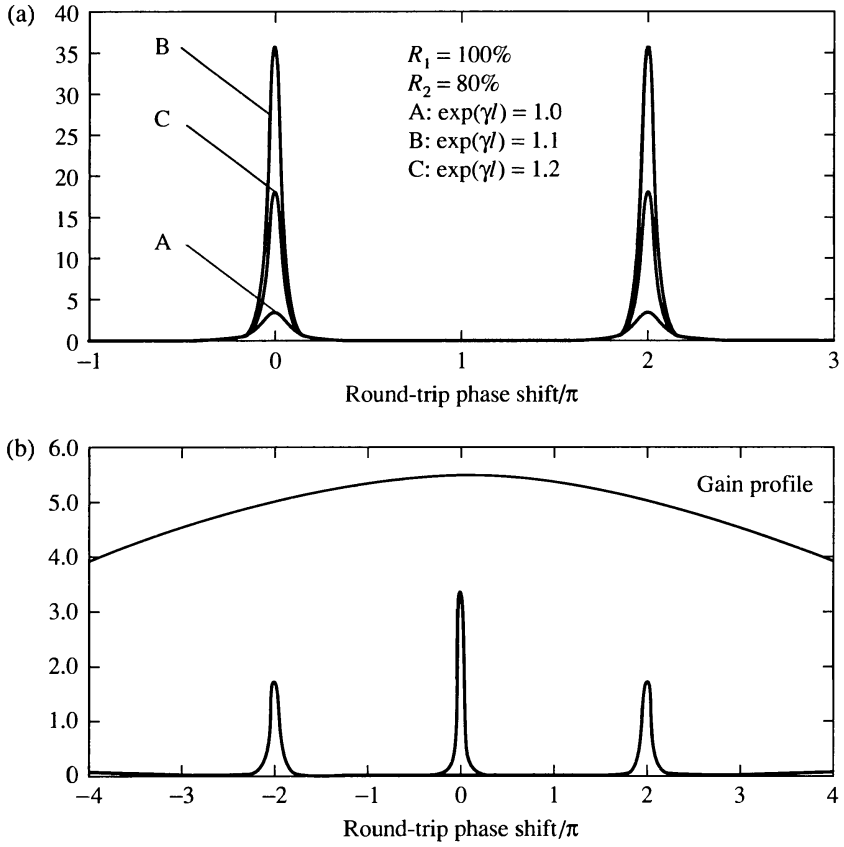


Figure 6.2 Output intensity as a function of the round-trip phase shift ($2kl - 2m\pi$) in a Fabry–Perot resonator: (a) uniform gain and (b) a Lorentzian gain profile centered at a resonance frequency of the Fabry–Perot etalon.

inside the resonator. Thus, in the case of a homogeneously broadened gain medium, only one mode will oscillate.

It is especially interesting to note the narrowing of the peaks as the oscillation condition $\exp(2\gamma l)R_1R_2 = 1$ is approached. At oscillation, the peak transmission becomes infinity, and the width of the transmission peak becomes zero. So the spectral bandwidth of an ideal laser is zero. In practice, the exact laser frequency is determined by the cavity length l, which can change due to external perturbations (e.g., thermal, acoustic). This can lead to a change of the oscillation frequency with time. Furthermore, the presence of spontaneous emission also leads to a broadening of the spectral distribution of the laser output. This problem is explored further in Section 10.6.

6.2 OSCILLATION FREQUENCY

The phase part of the oscillation condition as given by Equation (6.1-13) is satisfied at an infinite set of frequencies, which correspond to the different values of the integer m. If, in addition, the gain condition (6.1-12) is satisfied at one or more of these frequencies, the laser will oscillate at this frequency. To solve for the oscillation frequency, we rewrite Equation (6.1-13) as

$$kl\left(1 + \frac{\chi'(\nu)}{2n^2}\right) = m\pi \tag{6.2-1}$$

We introduce

$$\nu_m = \frac{mc}{2nl} \tag{6.2-2}$$

so that it corresponds to the mth resonance frequency of the passive ($N_2 - N_1 = 0$) resonator. Using Equations (5.4-10a) and (5.4-10b), we can write

$$\chi'(\nu) = \frac{2(\nu_0 - \nu)}{\Delta\nu}\chi''(\nu) \tag{6.2-3}$$

where $\Delta\nu$ is the FWHM bandwidth of the gain profile, and

$$\gamma(\nu) = -\frac{k\chi''(\nu)}{n^2} \tag{6.2-4}$$

We obtain from Equation (6.2-1) the oscillation frequency of the laser,

$$\nu\left[1 - \left(\frac{\nu_0 - \nu}{\Delta\nu}\right)\frac{\gamma(\nu)}{k}\right] = \nu_m \tag{6.2-5}$$

where ν_0 is the center frequency of the atomic lineshape function (gain profile). Let us assume that the laser resonator length l is adjusted so that one of its resonance frequencies ν_m is very near ν_0. We anticipate that the oscillation frequency ν will also be close to ν_m and take advantage of the fact that when $\nu \sim \nu_0$, the gain constant $\gamma(\nu)$ is a slowly varying function of ν; see Figure 5.2 for κ or $\chi''(\nu)$, which is proportional to $\gamma(\nu)$. We can consequently replace $\gamma(\nu)$ in Equation (6.2-5) by $\gamma(\nu_m)$, and $(\nu_0 - \nu)$ by $(\nu_0 - \nu_m)$, obtaining

$$\nu = \nu_m - (\nu_m - \nu_0)\frac{\gamma(\nu_m)c}{2\pi n\,\Delta\nu} \tag{6.2-6}$$

as the solution for the oscillation frequency ν.

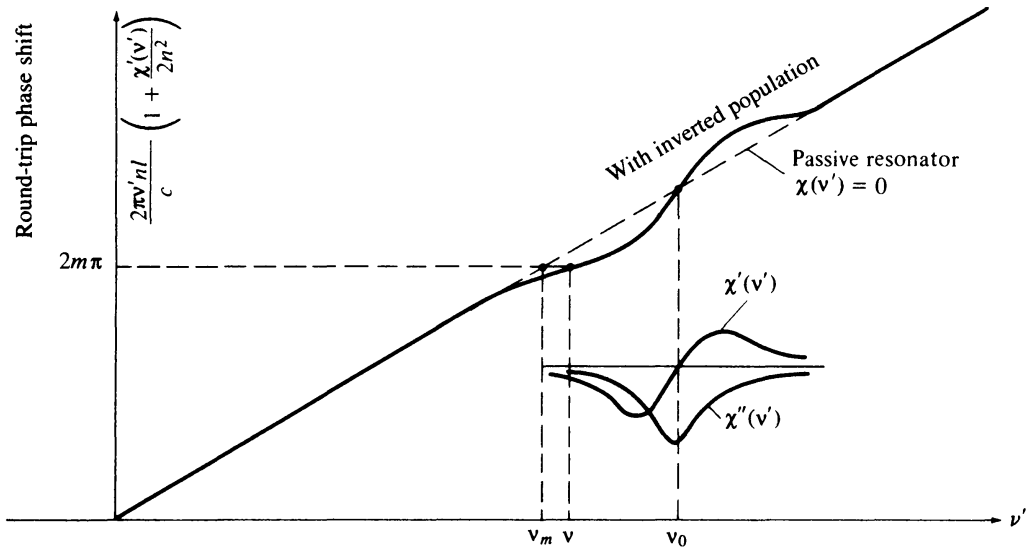


Figure 6.3 Graphical illustration of the laser frequency condition (Equation (6.2-1)) showing how the atomic dispersion $\chi'(\nu)$ "pulls" the laser oscillation frequency, ν, from the passive resonator value, ν_m, toward that of the atomic resonance at ν_0.

We can recast Equation (6.2-6) in a slightly different, and easier to use, form by starting with the gain threshold condition (6.1-12). Taking, for simplicity, $r_1 = r_2 = \sqrt{R}$ and assuming that $R \sim 1$ and $\alpha = 0$, we can write Equation (6.1-14) as

$$\gamma_t(\nu) \approx \frac{1-R}{l}$$

We also take advantage of the relation, according to Equation (4.2-4),

$$\Delta\nu_{1/2} \approx \frac{c(1-R)}{2\pi nl}$$

which relates the passive resonator linewidth $\Delta\nu_{1/2}$ to mirror reflectivity (with $R \sim 1$) and write Equation (6.2-6) as

$$\nu = \nu_m - (\nu_m - \nu_0)\frac{\Delta\nu_{1/2}}{\Delta\nu} \tag{6.2-7}$$

A study of Equation (6.2-7) shows that if the passive cavity resonance ν_m coincides with the atomic line center—that is, $\nu_m = \nu_0$—oscillation takes place at $\nu = \nu_0$. If $\nu_m \neq \nu_0$, oscillation takes place near ν_m but is shifted slightly toward ν_0. This phenomenon is referred to as *frequency pulling* and is demonstrated by Figure 6.3.

The frequency pulling is a result of dispersion in the spectral regime near the line center. Based on the discussion in Chapter 4, we can write the free spectral range (FSR) as

$$(\Delta\nu)_{\text{FSR}} = \frac{c}{2n_g l} \tag{6.2-8}$$

where n_g is the group index given by

$$n_g = n_a + \omega\frac{dn_a}{d\omega} = n_a + \nu\frac{dn_a}{d\nu} \tag{6.2-9}$$

where n_a is the index of refraction including the contribution from the active medium,

$$n_a = n + \tfrac{1}{2}\chi' \tag{6.2-10}$$

In the spectral regime near the line center of an inverted system, $\nu(dn_a/d\nu)$ is a positive number. This is reflected in the increase in the slope near the line center as shown in Figure 6.3. As a result of the positive slope, the group index is larger than the refractive index. Thus the free spectral range is expected to be smaller in the spectral regime near the line center where laser oscillation occurs. The smaller free spectral range is consistent with the pulling of the oscillation frequency toward the line center.

6.3 THREE- AND FOUR-LEVEL LASERS

Lasers are commonly classified into "three-level" or "four-level" lasers. An idealized model of a four-level laser is shown in Figure 6.4. The feature characterizing this laser is that the separation E_1 of the terminal laser level from the ground state is large enough that at the temperature T at which the laser is operated, $E_1 \gg kT$. This guarantees that the thermal equilibrium population of level 1 can be neglected. If, in addition, the lifetime t_1 of atoms in level 1 is short compared to t_2, we can neglect N_1 compared to N_2 and the threshold condition (6.1-18) is satisfied when

$$N_2 \simeq N_t \tag{6.3-1}$$

Therefore laser oscillation begins when the upper laser level acquires, by pumping, a population density equal to the threshold value N_t.

A three-level laser is one in which the lower laser level is either the ground state or a level whose separation E_1 from the ground state is small compared to kT, so that at thermal equilibrium a substantial fraction of the total population occupies this level. An idealized three-level laser system is shown in Figure 6.5.

At a pumping level that is strong enough to create a population $N_2 = N_1 = N_0/2$ in the upper laser level,[1] the optical gain γ is zero, since $\gamma \propto N_2 - N_1 = 0$. To satisfy the oscillation condition, the pumping rate has to be further increased until

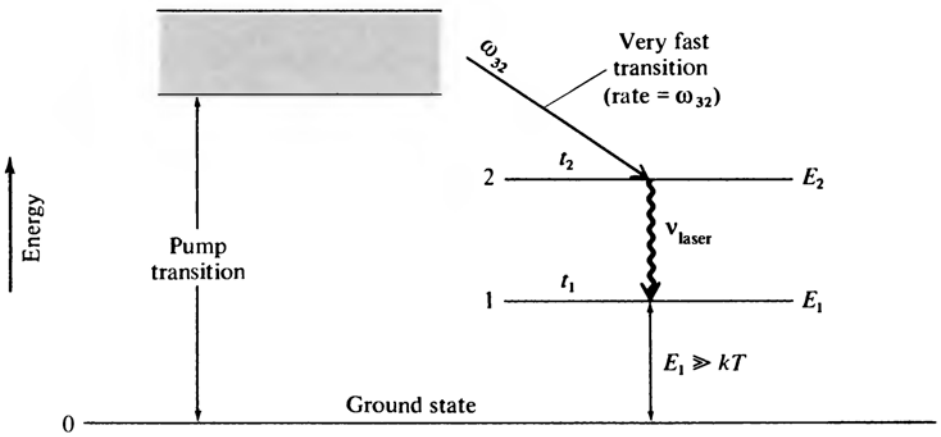


Figure 6.4 Energy-level diagram of an idealized four-level laser.

[1] Here we assume that because of the very fast transition rate ω_{32} out of level 3, the population of this level is negligible and $N_1 + N_2 = N_0$, where N_0 is the density of the active atoms.

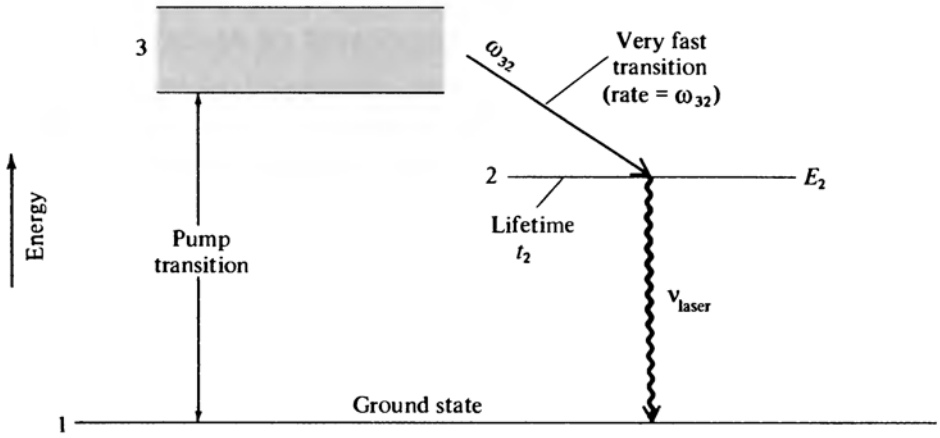


Figure 6.5 Energy-level diagram of an idealized three-level laser.

$$N_2 = \frac{N_0}{2} + \frac{N_t}{2}$$

and

$$N_1 = \frac{N_0}{2} - \frac{N_t}{2} \tag{6.3-2}$$

so $N_2 - N_1 = N_t$. Since in most laser systems $N_0 \gg N_t$, we find by comparing Equation (6.3-1) to (6.3-2) that the pump rate at threshold in a three-level laser must exceed that of a four-level laser—all other factors being equal—by

$$\frac{(N_2)_{3\text{-level}}}{(N_2)_{4\text{-level}}} \sim \frac{N_0}{2N_t}$$

In the example given in the next chapter, we will find that in the case of the ruby laser this factor is ~100.

The need to maintain about $N_0/2$ atoms in the upper level of a three-level laser calls for a *minimum* expenditure of power of

$$(P_s)_{3\text{-level}} = \frac{N_0 h\nu V}{2t_2} \tag{6.3-3}$$

and of

$$(P_s)_{4\text{-level}} = \frac{N_t h\nu V}{t_2} \tag{6.3-4}$$

in a four-level laser. V is the volume. The last two expressions are derived by multiplying the decay rate (atoms per second) from the upper level at threshold, which is $N_0V/2t_2$ and N_tV/t_2 in the two cases, by the energy $h\nu$ per transition. If the decay rate per atom t_2^{-1} (s^{-1}) from the upper level is due to spontaneous emission only, we can replace t_2 by t_{spont}. P_s is then equal to the power emitted through fluorescence by atoms within the (mode) volume V at threshold. We will refer to it as the *critical fluorescence* power. In the case of the four-level laser we use Equation (6.1-18) for N_t and obtain

$$(P_s)_{4\text{-level}} = \frac{N_t h\nu V}{t_2} = \frac{8\pi n^3 h\, \Delta\nu\, V}{\lambda^3 t_c} \frac{t_{spont}}{t_2} \tag{6.3-5}$$

where $\Delta\nu \cong 1/g(\nu_0)$ is the width of the laser transition lineshape.

EXAMPLE: CRITICAL FLUORESCENCE POWER OF AN ND^{3+}:GLASS LASER

The critical fluorescence power of an Nd^{3+}:glass laser is calculated using the following data:

$$l = 10 \text{ cm}$$

$$V = 10 \text{ cm}^3$$

$$\lambda = 1.06 \times 10^{-6} \text{ m}$$

R = (mirror reflectivity) = 0.95

$$n \simeq 1.5$$

$$t_c \simeq \frac{nl}{(1 - R)c} = 10^{-8} \text{ s}$$

$$\Delta\nu = 3 \times 10^{12} \text{ Hz}$$

The Nd^{3+} : glass is a four-level laser system (see Figure 6.49), since level 1 is about 2000 cm^{-1} above the ground state so that at room temperature $E_1 \approx 10kT$. We can thus use Equation (6.3-5), obtaining $N_t = 8.5 \times 10^{15}$ cm^{-3} and

$$P_s \simeq 150 \text{ W}$$

6.4 POWER IN LASER OSCILLATORS

In Section 6.1 we derived an expression for the threshold population inversion N_t at which the laser gain becomes equal to the losses. We would expect that as the pumping intensity is increased beyond the point at which $N_2 - N_1 = N_t$ the laser will break into oscillation and emit power. In this section, we obtain the expression relating the laser power output to the pumping intensity. We also treat the problem of optimum coupling—that is, of the mirror transmission that results in the maximum power output.

Rate Equations

Consider an ideal four-level laser such as the one shown in Figure 6.4. We take $E_1 \gg kT$ so that the thermal population of the lower laser level 1 can be neglected. We assume that the critical inversion density N_t is very small compared to the ground-state population, so during oscillation the latter is hardly affected. We can consequently characterize the pumping intensity by R_2 and R_1, the density of atoms pumped per second into levels 2 and 1, respectively. Process R_1, which populates the lower level 1, causes a reduction of the gain and is thus detrimental to the laser operation. In many laser systems, such as discharge gas lasers, considerable pumping into the lower laser level is unavoidable, and therefore a realistic analysis of such systems must take R_1 into consideration.

The rate equations that describe the populations of levels 1 and 2 become

$$\frac{dN_2}{dt} = -N_2\omega_{21} - W_i(N_2 - N_1) + R_2 \tag{6.4-1}$$

$$\frac{dN_1}{dt} = -N_1\omega_{10} + N_2\omega_{21} + W_i(N_2 - N_1) + R_1 \tag{6.4-2}$$

ω_{ij} is the decay rate per atom from level i to j; thus the density of atoms per second undergoing decay from i to j is $N_i\omega_{ij}$. If the decay rate is due entirely to spontaneous transitions, then ω_{ij} is equal to the Einstein A_{ij} coefficient introduced in Section 5.6. W_i is the probability per unit time that an atom in level 2 will undergo an *induced* (stimulated) transition to level 1 (or vice versa). W_i, given by Equation (5.6-15), is proportional to the energy density of the radiation field inside the cavity.

Implied in the foregoing rate equations is the fact that we are dealing with a homogeneously broadened system. In an inhomogeneously broadened atomic transition, atoms with different transition frequencies $(E_2 - E_1)/h$ experience different induced transition rates and a single parameter W_i is not sufficient to characterize them.

In a steady-state situation, we have $\dot{N}_1 = \dot{N}_2 = 0$. In this case, we can solve Equations (6.4-1) and (6.4-2) for N_1 and N_2, obtaining

$$N_2 - N_1 = \frac{R_2[1 - (\omega_{21}/\omega_{10})(1 + R_1/R_2)]}{W_i + \omega_{21}} \tag{6.4-3}$$

A necessary condition for population inversion in our model is thus $\omega_{21} < \omega_{10}$, which is equivalent to requiring that the lifetime of the upper laser level ω_{21}^{-1} exceed that of the lower one. The effectiveness of the pumping is, according to Equation (6.4-3), reduced by the finite pumping rate R_1 and lifetime ω_{10}^{-1} of level 1 to an effective value

$$R = R_2\left[1 - \frac{\omega_{21}}{\omega_{10}}\left(1 + \frac{R_1}{R_2}\right)\right] \tag{6.4-4}$$

so Equation (6.4-3) can be written

$$N_2 - N_1 = \frac{R}{W_i + \omega_{21}} \tag{6.4-5}$$

Below the oscillation threshold the induced transition rate W_i is zero (since the oscillation energy density is zero) and $N_2 - N_1$ is, according to Equation (6.4-5), proportional to the pumping rate R. This state of affairs continues until $R = N_t\omega_{21}$, at which point $N_2 - N_1$ reaches the threshold value (see Equation (6.1-18))

$$N_t = \frac{8\pi n^3 \nu^2 t_{\text{spont}}}{c^3 t_c g(\nu_0)} = \frac{8\pi n^3 \nu^2 t_{\text{spont}}\,\Delta\nu}{c^3 t_c} \tag{6.4-6}$$

This is the point at which the gain at ν_0 due to the inversion is large enough to make up *exactly* for the cavity losses (the criterion that was used to derive N_t). Further increase of $N_2 - N_1$ with pumping is impossible in a *steady-state situation*, since it would result in a rate of induced (energy) emission that exceeds the losses so that the field energy stored in the resonator will increase with time in violation of the steady-state assumption.

This argument suggests that, under steady-state conditions, $N_2 - N_1$ must remain equal to N_t regardless of the amount by which the threshold pumping rate is exceeded. An examination of Equation (6.4-5) shows that this is possible, provided W_i is allowed to increase once R exceeds its threshold value $\omega_{21}N_t$, so that the equality

$$N_t = \frac{R}{W_i + \omega_{21}} \tag{6.4-7}$$

is satisfied. Since, according to Equation (5.6-15), W_i is proportional to the energy density in the resonator, Equation (6.4-7) relates the electromagnetic energy stored in the resonator to

the pumping rate R. To derive this relationship we first solve Equation (6.4-7) for W_i, obtaining

$$W_i = \frac{R}{N_t} - \omega_{21}, \quad R \geq N_t\omega_{21} \tag{6.4-8}$$

The total power generated by stimulated emission is

$$P_e = (N_tV)W_ih\nu \tag{6.4-9}$$

where V is the volume of the oscillating mode. Using Equation (6.4-8) in (6.4-9) gives

$$\frac{P_e}{Vh\nu} = N_t\omega_{21}\left(\frac{R}{N_t\omega_{21}} - 1\right), \quad R \geq N_t\omega_{21} \tag{6.4-10}$$

This expression may be recast in a slightly different form, which we will find useful later on. We use expression (6.4-6) for N_t and, recalling that in our idealized model $\omega_{21}^{-1} = t_{\text{spont}}$, obtain

$$\frac{P_e}{Vh\nu} = N_t\omega_{21}\left(\frac{R}{p/t_c} - 1\right), \quad R \geq \frac{p}{t_c} \tag{6.4-11}$$

where

$$p = \frac{8\pi n^3\nu^2}{c^3 g(\nu_0)} = \frac{8\pi n^3\nu^2\,\Delta\nu}{c^3} \tag{6.4-12}$$

According to Equation (4.0-12), p corresponds to the density (m^{-3}) of radiation modes whose resonance frequencies fall within the atomic transition linewidth $\Delta\nu$—that is, the density of radiation modes that are capable of interacting with the transition.

Returning to the expression for the power output of a laser oscillator (6.4-11), we find that the term $R/(p/t_c)$ is the factor by which the pumping rate R exceeds its threshold value p/t_c. In addition, in an ideal laser system, $\omega_{21} = t_{\text{spont}}^{-1}$, so we can identify $N_t\omega_{21}h\nu V$ with the power P_s going into spontaneous emission at threshold, which is defined by Equation (6.3-5). We can consequently rewrite Equation (6.4-11) as

$$P_e = P_s\left(\frac{R}{R_t} - 1\right) \tag{6.4-13}$$

The main attraction of Equation (6.4-13) is in the fact that, in addition to providing an extremely simple expression for the power emitted by the laser atoms, it shows that for each unity increment of pumping, measured relative to the threshold value, the power increases by P_s. An experimental plot showing the linear relation predicted by Equation (6.4-13) is shown in Figure 6.6.

In the example of Section 6.3, which was based on an Nd^{3+}:glass laser, we obtained $P_s = 150$ W. We may expect on this basis that the power from this laser for, say, $(R/R_t) \simeq 2$ (i.e., twice above threshold) will be on the order of 150 W.

6.5 OPTIMUM OUTPUT COUPLING IN LASER OSCILLATORS

The total loss encountered by the oscillating laser mode can conveniently be attributed to two different sources: (a) the inevitable residual loss due to absorption and scattering in the laser material and in the mirrors, as well as diffraction losses in the finite diameter reflectors;

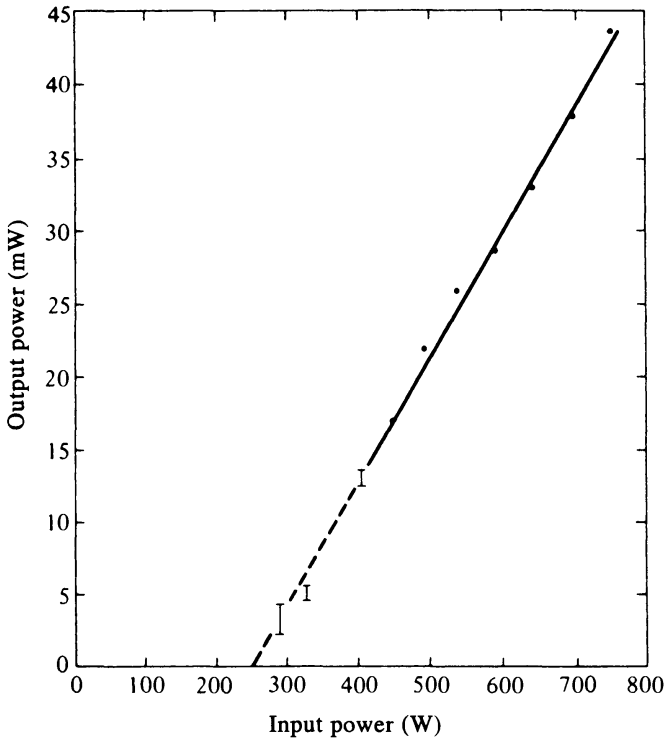


Figure 6.6 Plot of output power versus electric power input to a xenon lamp in a CW $CaF_2 : U^{3+}$ laser. Mirror transmittance at 2.61 μm is 0.2 percent, $T = 77$ K. (After Reference [2].)

(b) the (useful) loss due to coupling of output power through the partially transmissive reflector. It is obvious that loss (a) should be made as small as possible since it raises the oscillation threshold without contributing to the output power. The problem of the coupling loss (b), however, is more subtle. At zero coupling (i.e., both mirrors have zero transmission) the threshold will be at its minimum value and the power P_e emitted by the atoms will be maximum. But since none of this power is available as output, this is not a useful state of affairs. If, on the other hand, we keep increasing the coupling loss, the increasing threshold pumping will at some point exceed the actual pumping level. When this happens, oscillation will cease and the power output will again be zero. Between these two extremes there exists an optimum value of coupling (i.e., mirror transmission) at which the power output is a maximum.

The expression for the population inversion was shown in Equation (6.4-5) to have the form

$$N_2 - N_1 = \frac{R/\omega_{21}}{1 + W_i/\omega_{21}} \tag{6.5-1}$$

Since the exponential gain constant $\gamma(\nu)$ is, according to Equation (5.6-19), proportional to $N_2 - N_1$, we can use Equation (6.5-1) to write it as

$$\gamma = \frac{\gamma_0}{1 + W_i/\omega_{21}} \tag{6.5-2}$$

where γ_0 is the unsaturated ($W_i = 0$) gain constant (i.e., the gain exercised by a very weak field, so that $W_i \ll \omega_{21}$). We can use Equation (6.4-9) to express W_i in (6.5-2) in terms of the total emitted power P_e and then, in the resulting expression, replace $N_t V h\nu\omega_{21}$ by P_s. The result is

$$\gamma = \frac{\gamma_0}{1 + P_e/P_s} \tag{6.5-3}$$

where P_s, the saturation power, is given by Equation (6.3-4). The oscillation condition (6.1-12) can be written

$$e^{\gamma_t l}(1 - L) = 1 \tag{6.5-4}$$

where $L = 1 - r_1 r_2 \exp(-\alpha l)$ is the fraction of the intensity lost per pass. In the case of small losses ($L \ll 1$), Equation (6.5-4) can be written

$$\gamma_t l = L \tag{6.5-5}$$

According to the discussion in the introduction to this chapter, once the oscillation threshold is exceeded, the actual gain γ exercised by the laser oscillation is clamped at the threshold value γ_t regardless of the pumping. We can thus replace γ by γ_t in Equation (6.5-3) and, solving for P_e, obtain

$$P_e = P_s\left(\frac{g_0}{L} - 1\right) \tag{6.5-6}$$

where $g_0 = \gamma_0 l$ (i.e., the unsaturated gain per pass in nepers). P_e, we recall, is the *total* power given off by the atoms due to stimulated emission. The total loss per pass L can be expressed as the sum of the residual (unavoidable) loss L_i and the useful mirror transmission[2] T, so

$$L = L_i + T \tag{6.5-7}$$

The fraction of the total power P_e that is coupled out of the laser as useful output is thus $T/(T + L_i)$. Therefore, using Equation (6.5-6), we can write the (useful) power output as

$$P_o = P_s\left(\frac{g_0}{L_i + T} - 1\right)\frac{T}{T + L_i} \tag{6.5-8}$$

Replacing P_s in Equation (6.5-8) by the right side of Equation (6.3-5), and recalling from Equation (4.7-2) that for small losses

$$t_c = \frac{nl}{(L_i + T)c} = \frac{nl}{Lc} \tag{6.5-9}$$

Equation (6.5-8) becomes

$$P_o = \frac{8\pi n^2 h\nu\, \Delta\nu\, A}{\lambda^2 (t_2/t_{\text{spont}})} T\left(\frac{g_0}{L_i + T} - 1\right) = I_s A T\left(\frac{g_0}{L_i + T} - 1\right) \tag{6.5-10}$$

where $A = V/l$ is the cross-sectional area of the mode (assumed constant) and I_s is the saturation intensity as given in Equation (5.7-9). Maximizing P_0 with respect to T by setting $\partial P_0/\partial T = 0$ yields

$$T_{\text{opt}} = -L_i + \sqrt{g_0 L_i} \tag{6.5-11}$$

as the condition for the mirror transmission that yields the maximum power output.

The expression for the power output at optimum coupling is obtained by substituting Equation (6.5-11) for T in (6.5-10). The result, using Equation (5.7-9), is

$$(P_o)_{\text{opt}} = \frac{8\pi n^2 h\nu\, \Delta\nu\, A}{(t_2/t_{\text{spont}})\lambda^2}(\sqrt{g_0} - \sqrt{L_i})^2 = I_s A(\sqrt{g_0} - \sqrt{L_i})^2 \equiv S(\sqrt{g_0} - \sqrt{L_i})^2 \tag{6.5-12}$$

---

[2] For the sake of simplicity we can imagine one mirror as being perfectly reflecting, whereas the second (output) mirror has a transmittance T.

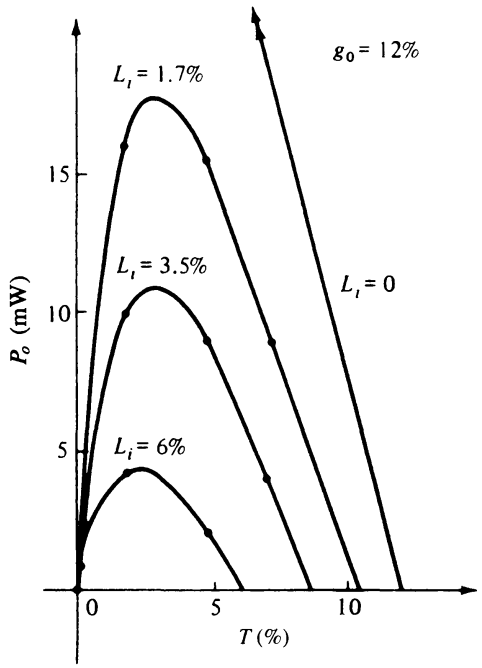


Figure 6.7 Useful power output (P_o) versus mirror transmission T for various values of internal loss L_i in an He–Ne 6328 Å laser. (After Reference [3].)

where the parameter $S = I_s A$ is defined by Equation (6.5-12) and is independent of the excitation level (pumping) or losses.

Theoretical plots of Equation (6.5-10) with L_i as a parameter are shown in Figure 6.7. Also shown are experimental data points obtained in a He–Ne 6328 Å laser. Note that the value of g_0 is given by the intercept of the $L_i = 0$ curve and is equal to 12%. The existence of an optimum coupling resulting in a maximum power output for each L_i is evident.

It is instructive to consider what happens to the energy $\mathscr{E}$ stored in the laser resonator as the coupling T is varied. A little thinking will convince us that $\mathscr{E}$ is proportional to P_o/T.[3] A plot of P_o (taken from Figure 6.7) and $\mathscr{E} \propto P_o/T$ as a function of the coupling T is shown in Figure 6.8. As we may expect, $\mathscr{E}$ is a monotonically decreasing function of T.

6.6 MULTIMODE LASER OSCILLATION AND MODE LOCKING

In this section, we discuss the effect of homogeneous or inhomogeneous broadening of the gain profile on the laser oscillation. As a result of the broadening, several modes can achieve enough gain to reach oscillation. We discuss both single-mode oscillation and multimode oscillation. In addition, we will discuss the locking of phases of multiple longitudinal modes to achieve short laser pulses. We start by reminding ourselves of some basic results pertinent to this discussion:

1. The actual gain constant prevailing inside a laser oscillator *at the oscillation frequency* ν is clamped, at steady state, at a value that is equal to the losses, according to Equation (6.1-14),

$$\gamma_t(\omega) = \alpha - \frac{1}{l} \ln r_1 r_2 \tag{6.6-1}$$

[3] The internal one-way power P_i incident on the mirrors is related, by definition, to P_o by $P_o = P_i T$. The total energy $\mathscr{E}$ is proportional to P_i.

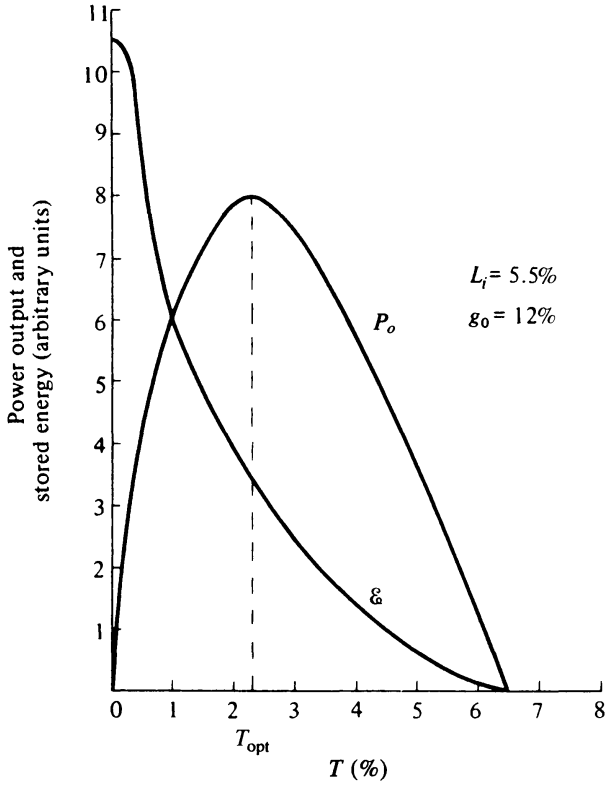


Figure 6.8 Power output P_o and stored energy $\mathscr{E}$ plotted against mirror transmission T.

where l is the length of the gain medium as well as the distance between the mirrors, which are taken here to be the same, α is the linear attenuation coefficient in the resonator (or cavity), and r_1 and r_2 are amplitude reflection coefficients of the mirror.

2. The gain constant of a distributed laser medium is given, according to Equation (5.6-19), by

$$\gamma(\nu) = (N_2 - N_1)\frac{c^2}{8\pi n^2 \nu^2 t_{spont}} g(\nu) \tag{6.6-2}$$

3. The optical resonator can support oscillations, provided sufficient gain is present to overcome losses, at resonance frequencies ν_q of the resonator separated according to Equation (4.6-3) by

$$\nu_{q+1} - \nu_q = \frac{c}{2nl} \tag{6.6-3}$$

where q is an integer, and we ignore the high-order transverse modes and the dispersion. Now consider what happens to the gain constant $\gamma(\nu)$ inside a laser oscillator as the pumping is increased from some value below threshold. Operationally, we can imagine an extremely weak wave of frequency ν launched into the laser medium and then measuring the gain constant $\gamma(\nu)$ as "seen" by this signal as ν is varied.

We treat first the case of a laser with a homogeneous broadening of the gain medium. Below threshold the inversion $(N_2 - N_1)$ is proportional to the pumping rate and $\gamma(\nu)$, which is given by Equation (6.6-2), is proportional to $g(\nu)$. This situation is illustrated by curve A in Figure 6.9a. The resonance frequencies as given by Equation (6.6-3) of the passive resonator are shown in Figure 6.9b. As the pumping rate is increased, the point is reached at which the gain per pass at the center resonance frequency ν_0 is equal to the average loss per

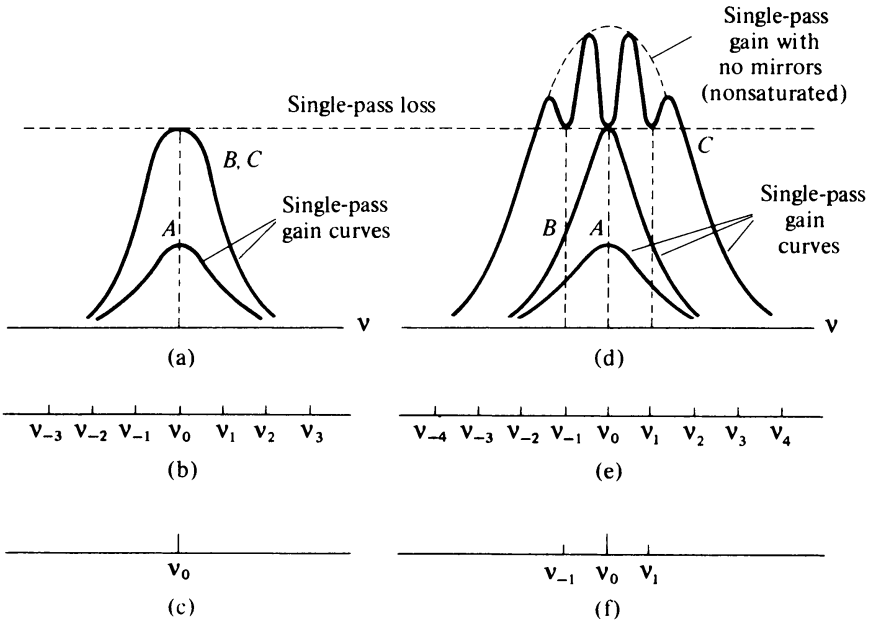


Figure 6.9 (a) Single-pass gain curves for a homogeneous atomic system (*A*—below threshold; *B*—at threshold; *C*—well above threshold). (b) Mode spectrum of optical resonator. (c) Oscillation spectrum (only one mode oscillates). (d) Single-pass gain curves for an inhomogeneous atomic system (*A*—below threshold; *B*—at threshold; *C*—well above threshold). (e) Mode spectrum of optical resonator. (f) Oscillation spectrum for pumping level *C*, showing three oscillating modes.

pass. This is shown in curve *B*. At this point, oscillation at ν_0 starts. An increase in the pumping cannot increase the inversion since this will cause $\gamma(\nu_0)$ to increase beyond its clamped value as given by Equation (6.6-1). Since the spectral lineshape function $g(\nu)$ describes the response of each individual atom, all the atoms being identical, it follows that the gain profile $\gamma(\nu)$ above threshold as in curve *C* is identical to that at threshold curve *B*. The result of further pumping only leads to an increase of the oscillation intensity, while the gain is clamped at the threshold gain value of Equation (6.6-1). Further increase in pumping, and the resulting increase in intracavity optical intensity, will eventually cause an additional broadening of $\gamma(\nu)$ due to the shortening of the lifetime by induced emission. The gain at other frequencies—such as ν_{-1}, ν_1, ν_{-2}, ν_2, and so forth—remains below the threshold value so that the ideal homogeneously broadened laser can oscillate only at a single mode.

In the extreme inhomogeneous case, the individual atoms can be considered as being all different from one another and acting independently. The lineshape function $g(\nu)$ reflects the distribution of the transition frequencies of the individual atoms. The gain profile $\gamma(\nu)$ below threshold is proportional to $g(\nu)$, and its behavior is similar to that of the homogeneous case. Once the threshold is reached as in curve *B*, the gain at ν_0 remains clamped at the threshold value. There is no reason, however, why the gain at other frequencies should not increase with further pumping. This gain is due to atoms that do not communicate with those contributing to the gain at ν_0. Further pumping will thus lead to oscillation at additional longitudinal-mode frequencies as shown in curve *C*. Since the gain at each oscillating frequency is clamped, the gain profile curve acquires depressions at the oscillation frequencies. This phenomenon is referred to as "hole burning".[4] More discussion on hole burning will be given in Section 6.10.

A plot of the output frequency spectrum showing the multimode oscillation of a He–Ne 6328 Å laser is shown in Figure 6.10.

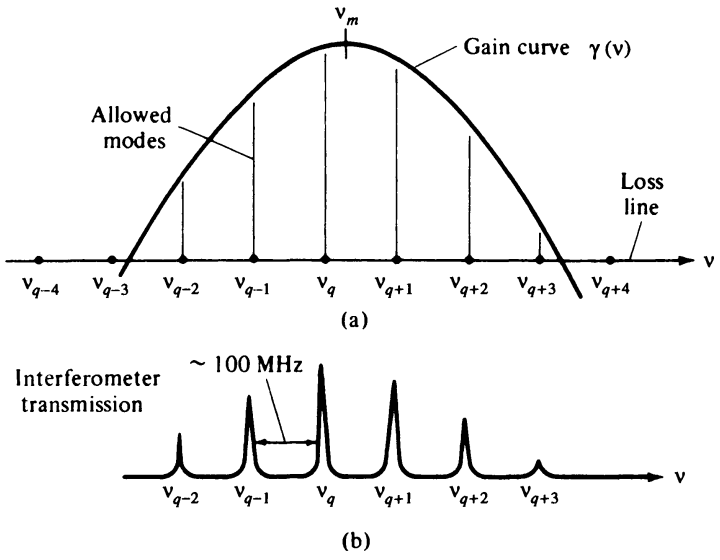


Figure 6.10 (a) Inhomogeneously broadened Doppler gain curve of the 6328 Å Ne transition and position of allowed longitudinal-mode frequencies. (b) Intensity versus frequency profile of an oscillating He–Ne laser. Six modes have sufficient gain to oscillate (After Reference [5].)

Mode Locking

We have argued above that in an inhomogeneously broadened laser, oscillation can take place at a number of frequencies, which are separated by (assuming a refractive index of $n = 1$)

$$\omega_q - \omega_{q-1} = \frac{\pi c}{l} \equiv \Omega \quad (q = \text{integer}) \tag{6.6-4}$$

where Ω is the free spectral range (or the mode spacing) in angular frequency. Now consider the total optical electric field resulting from such multimode oscillation at some arbitrary point, say, next to one of the mirrors, in the optical resonator. It can be taken, using complex notation, as

$$E(t) = \sum_m C_m e^{i[(\omega_0 + m\Omega)t + \phi_m]} \tag{6.6-5}$$

where C_m is the amplitude of the mth mode, and ϕ_m is the phase of the mth mode. The summation is extended over all the oscillating modes and ω_0 is chosen, arbitrarily, as oscillation frequency of one of the modes (usually chosen as the one closest to the line center). One property of Equation (6.6-5) is that $|E(t)|$ is periodic in time with a period of $\tau \equiv 2\pi/\Omega = 2l/c$, which is the round-trip transit time inside the resonator. Using Equation (6.6-5), the field at $t + \tau$ can be written

$$\begin{aligned} E(t+\tau) &= \sum_m C_m \exp\left\{ i\left[(\omega_0 + m\Omega)\left(t + \frac{2\pi}{\Omega} \right) + \phi_m \right] \right\} \\ &= \sum_m C_m \exp\{i[(\omega_0 + m\Omega)t + \phi_m]\} \exp\left\{ i\left[2\pi\left(\frac{\omega_0}{\Omega} + m \right) \right] \right\} \\ &= E(t) \exp(i2\pi\omega_0/\Omega) \end{aligned} \tag{6.6-6}$$

Not that $E(t + \tau)$ is identical to $E(t)$, except for a constant phase factor.

Note that the periodic property of $E(t)$ depends on the fact that the modes are equally spaced and the phases ϕ_m are fixed. In typical lasers the phases ϕ_m are likely to vary randomly with time. This causes the intensity of the laser output to fluctuate randomly and greatly reduces its usefulness for many applications where temporal coherence is important. It should

be noted that this fluctuation takes place because of random interference between modes and not because of intensity fluctuations of individual modes.

There are two ways in which the laser can be made coherent. First, make it possible for the laser to oscillate at a single frequency only so that mode interference is eliminated. This can be achieved in a variety of ways, including shortening the resonator length l, thus increasing the mode spacing ($\Omega = \pi c/l$) to a point where only one mode has sufficient gain to oscillate. The second approach is to force the modes' phases ϕ_m to maintain their relative values (ideally zero, so that they all oscillate in phase). This is the "mode locking" technique proposed and demonstrated in the early history of the laser [6–8]. This mode locking causes the oscillation intensity to consist of a periodic train with a period of $\tau = 2l/c = 2\pi/\Omega$.

One of the most useful forms of mode locking results when the phases ϕ_m are made equal to zero. To simplify the analysis of this case, assume that there are N oscillating modes with equal amplitudes. Taking $C_m = 1/\sqrt{N}$ and $\phi_m = 0$ in Equation (6.6-5), we obtain

$$E(t) = \frac{1}{\sqrt{N}} \sum_{m=1}^{N} e^{i(\omega_0 + m\Omega)t} = \frac{1}{\sqrt{N}} e^{i[(\omega_0 + (N+1)\Omega/2]t} \frac{\sin(N\Omega t/2)}{\sin(\Omega t/2)} \tag{6.6-7}$$

where the field is normalized to have a constant energy (independent of N). The last equality in the above equation is obtained by summing up the geometric series. The average laser power output is proportional to $E(t)E^*(t)$ and is given by

$$P(t) \propto \frac{1}{N} \frac{\sin^2(N\Omega t/2)}{\sin^2(\Omega t/2)} \tag{6.6-8}$$

where the averaging is performed over a time that is long compared with the optical period $2\pi/\omega_0$ but short compared with the modulation period $2\pi/\Omega$.

Some of the analytic properties of $P(t)$ are immediately apparent.

1. The power is emitted in the form of a train of pulses with a period $\tau = 2l/c$, that is, the round-trip transit time.
2. The peak power, $P(s\tau)$ (for $s = 0, 1, 2, 3, \ldots$), is equal to N times the average power, where N is the number of modes locked together.
3. The peak field amplitude is equal to N times the amplitude of a single mode.
4. The pulse width of the main peaks, defined as the time from the peak to the first zero, is $\tau_0 = \tau/N$. This is approximately the FWHM of the main peaks of $P(t)$ (for $N \gg 1$) (see Problem 6.18). There are $(N - 2)$ sidelobes between the neighboring main peaks.

The number of oscillating modes can be estimated by $N \cong \Delta\omega/\Omega$—that is, the ratio of the transition lineshape width $\Delta\omega$ (or gain bandwidth) to the frequency spacing Ω between the modes. Using this relation, as well as $\tau = 2\pi/\Omega$ in $\tau_0 = \tau/N$, we obtain

$$\tau_0 \sim \frac{2\pi}{\Delta\omega} = \frac{1}{\Delta\nu} = \frac{\tau}{N} \tag{6.6-9}$$

where $\Delta\nu$ is the gain bandwidth. Thus the temporal length of the mode-locked laser pulses is approximately the inverse of the gain linewidth.

A theoretical plot of $\sqrt{P(t)}$ as given by Equation (6.6-8) for the case of eight modes ($N = 8$) is shown in Figure 6.11. The ordinate may also be considered as being proportional to the instantaneous field amplitude. The foregoing discussion was limited to consideration of mode locking as a function of time. It is clear, however, that since the solution of Maxwell's equations in the cavity involves traveling waves (a standing wave can be considered as the

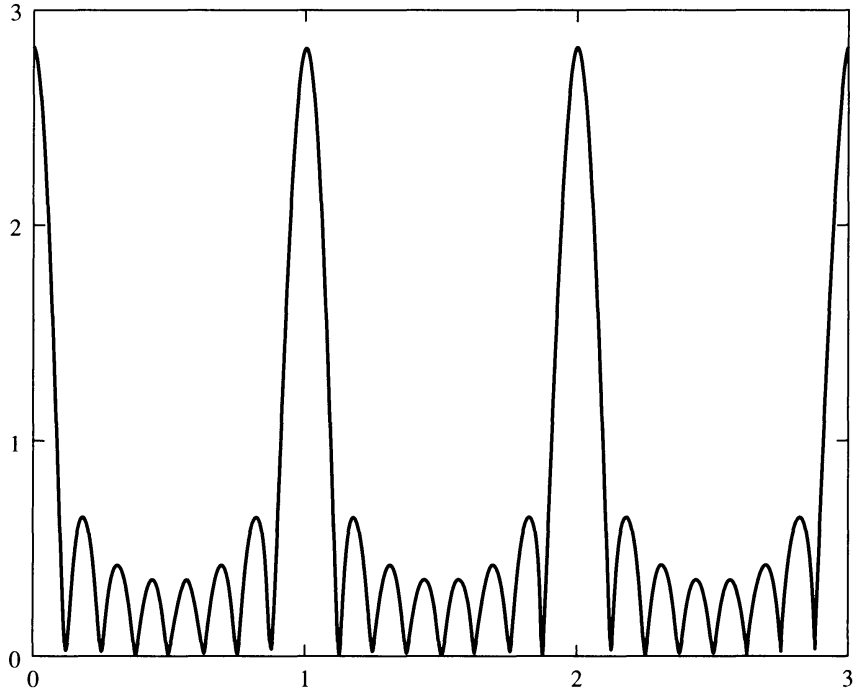


Figure 6.11 Theoretical plot of optical field amplitude $\sqrt{P(t)}$: $|\sin(N\Omega t/2)/\sin(\Omega t/2)|$ resulting from phase locking of eight ($N = 8$) equal-amplitude modes separated from each other by a frequency interval $\Omega = 2\pi/\tau$.

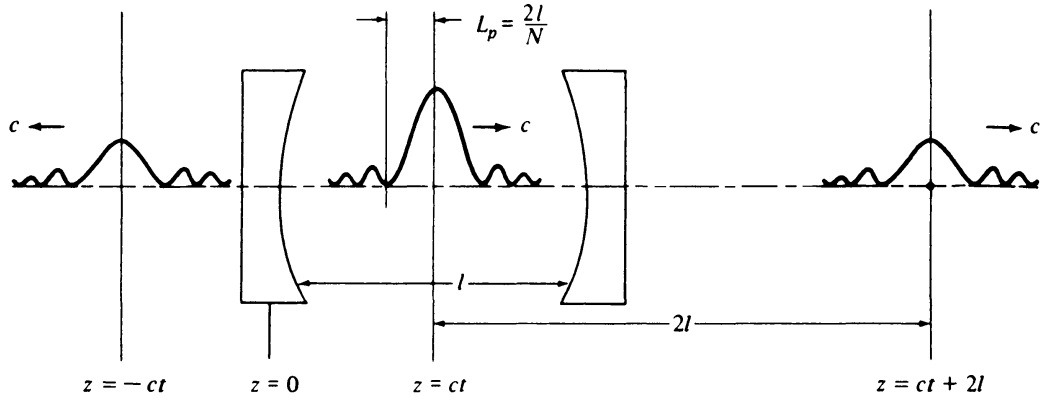


Figure 6.12 Traveling pulse of energy resulting from the mode locking of N laser modes.

sum of two waves traveling in opposite directions), mode locking causes the oscillation energy of the laser to be condensed into a packet that travels back and forth between the mirrors with the velocity of light c (see Figure 6.12). The pulsation period $\tau = 2l/c$ corresponds simply to the time interval between two successive arrivals of the pulse at the mirror. The spatial length of the pulse L_p must correspond to its time duration multiplied by its group velocity c. Using $\tau_0 = \tau/N$, we obtain

$$L_p = c\tau_0 = \frac{c\tau}{N} = \frac{2l}{N} \tag{6.6-10}$$

Figure 6.12 shows a schematic drawing of the spatial profile of the mode-locked pulses at a given instant in time.

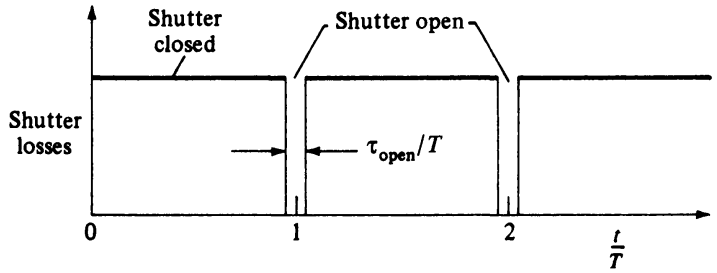


Figure 6.13 Periodic losses introduced by a shutter to induce mode locking. The presence of these losses favors the choice of mode phases that result in a pulse passing through the shutter during open intervals—that is, mode locking.

Methods of Mode Locking

In the preceding discussion we considered the consequences of fixing the phases of the longitudinal modes of a laser mode locking. Mode locking can be achieved by modulating the losses (or gain) of the laser at an angular frequency $\Omega = \pi c/l$, which is equal to the intermode frequency spacing. The theoretical proof of mode locking by loss modulation [2, 6, 7] is rather formal, but a good plausibility argument can be made as follows: As a form of loss modulation consider a thin shutter inserted inside the laser resonator. Let the shutter be closed (high optical loss) most of the time except for brief periodic openings for a duration of τ_{open} every $\tau = 2l/c$ s. This situation is illustrated in Figure 6.13.

A single laser mode will not oscillate in this case because of the high losses (we assume that τ_{open} is too short to allow the oscillation to build up during each opening). The same applies to multimode oscillation with arbitrary phases. There is one exception, however. If the phases were "locked" as in Equation (6.6-7), the energy distribution inside the resonator would correspond to that shown in Figure 6.12 and would consist of a narrow ($L_p = 2l/N$) traveling pulse. If this pulse should arrive at the shutter's position when it is open and if the pulse (temporal) length τ_0 is short compared to the opening time τ_{open}, the mode-locked pulse will be "unaware" of the shutter's existence and, consequently, *will not be attenuated by it.* We may thus reach the conclusion that loss modulation causes mode locking through some kind of "survival of the fittest" mechanism. In reality, the periodic shutter chops off any intensity tails acquired by the mode-locked pulses due to a "wandering" of the phases from their ideal ($\phi_n = 0$) values. This has the effect of continuously restoring the phases.

An experimental setup used to mode lock a He–Ne laser is shown in Figure 6.14; the periodic loss [8] is introduced by Bragg diffraction of a portion of the laser intensity from a standing acoustic wave. The standing-wave nature of the acoustic oscillation causes the strain to have a form $S(z, t) = S_0 \cos(\omega_a t)\cos(k_a z)$, where the acoustic velocity is $v_a = \omega_a/k_a$, and S_0 is a constant amplitude. Since the change in the index of refraction is, to first order, proportional to the strain $S(z, t)$, a phase diffraction grating is created with a spatial period $2\pi/k_a$, which is equal to the acoustic wavelength. The diffraction loss of the incident laser beam due to the grating reaches its peak twice in each acoustic period when $S(z, t)$ has its maximum and minimum values. The loss modulation frequency is thus $2\omega_a$, and mode locking occurs when $2\omega_a = \Omega$, where Ω is the angular frequency separation between two longitudinal laser modes.

Figure 6.15 shows the pulses resulting from mode locking a rhodamine 6G dye laser. Mode locking occurs spontaneously in some lasers if the optical path contains a saturable absorber (an absorber whose opacity decreases with increasing optical intensity). This method is used to induce mode locking in the high-power pulsed solid-state lasers [10, 11]

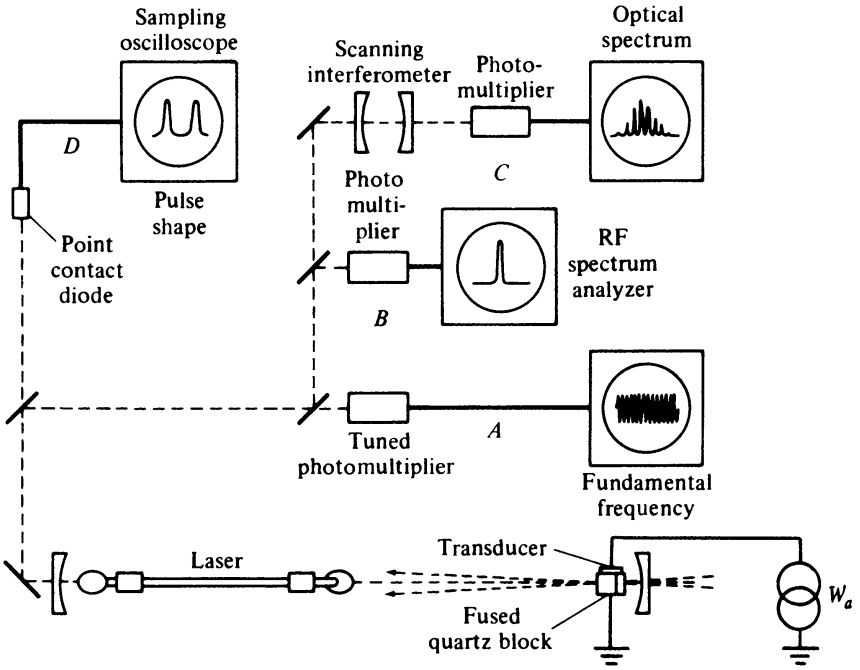


Figure 6.14 Experimental setup for laser mode locking by acoustic (Bragg) loss modulation. The loss is due to Bragg diffraction of the main laser beam by a standing acoustic wave. Parts *A*, *B*, *C*, and *D* of the experimental setup are designed to display the fundamental component of the intensity modulation, the power spectrum of the intensity modulation, the power spectrum of the optical field $E(t)$, and the optical intensity, respectively. (After Reference [9].)

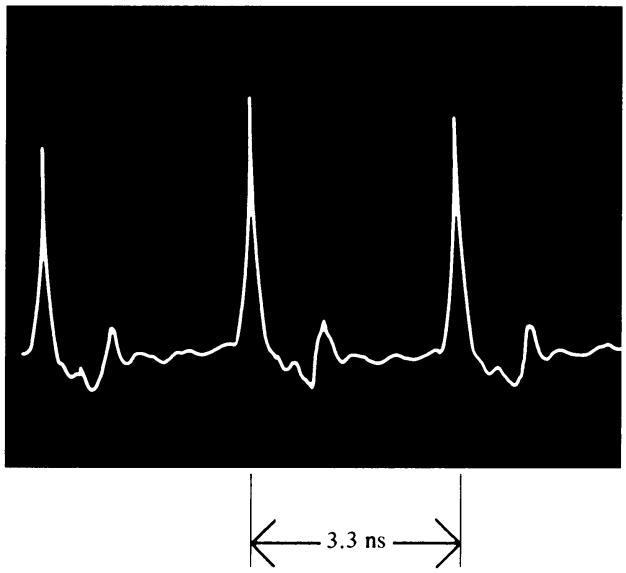


Figure 6.15 Power output as a function of time of a mode-locked dye laser, using rhodamine 6G. The oscillation is at $\lambda = 0.61$ μm. The pulse width is detector limited. (After Reference [9].)

and in continuous dye lasers. This is due to the fact that such a dye will absorb less power from a mode-locked train of pulses than from a random phase oscillation of many modes [2], since the first form of oscillation leads to the highest possible peak intensities, for a given average power from the laser, and is consequently attenuated less severely. From arguments identical to those advanced in connection with the periodic shutter (Figure 6.13), it follows that the presence of a saturable absorber in the laser cavity will "force" the laser, by a "survival of the fittest" mechanism, to lock its modes' phases as in Equation (6.6-7).

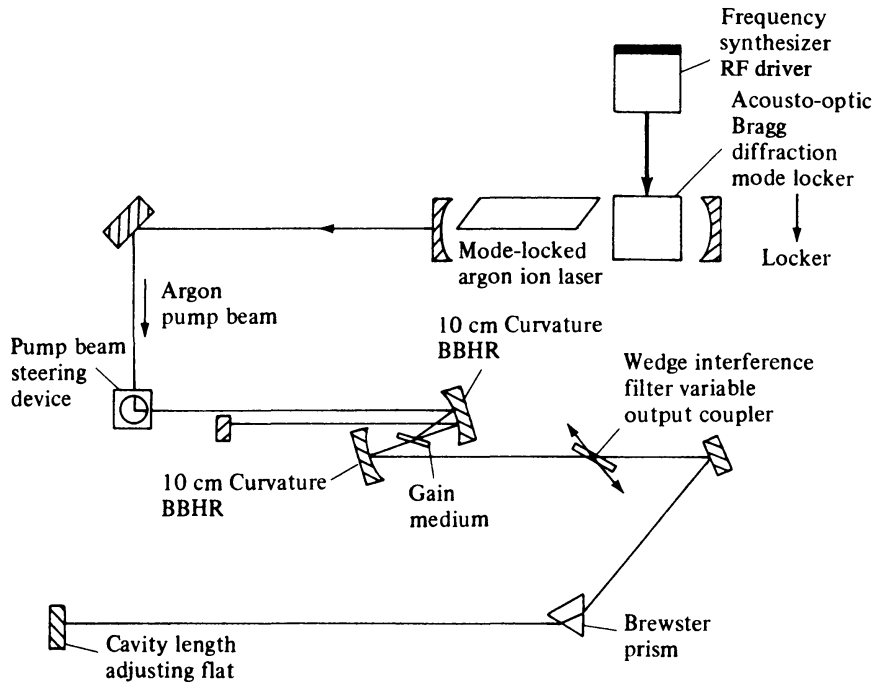


Figure 6.16 Synchronously mode-locked dye laser configuration. (After Reference [19].)

Some of the shortest mode-locked pulses to date were obtained from dye lasers employing rhodamine 6G as the gain medium. The mode locking is caused by synchronous gain modulation that is due to the fact that the pumping (blue-green) argon gas laser is itself mode locked. The pump pulses are synchronized exactly to the pulse repetition rate of the dye laser. (This requires that both lasers have precisely the same optical resonator length.) When this is done, the dye laser gain medium will be pumped once in each round-trip period so that the pumping pulse and the mode-locked pulse overlap spatially and temporally in the dye cell.

Additional sharpening of the mode-locked pulses can result from inclusion of a saturable (dye) absorber in the resonator. A sketch of a synchronously mode-locked dye laser configuration is shown in Figure 6.16. Additional amplification of the output pulses of the dye laser by a sequence of three to four dye laser amplifier cells (consisting of rhodamine 6G pumped by the pulsed second harmonic of Q-switched Nd^{3+}:YAG lasers) has yielded subpicosecond pulses with peak power exceeding 10^9 W. Mode-locked optical pulses with pulse duration of $\sim 30 \times 10^{-15}$ s have been obtained [12]. These pulses have been narrowed down further to $\sim 6 \times 10^{-15}$ s by the use of nonlinear optical techniques that are described in Section 6.8.

Ultrashort mode-locked pulses are now used in an ever-widening circle of applications involving the measurement and study of short-lived molecular and electronic phenomena. The use of ultrashort optical pulses has led to an improvement of the temporal resolution of such experiments by more than three orders of magnitude. For a description of many of these applications as well as of the many methods used to measure the pulse duration, the student should consult References [13–15].

Mode locking in semiconductor lasers is of particular interest owing to the very large gain bandwidth in these media. These lasers offer potential operation in the 10–20 fs range although present results are far from this goal [16, 17]. Of special interest is the possibility of controlling the gain and loss by means of multiple electrodes [18]. Table 6.1 lists some of the lasers commonly used in mode locking, along with the achievable pulse durations.

TABLE 6.1 Some Laser Systems, Their Gain Linewidth $\Delta\nu$, and the Length of Their Pulses in Mode-Locked Operation

Laser Medium	$\Delta\nu$ (Hz)	$(\Delta\nu)^{-1}$ (second)	Observed τ_0 (second)
He–Ne (0.6328 μm)	$\sim 1.5 \times 10^9$	6.66×10^{-10}	$\sim 6 \times 10^{-10}$
Nd:YAG (1.06 μm)	$\sim 1.2 \times 10^{10}$	8.34×10^{-11}	$\sim 7.6 \times 10^{-11}$
Ruby (0.6934 μm)	6×10^{10}	1.66×10^{-11}	$\sim 1.2 \times 10^{-11}$
Nd^{3+}:glass	3×10^{12}	3.33×10^{-13}	$\sim 3 \times 10^{-13}$
Rhodamine 6G dye laser (0.6 μm)	10^{13}	10^{-13}	3×10^{-14}
Diode laser	10^{13}	10^{-13}	4×10^{-13}

Theory of Mode Locking

We have argued earlier that the introduction of a time-periodic loss, or gain, into a laser oscillator will cause phase locking between the otherwise independent, longitudinal modes of the resonator. A necessary condition is that the frequency of this modulation be equal to an integral multiple of the intermode frequency separation (free spectral range) of the resonator. This follows from the gating argument leading to Figure 6.13 or, equivalently, from the requirement that optical sidebands at $\omega_0 \pm \omega_m$, which arise from loss modulation of optical mode ω_0, coincide in frequency with the resonance frequency of the modes ω_q. The theoretical treatment of this problem consists of solving Maxwell's equations for oscillation of an optical resonator with gain (due to inversion) and a periodic loss modulation. The periodic loss modulation leads to a coupling among the modes, and eventually a locking of the phases of the modes. A detailed derivation is presented, for example, in References [7, 20]. The key result is that when the modulation frequency ω_m is equal to the intermode frequency separation, $\Omega = \pi c/l$, we obtain a solution wherein all the modes have the *same* phase. This corresponds to the ultra short pulse form of mode locking discussed at the beginning of this section.

In what follows, we describe an elementary theory of mode locking via a periodic loss modulation. We represent the loss element as a transmission element with a time-varying amplitude modulation. Let the round-trip amplitude transmission coefficient of the loss modulator be written

$$T(t) = \sigma_0 + 2\sigma_1 \cos(\Omega t) = \sigma_0 + 2\sigma_1 \cos\left(\frac{2\pi}{\tau} t\right) \tag{6.6-11}$$

where σ_0 and σ_1 are constants, τ is the round-trip flight time inside the resonator, and $\Omega = 2\pi/\tau$ is the angular frequency of the longitudinal mode spacing. For simplicity in the discussion, we assume that $T(t)$ is a round-trip transmission modulation with real constants σ_0 and σ_1. Let the field amplitude at the input of the loss modulator be written

$$E_{\text{in}}(t) = \sum_{n=1}^{N} C_n e^{in\Omega t} \tag{6.6-12}$$

where C_n (the mode amplitudes) are arbitrary constants, and N is the number of modes involved. For simplicity, we ignore the constant reference frequency ω_0. Physically, N is the total number of modes that are oscillating in the resonator. The field amplitude of the output after the transmission modulation element can be written

$$E_{\text{out}}(t) = T(t)E_{\text{in}}(t) = \sum_{n=1}^{N} C'_n e^{in\Omega t} \tag{6.6-13}$$

where C'_n are constants. Using Equation (6.6-11) for $T(t)$ in Equation (6.6-12), we obtain

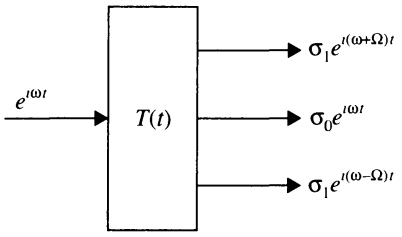


Figure 6.17 A single frequency (ω) input at the modulator leads to an output beam consisting of three frequency components (ω, $\omega + \Omega$, $\omega - \Omega$).

$$E_{\text{out}}(t) = T(t)E_{\text{in}}(t) = \sum_{n=1}^{N} (\sigma_0 C_n + \sigma_1 C_{n+1} + \sigma_1 C_{n-1})\, e^{in\Omega t} \tag{6.6-14}$$

We notice that each mode is coupled with its neighboring modes, when the modulation frequency is equal to the longitudinal mode spacing of the resonator. This coupling mechanism is illustrated in Figure 6.17. For each mode transmitting through the periodic loss modulator, two side bands ($\omega \pm \Omega$) are generated.

As the wave bounces back and forth inside the resonator, all the modes are coupled eventually. Thus the time modulation creates a mechanism that couples all the modes. We will now show that the coupling leads to a constant phase relationship between the modes. This is the origin of mode locking.

If we represent the field amplitude in terms of a column vector of the constants C_n, then the input–output relationship (Equations (6.6-13) and (6.6-14)) can be written

$$\begin{bmatrix} \sigma_0 & \sigma_1 & 0 & \cdots & 0 \\ \sigma_1 & \sigma_0 & \sigma_1 & \cdots & 0 \\ 0 & \sigma_1 & \sigma_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_0 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_N \end{bmatrix} = \begin{bmatrix} C_1' \\ C_2' \\ C_3' \\ \vdots \\ C_N' \end{bmatrix} \tag{6.6-15}$$

We note that all the elements of the matrix are zero except the diagonal elements and those immediately next to the diagonal elements. This is a result of the particular modulation function that creates only one upper and one lower sideband, leading to nearest-neighbor mode coupling. In an inhomogeneously broadened laser system, we may assume an equal gain for the modes that are oscillating. Let g be the round-trip net gain coefficient for the field amplitude; then the round-trip input–output relationship becomes

$$\begin{bmatrix} \sigma_0 & \sigma_1 & 0 & \cdots & 0 \\ \sigma_1 & \sigma_0 & \sigma_1 & \cdots & 0 \\ 0 & \sigma_1 & \sigma_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_0 \end{bmatrix} \begin{bmatrix} g & 0 & 0 & \cdots & 0 \\ 0 & g & 0 & \cdots & 0 \\ 0 & 0 & g & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & g \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_N \end{bmatrix} = \begin{bmatrix} C_1' \\ C_2' \\ C_3' \\ \vdots \\ C_N' \end{bmatrix}$$

or equivalently,

$$\begin{bmatrix} g\sigma_0 & g\sigma_1 & 0 & \cdots & 0 \\ g\sigma_1 & g\sigma_0 & g\sigma_1 & \cdots & 0 \\ 0 & g\sigma_1 & g\sigma_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & g\sigma_0 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_N \end{bmatrix} = \begin{bmatrix} C_1' \\ C_2' \\ C_3' \\ \vdots \\ C_N' \end{bmatrix} \tag{6.6-16}$$

At steady state, the field must reproduce itself after one round-trip. In other words, the output field is proportional to the input field. We can write

$$E_{\text{out}}(t) = \eta E_{\text{in}}(t) \tag{6.6-17}$$

where η is a constant. Physically, η is a round-trip amplitude transmission coefficient of the field inside the resonator. Thus the matrix equation can now be written

$$\begin{bmatrix} g\sigma_0 & g\sigma_1 & 0 & \cdots & 0 \\ g\sigma_1 & g\sigma_0 & g\sigma_1 & \cdots & 0 \\ 0 & g\sigma_1 & g\sigma_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & g\sigma_0 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_N \end{bmatrix} = \eta \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_N \end{bmatrix} \tag{6.6-18}$$

We note that the steady-state oscillation is an eigenvector of the matrix equation and η is an eigenvalue of the matrix equation. The eigenvectors represent a superposition of the N modes in the resonator and can be viewed as the "supermodes." This matrix problem can be solved by rewriting the equation as the following recurrence equation,

$$g\sigma_0 C_n + g\sigma_1 C_{n+1} + g\sigma_1 C_{n-1} = \eta C_n \tag{6.6-19}$$

or equivalently,

$$g\sigma_1 C_{n+1} + (g\sigma_0 - \eta) C_n + g\sigma_1 C_{n-1} = 0 \tag{6.6-20}$$

subject to the boundary condition

$$C_0 = C_{N+1} = 0 \tag{6.6-21}$$

The solution for the mode amplitudes subject to the preceding boundary condition can be written

$$C_n = \sqrt{\frac{2}{N+1}} \sin\left(\frac{ns\pi}{N+1}\right) \tag{6.6-22}$$

where s is an integer ($s = 1, 2, 3, \ldots, N$) and the square root $\sqrt{2/(N+1)}$ is a normalization factor. The eigenvalues are given by

$$\eta = g\sigma_0 + 2g\sigma_1 \cos\left(\frac{s\pi}{N+1}\right), \quad s = 1, 2, 3, \ldots, N \tag{6.6-23}$$

It is interesting to note that the eigenvectors as given by Equation (6.6-22) are independent of the absolute magnitude of (g, σ_0, σ_1). Physically, this means mode locking can be achieved with an arbitrary depth of amplitude modulation (σ_1). With η being the effective round-trip amplitude transmission coefficient, we note that the fundamental solution (with $s = 1$) has the maximum effective transmission (assuming positive σ_0 and σ_1). For this solution, the mode amplitudes are given by

$$C_n = \sqrt{\frac{2}{N+1}} \sin\left(\frac{n\pi}{N+1}\right), \quad n = 1, 2, 3, \ldots, N \tag{6.6-24}$$

Not that the amplitudes are all real. In other words, all phases of the mode amplitudes are locked at zero. We can also say that all modes are phase locked.

We can now construct the temporal profile of the steady-state field inside the resonator for this solution. The field can be written, according to Equation (6.6-24),

$$E(t) = \sum_{n=1}^{N} C_n e^{in\Omega t} = \sqrt{\frac{2}{N+1}} \sum_{n=1}^{N} \sin\left(\frac{n\pi}{N+1}\right) e^{in\Omega t} \tag{6.6-25}$$

After a few steps of algebra, we obtain

$$E(t) = \frac{1}{2}\sqrt{\frac{2}{N+1}} \left(\frac{\sin\frac{N}{2}(\Omega t + \alpha)}{\sin\frac{1}{2}(\Omega t + \alpha)} + \frac{\sin\frac{N}{2}(\Omega t - \alpha)}{\sin\frac{1}{2}(\Omega t - \alpha)} \right) e^{i[(N+1)/2]\Omega t} \tag{6.6-26}$$

where α is given by

$$\alpha = \frac{\pi}{N+1} \tag{6.6-27}$$

Equation (6.6-26) is slightly different from the previous expression, Equation (6.6-7), where constant mode amplitudes were assumed. The field consists of two pulses, each with a width of τ/N, centered at $\Omega t = \alpha, -\alpha$, respectively. As a result, the mode-locked pulse is broader and the sidelobes are lower. The field (6.6-26) is periodic with a period of τ. Peak intensity occurs at $t = 0, \tau, 2\tau, 3\tau, 4\tau, \ldots$. If we define the pulse width of the main peaks as the time from the peak to the first zero, it can easily be shown that, for large N, the width of the mode-locked pulse (6.6-26) is

$$\tau_0 = \frac{3}{2}\frac{\tau}{N} \tag{6.6-28}$$

where we recall that τ is the round-trip flight time in the cavity. This is approximately the full width at half maximum of the pulse (see Problem 6.18). We note that the pulse width is approximately 50% broader due to the tapering of the mode amplitudes as given by Equation (6.6-24). Figure 6.18 is a plot of the amplitude profile of the steady-state field inside the resonator.

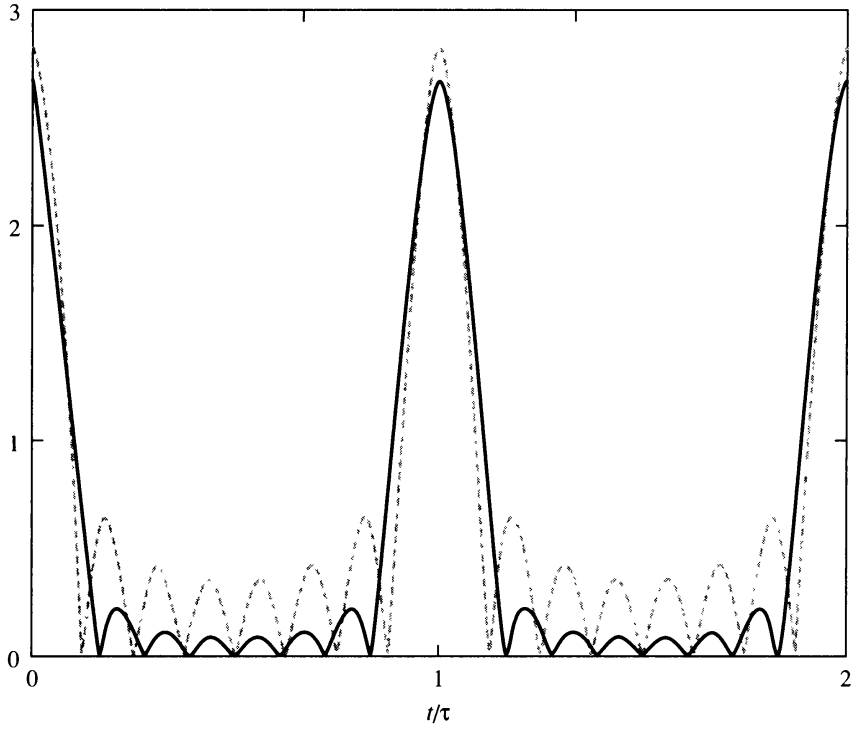


Figure 6.18 Normalized field amplitude (absolute value) profile of mode-locked laser beam with $N = 8$. The solid curve is from Equation (6.6-25), and the dotted curve is from Equation (6.6-7). The field amplitudes are normalized so that $\sum |C_n|^2 = 1$.

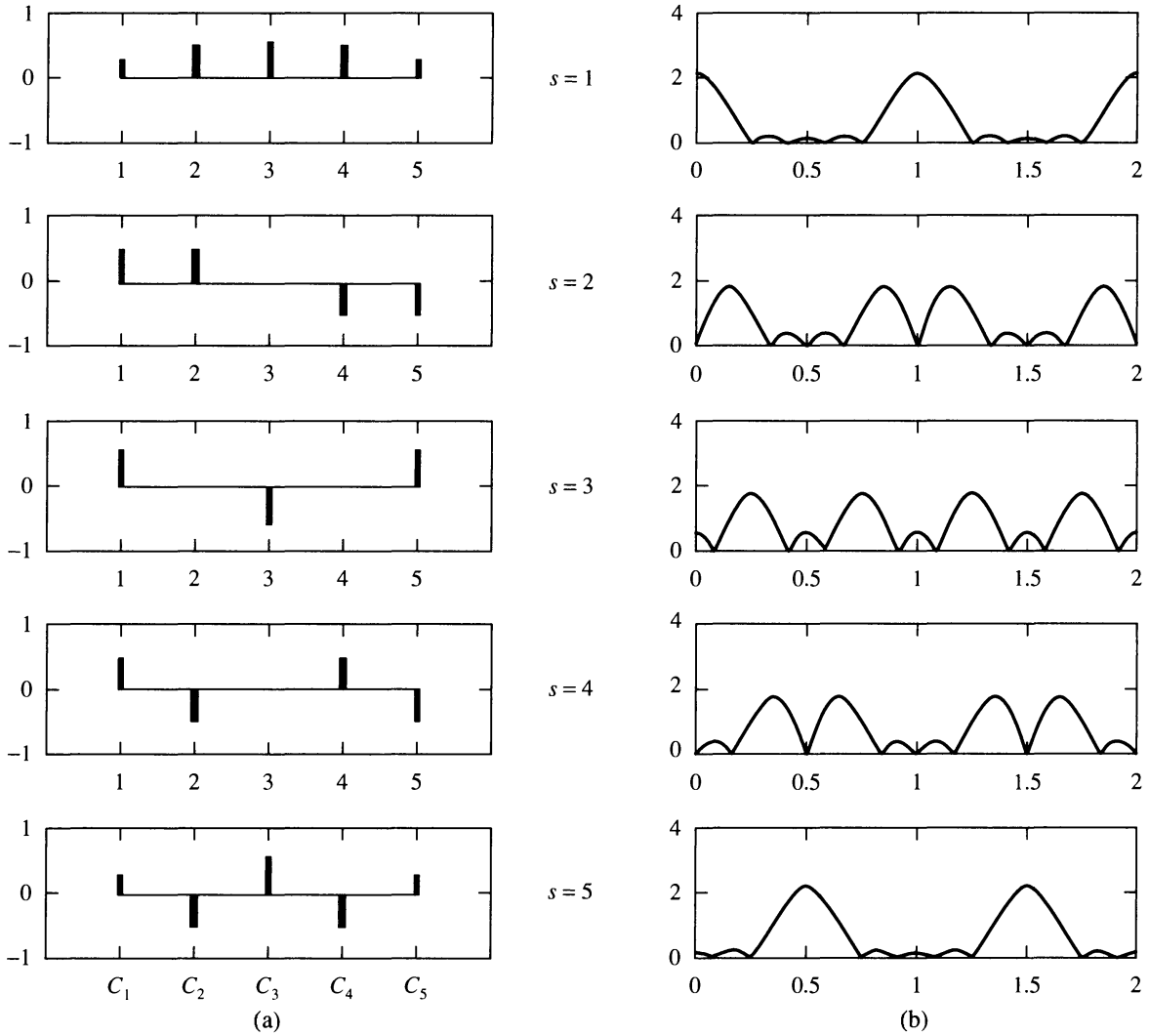


Figure 6.19 Normalized field amplitude (absolute value) profile of (a) the mode coefficients C_n with $N = 5$ and (b) the supermodes. The field amplitudes are normalized so that $\sum|C_n|^2 = 1$. The top row is for the supermode with $s = 1$, and the bottom row is for the supermode with $s = 5$.

Strictly speaking, only the supermode with $\eta = 1$ can sustain oscillation. In general, there is only one supermode whose round-trip amplitude transmission coefficient equals one. Depending on the pumping level and the gain coefficient, any one of the supermodes can oscillate. Figure 6.19 shows the normalized field amplitude (absolute value) profile of the supermodes with $N = 5$. The field amplitude of the supermodes can be written

$$E(t) = \frac{1}{2}\sqrt{\frac{2}{N+1}}\left(\frac{\sin\frac{N}{2}(\Omega t + s\alpha)}{\sin\frac{1}{2}(\Omega t + s\alpha)} + \frac{\sin\frac{N}{2}(\Omega t - s\alpha)}{\sin\frac{1}{2}(\Omega t - s\alpha)}\right)e^{i[(N+1)/2]\Omega t} \tag{6.6-29}$$

where $s = 1, 2, 3, \ldots, N$.

We note that the main peaks of the supermode with $s = 1$ coincide with the maximum transmission windows of the amplitude modulator. This is consistent with the maximum averaged transmission. The main peaks of the supermode with $s = 5$ coincide with the minimum transmission windows of the amplitude modulator. Figure 6.20 illustrates the absolute value of the field amplitude for the supermodes with $s = 1$ and $s = 5$. Note the temporal locations

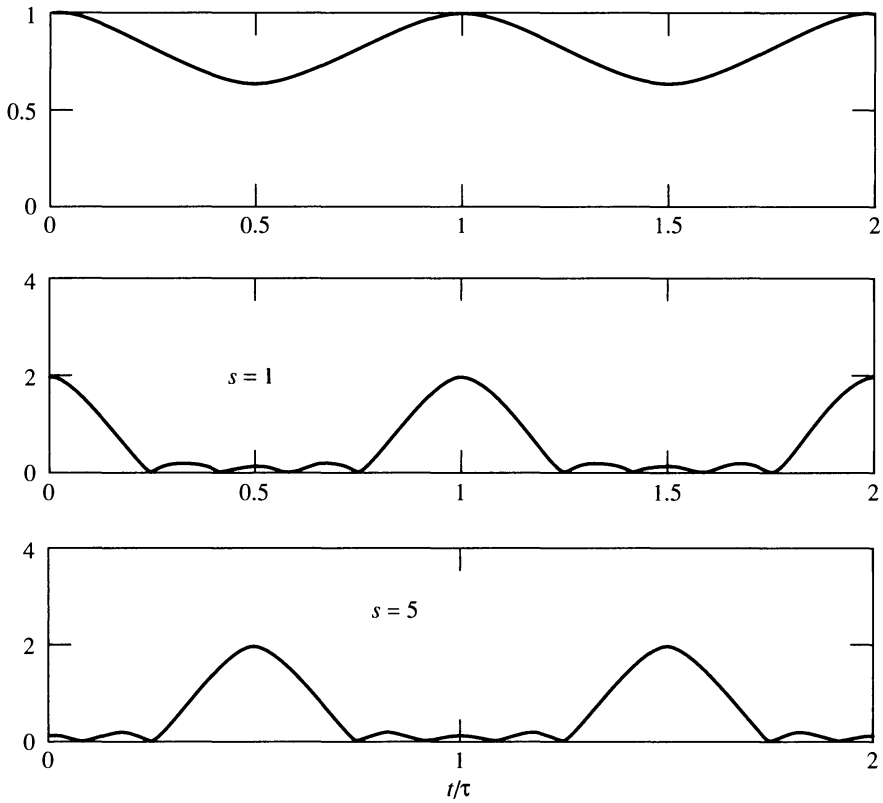


Figure 6.20 Theoretical plots of the absolute values of the field amplitude for the supermodes with $s = 1$ and $s = 5$. The top plot is the amplitude transmission as a function of time (in units of τ). The lower two plots are the absolute values of the field of supermodes $s = 1$ and $s = 5$, respectively.

of the main peaks of these modes relative to the maximum transmission windows of the amplitude modulator.

6.7 MODE LOCKING IN HOMOGENEOUSLY BROADENED LASER SYSTEMS

The analysis of mode locking in inhomogeneous laser systems in Section 6.6 assumed that the role of internal modulation was that of locking together the phases of modes that, in the absence of modulation, oscillate with random phases. In the case of homogeneous broadening, only one mode can normally oscillate. Experiments, however, reveal that mode locking leads to short pulses in a manner quite similar to that described in Section 6.6 and therefore must involve multimode oscillation. One way to reconcile the two points of view and the experiments is to realize that *in the presence of internal modulation*, power is transferred continuously from the high-gain mode to those of lower gain (i.e., those that would not normally oscillate). This power can be viewed simply as that of the sidebands at $(\omega_0 \pm n\Omega)$ of the mode at ω_0 created by a modulation at frequency Ω. Armed with this understanding, we see that the physical phenomenon is not only one of mode locking but also one of mode generation. The net result, however, is that of a large number of oscillating modes with equal frequency spacing and fixed phases, as in the inhomogeneous case, leading to ultrashort pulses.

The analytical solution to this case [21–23] follows an approach used originally to analyze short pulses in traveling wave microwave oscillators [24]. Referring to Figure 6.21, we consider an optical resonator with mirror reflectivity R_1 and R_2 that contains, in addition to the gain medium, a periodically modulated loss cell. The method of solution is to follow

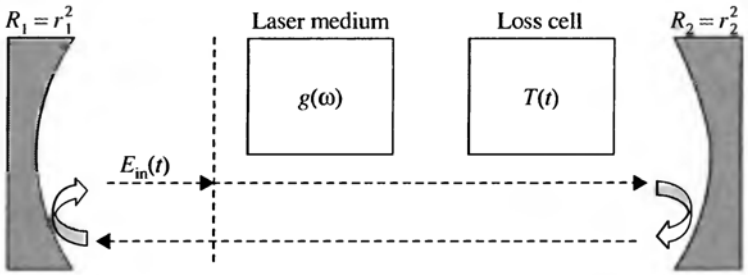


Figure 6.21 The experimental arrangement assumed in the theoretical analysis of mode locking in homogeneously broadened lasers. An input field $E_{in}(t)$ at an arbitrary reference point will experience an amplitude gain of $g(\omega)$ and an amplitude loss modulation of $T(t)$ after one round-trip and becomes $E_{out}(t) = g(\omega)T(t)\,E_{in}(t)$.

one pulse through a complete round trip through the resonator and to require that the pulse reproduce itself. For simplicity in the discussion, we assume a round-trip gain profile of $g(\omega)$ for the field amplitude and a round-trip amplitude transmission modulation function of $T(t)$ for the loss cell.

The elementary theory of mode locking described in the previous section must be modified to account for the different gain in a homogeneously broadened gain profile. Here, again, we assume a round-trip amplitude transmission function of

$$T(t) = \sigma_0 + 2\sigma_1 \cos(\Omega t) = \sigma_0 + 2\sigma_1 \cos\left(\frac{2\pi}{\tau}t\right) \tag{6.7-1}$$

for the loss cell. Where we recall that σ_0 and σ_1 are constants, τ is the round-trip flight time inside the resonator, and Ω is the mode spacing in angular frequency. For simplicity, we assume that σ_0 and σ_1 are real constants. The constant σ_1 is the so-called modulation depth. Let the round-trip amplitude gain be written

$$g_n = g(\omega_0 + n\Omega) \tag{6.7-2}$$

for the nth mode, where ω_0 is a constant reference frequency. Let the field amplitude at an arbitrary reference point inside the resonator be written

$$E_{in}(t) = \sum_{n=1}^{N} C_n e^{in\Omega t} \tag{6.7-3}$$

where C_n is the amplitude of the nth mode and N is the total number of modes involved. Here again, we ignore the constant reference frequency ω_0. The field amplitude after one round-trip inside the resonator can be written

$$E_{out}(t) = g(\omega)T(t)E_{in}(t) = \sum_{n=1}^{N} C_n' e^{in\Omega t} \tag{6.7-4}$$

where we have lumped the mirror reflectivity into the net gain $g(\omega)$, and C_n are constants.

Using Equations (6.7-1) and (6.7-2), we obtain

$$E_{out}(t) = g(\omega)T(t)E_{in}(t) = \sum_{n=1}^{N} (g_n\sigma_0 C_n + \sigma_1 g_{n+1}C_{n+1} + \sigma_1 g_{n-1}C_{n-1})\, e^{in\Omega t} \tag{6.7-5}$$

We notice that, again, each mode is coupled with its neighboring modes, when the modulation frequency is equal to the mode spacing of the resonator. As the wave bounces back and forth inside the resonator, all the modes are coupled eventually. Thus the time modulation creates a mechanism that couples all the modes. As a result of the coupling, energy exchange

occurs among the modes. We will now show that the coupling leads to a constant phase relationship between the modes. This is the origin of mode locking.

If we represent the field amplitude in terms of a column vector of the constants C_n, then the input–output relationship can be written

$$\begin{bmatrix} \sigma_0 & \sigma_1 & 0 & \cdots & 0 \\ \sigma_1 & \sigma_0 & \sigma_1 & \cdots & 0 \\ 0 & \sigma_1 & \sigma_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_0 \end{bmatrix} \begin{bmatrix} g_1 & 0 & 0 & \cdots & 0 \\ 0 & g_2 & 0 & \cdots & 0 \\ 0 & 0 & g_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & g_N \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_N \end{bmatrix} = \begin{bmatrix} C_1' \\ C_2' \\ C_3' \\ \vdots \\ C_N' \end{bmatrix} \tag{6.7-6}$$

Here the second diagonal matrix accounts for the different gains for each of the modes in a homogeneously broadened medium. Multiplying the two matrices, we obtain

$$\begin{bmatrix} g_1\sigma_0 & g_2\sigma_1 & 0 & \cdots & 0 \\ g_1\sigma_1 & g_2\sigma_0 & g_3\sigma_1 & \cdots & 0 \\ 0 & g_2\sigma_1 & g_3\sigma_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & g_N\sigma_0 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_N \end{bmatrix} = \begin{bmatrix} C_1' \\ C_2' \\ C_3' \\ \vdots \\ C_N' \end{bmatrix} \tag{6.7-7}$$

At steady state, the field must reproduce itself after one round-trip. Thus the output field is proportional to the input field. We can write

$$E_{\text{out}}(t) = \eta E_{\text{in}}(t) \tag{6.7-8}$$

where η is a constant. Physically, η is the net round-trip amplitude transmission coefficient of the combination of the gain medium, mirrors, and the loss cell in the resonator. Thus the matrix equation can now be written

$$\begin{bmatrix} g_1\sigma_0 & g_2\sigma_1 & 0 & \cdots & 0 \\ g_1\sigma_1 & g_2\sigma_0 & g_3\sigma_1 & \cdots & 0 \\ 0 & g_2\sigma_1 & g_3\sigma_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & g_N\sigma_0 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_N \end{bmatrix} = \eta \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_N \end{bmatrix} \tag{6.7-9}$$

We note that the steady-state oscillation is an eigenvector of the matrix equation and η is an eigenvalue of the matrix equation. The eigenvectors represent a superposition of the N modes in the resonator and can be viewed as the supermodes. This matrix problem is equivalent to the following recurrence equation

$$g_n\sigma_0 C_n + g_{n+1}\sigma_1 C_{n+1} + g_{n-1}\sigma_1 C_{n-1} = \eta C_n \tag{6.7-10}$$

or equivalently,

$$g_{n+1}\sigma_1 C_{n+1} + (g_n\sigma_0 - \eta)C_n + g_{n-1}\sigma_1 C_{n-1} = 0 \tag{6.7-11}$$

subject to the boundary condition

$$C_0 = C_{N+1} = 0 \tag{6.7-12}$$

Closed-form solutions of Equation (6.7-10) subject to the preceding boundary condition are, in general, not easy to obtain. Since the matrix in Equation (6.7-9) is a product of a real symmetric matrix and a real diagonal matrix, all eigenvalues and eigenvectors are real. It can

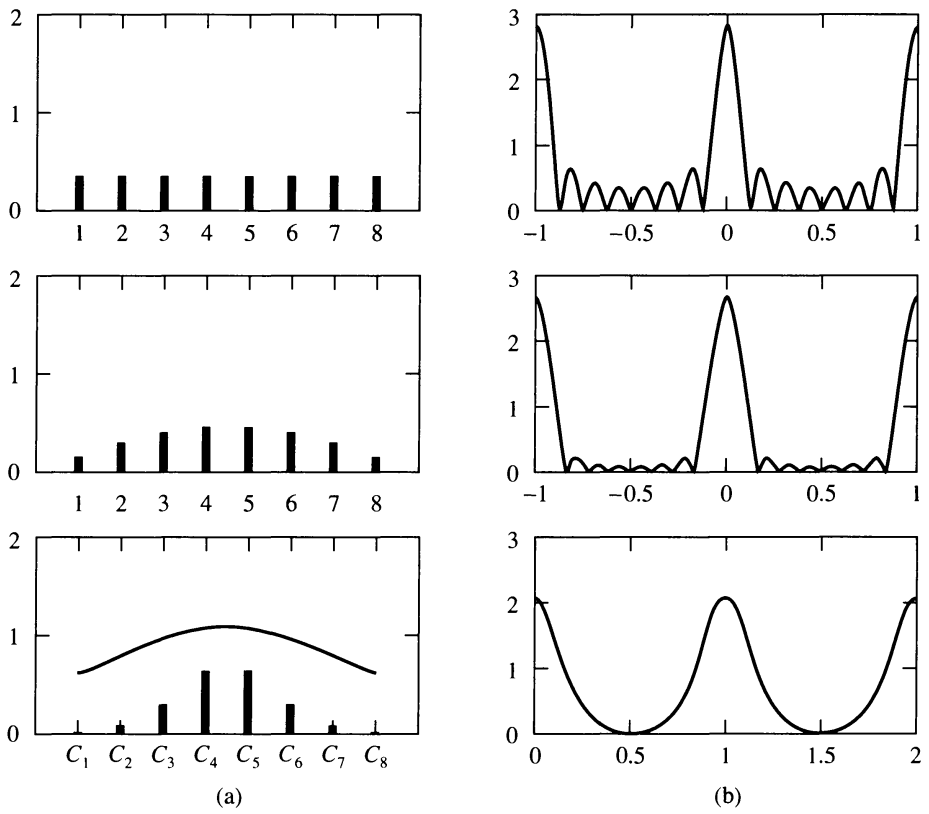


Figure 6.22 (a) Mode amplitudes and (b) absolute values of the field of the fundamental supermode of mode-locked laser pulses. The top row if for the case when all modes have the same amplitude. The middle row is for the case of inhomogeneous broadening with $C_n \propto \sin[n\pi/(N+1)]$. The bottom row is for the case of homogeneous broadening. The curve in the lower left box shows the gain profile of the homogeneous broadening.

be shown that the components of the eigenvector corresponding to the largest eigenvalue (the fundamental supermode) are all phase locked with the same phase. A simple computer program can be employed to obtain all the supermodes and their corresponding eigenvalues. As a result of the homogeneous broadening of the gain profile, the mode amplitudes C_n are further tapered (see Figure 6.22). This leads to a broadening of the main peak and the suppression of the sidelobes. Figure 6.22 shows a comparison of the three different mode-locked pulses and the corresponding mode amplitudes for $N = 8$.

The pulse width of the mode-locked laser with a homogeneous broadening can be estimated as follows. According to Equation (6.7-5), we can write the nth Fourier component of the pulse as

$$C_n' = g_n\sigma_0 C_n + g_{n+1}\sigma_1 C_{n+1} + g_{n-1}\sigma_1 C_{n-1} \tag{6.7-13}$$

The round-trip amplitude change can thus be written

$$\Delta C_n = C_n' - C_n = (g_n\sigma_0 + 2g_n\sigma_1 - 1)C_n + \sigma_1(g_{n+1}C_{n+1} - 2g_nC_n + g_{n-1}C_{n-1}) \tag{6.7-14}$$

where we recall that g_n is the round-trip gain for the nth frequency component. We consider the case of a Lorentzian gain profile,

$$g_n = 1 + \frac{g_0}{1 + \left(\dfrac{n\Omega}{\Omega_g}\right)^2} - L = 1 + g_0 - L - g_0\left(\frac{n\Omega}{\Omega_g}\right)^2 \tag{6.7-15}$$

where g_0 is a constant (the peak round-trip fractional gain), L is the round-trip fractional loss, and Ω_g is the gain bandwidth. Equation (6.7-14) can be transformed into a differential equation by introducing three approximations [25]: (1) the frequency dependence of the gain profile can be expanded to second order in $(\omega - \omega_0) = n\Omega$; (2) the discrete frequency spectrum with Fourier component at $n\Omega$ is replaced with a continuum spectrum, a function of $(\omega - \omega_0) = n\Omega$; and (3) the sum $(g_{n+1}C_{n+1} - 2g_nC_n + g_{n-1}C_{n-1})$ can be replaced by a second-order derivative with respective to frequency if the spectrum is very dense (usually thousands of modes are involved in mode locking). Thus we obtain

$$\Delta C(\omega) = (g_0 - L)C - g_0\left(\frac{\omega - \omega_0}{\Omega_g}\right)^2 C + \sigma_1\Omega^2\frac{d^2gC}{d\omega^2} \tag{6.7-16}$$

where we assume that $\sigma_0 + 2\sigma_1 = 1$ (which corresponds to a peak transmission of 100% for the loss cell, according to Equation (6.7-1)), and Ω is the modulation frequency as well as the intermode frequency spacing. In the steady state, the round-trip change of the amplitude is zero. Hence the mode-locked pulse must be a solution of the differential equation

$$(g_0 - L)C - g_0\left(\frac{\omega - \omega_0}{\Omega_g}\right)^2 C + \sigma_1\Omega^2\frac{d^2gC}{d\omega^2} = 0 \tag{6.7-17}$$

where C is now the Fourier transform of the mode-locked field $E_{in}(t)$ (see Equation (6.7-3)).

The solution of Equation (6.7-17) is a Gaussian pulse (see Problem 6.16)

$$C(\omega - \omega_0) = C_0\exp[-(\omega - \omega_0)^2\tau^2] \tag{6.7-18}$$

where the constant τ (pulse width) must satisfy the following equations:

$$\left(1 + \frac{2g_0}{\Omega_g^2\tau^2}\right)\tau^4 = \frac{g_0}{4\sigma_1(\Omega\Omega_g)^2} \quad \text{and} \quad g_0 - L = 2\sigma_1\Omega^2\tau^2\left(1 + \frac{g_0}{\Omega_g^2\tau^2}\right) \tag{6.7-19}$$

For mode-locked lasers with $g_0 \ll \Omega_g^2\tau^2$, the above equation becomes

$$\tau^4 = \frac{g_0}{4\sigma_1(\Omega\Omega_g)^2} \quad \text{and} \quad g_0 - L = 2\sigma_1\Omega^2\tau^2 \tag{6.7-20}$$

The first equation in (6.7-20) is known as the Kuizenga–Siegman formula [26] for the pulse width, which is inversely proportional to the geometric mean of the gain bandwidth and the modulation frequency. The second equation gives an expression for the round-trip net gain

$$g_0 - L = 2\sigma_1\Omega^2\tau^2 \tag{6.7-21}$$

The peak gain is greater than the loss. This is legitimate and does not cause instabilities, since the modulation leads to an increase of loss in the pulse wings. The temporal pulse profile $E(t)$ is thus given by

$$E(t) = \int_{-\infty}^{\infty} C(\omega)e^{i\omega t}d\omega = \frac{C_0}{\tau}\sqrt{\pi}e^{i\omega_0 t}\exp\left(-\frac{t^2}{4\tau^2}\right) \tag{6.7-22}$$

The full width at half-maximum of the intensity profile is thus given by

$$\tau_p = 2\tau\sqrt{2\ln 2} = 2\sqrt{\ln 2}\left(\frac{g_0}{\sigma_1}\right)^{1/4}\frac{1}{\sqrt{\Omega\Omega_g}} \tag{6.7-23}$$

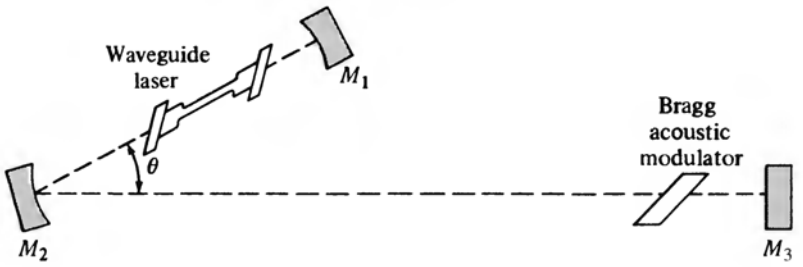


Figure 6.23 Schematic drawing of the mode-locking experiment in a high-pressure CO_2 laser. (After Reference [27].)

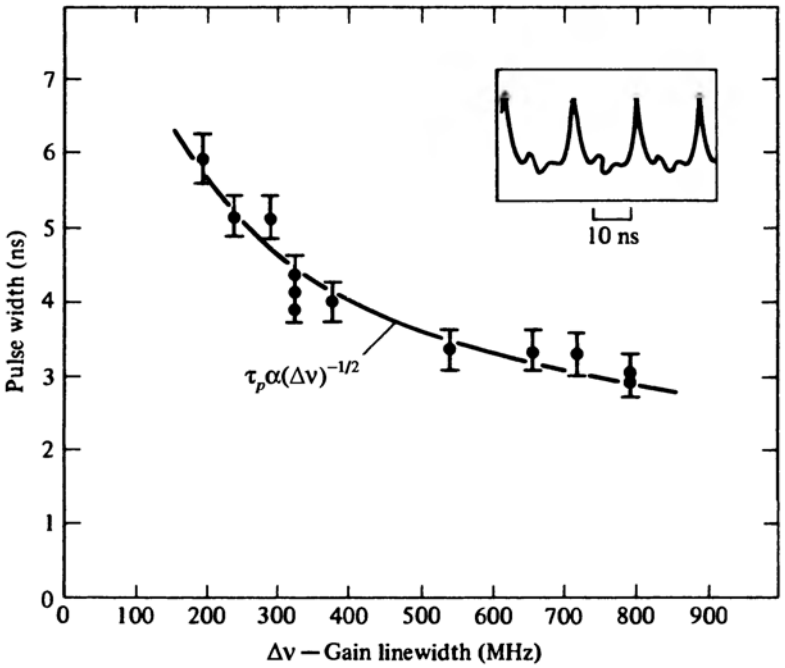


Figure 6.24 The dependence of the pulse width on the gain linewidth, $\Delta\nu = \Omega_g/2\pi$, that is controlled by varying the pressure ($\Delta\nu = 8 \times 10^8$ at 150 torr). (After Reference [27].)

where we recall that Ω is the angular frequency spacing between the longitudinal modes and Ω_g is the gain bandwidth.

An experimental setup demonstrating mode locking in a pressure-broadened CO_2 laser is sketched in Figure 6.23. The inverse square root dependence of τ_p on $\Delta\nu$ is displayed by the data of Figure 6.24, while the dependence on the modulation parameter σ_1 is shown in Figure 6.25.

Mode Locking by Phase Modulation

Mode locking can be induced by internal phase, rather than loss, modulation. This is usually done by using an electro-optic crystal inside the resonator oriented in the appropriate orientation such that the passing wave undergoes a phase delay proportional to the instantaneous electric field in the crystal. The frequency of the modulating signal is equal, as in the loss modulation case, to the inverse of the round-trip group delay time, that is, to the longitudinal intermode frequency separation.

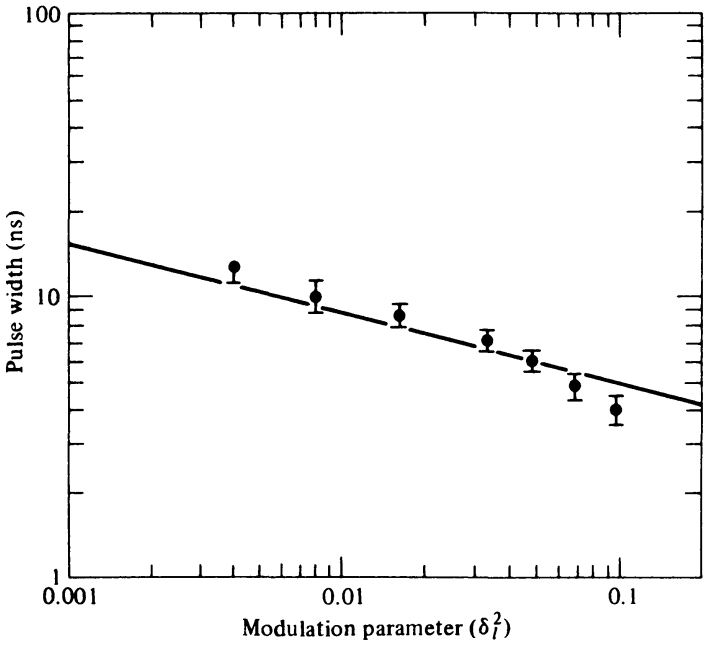


Figure 6.25 The mode-locked pulse width as a function of the modulation parameter, $\delta_l^2 = 2\sigma_1$. (After Reference [27].)

We consider a transmission function of a phase modulator as

$$T(t) = \exp(-i\delta \sin \Omega t) = \exp(-i\delta \sin 2\pi \Delta\nu_{\text{axial}}\, t) \tag{6.7-24}$$

where δ is a constant, known as the depth of modulation. Such an amplitude transmission function can be written in terms of its Fourier components as

$$T(t) = \exp(-i\delta \sin \Omega t) = \sum_{n=-\infty}^{+\infty} J_n(\delta)\, e^{-in\Omega t} \tag{6.7-25}$$

where $J_n(\delta)$ is the Bessel function. We note that a phase modulator will generate an infinite series of sidebands. These sidebands coincide with the modes of the resonator. Thus a periodically modulated phase modulator will couple all the modes of the resonator provided that the modulation frequency equals the longitudinal mode spacing.

Following the matrix analysis described earlier, we can obtain the supermodes of oscillation of the mode-locked laser by solving the following eigenvalue problem:

$$g\begin{bmatrix} J_0(\delta) & J_1(\delta) & J_2(\delta) & \cdots & J_{N-1}(\delta) \\ J_{-1}(\delta) & J_0(\delta) & J_1(\delta) & \cdots & J_{N-2}(\delta) \\ J_{-2}(\delta) & J_{-1}(\delta) & J_0(\delta) & \cdots & J_{N-3}(\delta) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ J_{-N+1}(\delta) & J_{-N+2}(\delta) & J_{-N+3}(\delta) & \cdots & J_0(\delta) \end{bmatrix}\begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_N \end{bmatrix} = \eta\begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_N \end{bmatrix} \tag{6.7-26}$$

where g is the round-trip net gain. The mode amplitudes C_n can be obtained by solving the eigenvalue problem. As a result of the antisymmetric property of the matrix, the eigenvectors and eigenvalues are in general complex. Although the mode amplitudes are locked, the complex amplitudes lead to "fixed" phase delays among the modes. The phase delays lead to a mode-locked pulse with a frequency chirping.

Passive Mode Locking Via Saturable Absorber

As mentioned in Section 6.6, a saturable absorber exhibits a low absorption for optical beams with high intensities, and a high absorption for optical beams with low intensities. As a result,

such an absorber will favor the transmission of short optical pulses with high peak intensities in the laser resonator. In what follows, we describe an elementary theory of passive mode locking by assuming a fast response of the saturable absorber [25]. In passive mode locking, the loss modulator is replaced by a saturable absorber. The loss modulation of the saturable absorber $s(t)$ in transmission through the absorber is given by

$$s(t) = \frac{s_0}{1 + I(t)/I_{\text{sat}}} \tag{6.7-27}$$

where s_0 is the unsaturated loss (i.e., loss at low intensities), $I(t)$ is the instantaneous intensity of the pulse, and I_{sat} is the saturation intensity. If the saturation is relatively weak, Equation (6.7-27) can be written

$$s(t) = s_0 - s_0 I(t)/I_{\text{sat}} \tag{6.7-28}$$

If we normalize the mode amplitude so that $|E(t)|^2 = \text{Power}$, then the transmission through the absorber can be written

$$T(t) = 1 - s(t) = 1 - s_0 + s_0 \frac{|E(t)|^2}{I_{\text{sat}} A_{\text{eff}}} = 1 - s_0 + \gamma |E(t)|^2 \tag{6.7-29}$$

where A_{eff} is the effective area of the mode and γ is the self amplitude modulation (SAM) coefficient. The self amplitude modulation is a direct result of the saturation behavior of the absorption. According to Equation (6.7-29), the transmission is an increasing function of the field intensity. The round-trip amplitude change in the case of passive mode locking can now be obtained by modifying Equation (6.7-16). Using the Fourier transform relationship (6.7-22), we can transform Equation (6.7-16) to the time domain. This leads to

$$\Delta E(t) = (g_0 - L)E(t) + \frac{g_0}{\Omega_g^2} \frac{\partial^2}{\partial t^2} E(t) - \sigma_1 \Omega^2 t^2 \int g C e^{i\omega t}\, d\omega \tag{6.7-30}$$

We replace the active mode-locking term $\sigma_1 \Omega^2 t^2 \int g C e^{i\omega t}\, d\omega$ with the self amplitude modulation (there's no active modulation). This leads to

$$\Delta E(t) = (g_0 - L)E(t) + \frac{g_0}{\Omega_g^2} \frac{\partial^2}{\partial t^2} E(t) + \gamma |E(t)|^2 E(t) \tag{6.7-31}$$

At steady state, $\Delta E(t) = 0$, the solution is a simple hyperbolic secant (see Problem 6.17):

$$E(t) = E_0 \operatorname{sech}(t/\tau)$$

where E_0 and τ are constants, with

$$\tau^2 = \frac{2g_0}{\gamma E_0^2 \Omega_g^2} \tag{6.7-32}$$

and

$$L - g_0 = \frac{g_0}{\Omega_g^2 \tau^2} \tag{6.7-33}$$

The full width at half-maximum of the intensity profile is thus given by

$$\tau_p = 1.67\tau = 1.67 \sqrt{\frac{2g_0}{\gamma}} \frac{1}{E_0 \Omega_g} \tag{6.7-34}$$

It is interesting to note that expression (6.7-32) for the pulse width of passive mode locking is related to that of active mode locking (6.7-19). In active mode locking, $\sigma_1\Omega^2$ is proportional to the curvature of the amplitude modulation (second-order derivative of transmission). In the case of self amplitude modulation in passive mode locking, the curvature is proportional to $\gamma E_0^2/\tau^2$. Thus Equations (6.7-32) and (6.7-19) are indeed related. This comparison also explains why passive mode locking can result in much shorter pulse widths for the same gain bandwidth Ω_g. As the pulse gets shorter, the curvature of self-modulation increases as $\gamma E_0^2/\tau^2$, whereas it remains unchanged ($\sigma_1\Omega^2$) for active mode locking. In the resonator, the net gain is negative preceding and following the pulse (i.e., at low intensities, according to Equation (6.7-33)). At the peak of the pulse, the net gain is positive due to bleaching of the saturable absorber.

6.8 PULSE LENGTH MEASUREMENT AND NARROWING OF CHIRPED PULSES

The problem of measuring the duration of mode-locked ultrashort pulses is of great practical and theoretical interest. Since the fastest conventional optical detectors possess response times of $\sim 2 \times 10^{-11}$ s, it is impossible to use these optical detectors to measure directly the short ($\tau < 10^{-11}$ s) mode-locked pulses. A number of techniques invented for this purpose all take advantage of some nonlinear process to obtain a spatial autocorrelation trace of the optical intensity pulse. The measurement of a pulse of duration, say, $\tau_0 = 10^{-12}$ s is thus replaced with measuring the spatial extent of an autocorrelation trace of length $c\tau_0 = 0.3$ mm, which is a relatively simple task.

In what follows we will describe one such method, the one most widely used, that is based on the phenomenon of optical second harmonic generation. The process of second harmonic generation is developed in detail in Chapter 8. It will suffice for the purpose of the present discussion to state that when an optical pulse

$$e_1(t) = \mathrm{Re}[\mathscr{E}_1(t)e^{i\omega t}] \tag{6.8-1}$$

is incident on a nonlinear optical crystal, it generates an output optical pulse $e_2(t)$ at twice the frequency with

$$e_2(t) = \mathrm{Re}[\mathscr{E}_2(t)e^{2i\omega t}] \propto \mathrm{Re}[\mathscr{E}_1^2(t)e^{2i\omega t}] \tag{6.8-2}$$

A sketch of a second harmonic system for measuring the pulse length is shown in Figure 6.26. The laser emits a continuous stream of mode-locked pulses. Each individual pulse $\mathscr{E}(t)e^{i\omega t}$ is divided by a beam splitter into two equal intensity pulses. One of these pulses is advanced (or delayed) by τ seconds relative to the other. The two pulses recombine again in a nonlinear optical crystal. The second harmonic (2ω) pulse generated by the crystal is incident on a "slow" detector whose output current is integrated over a time long compared to the optical pulse duration.

The total optical field incident on the nonlinear crystal is the sum of the direct and retarded fields:

$$\begin{aligned} e_{\mathrm{tot}}(t) &= \mathrm{Re}\{[\mathscr{E}_1(t) + \mathscr{E}_1(t-\tau)e^{-i\omega\tau}]e^{i\omega t}\} = \mathrm{Re}[\mathscr{E}(t)e^{i\omega t}] \\ \mathscr{E}(t) &= \mathscr{E}_1(t) + \mathscr{E}_1(t-\tau)e^{-i\omega\tau} \end{aligned} \tag{6.8-3}$$

According to Equation (6.8-2), the second harmonic field radiated by the crystal has a complex amplitude that is proportional to the square of the complex amplitude $\mathscr{E}(t)$ of the incident fundamental field:

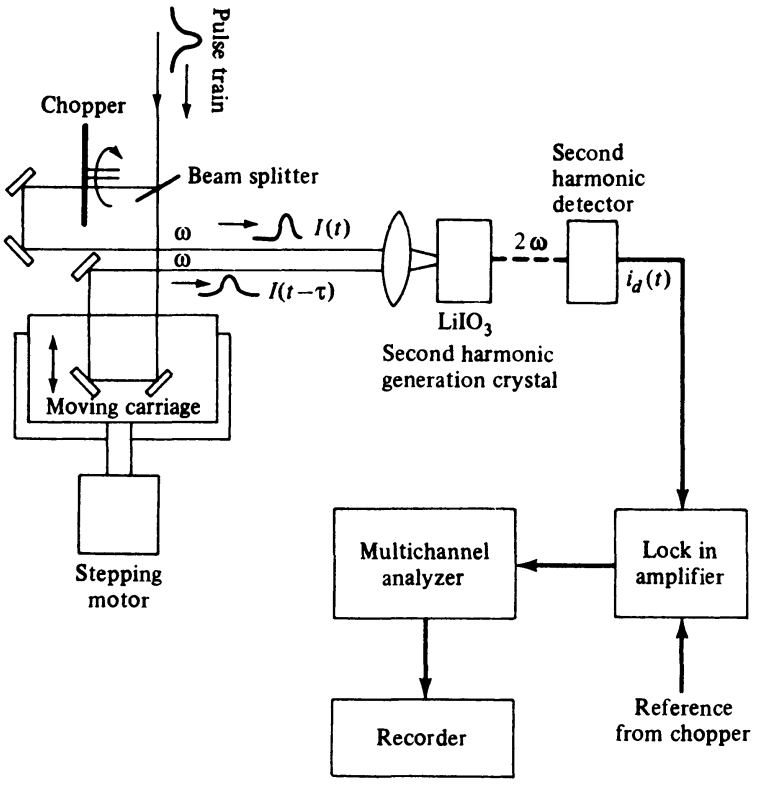


Figure 6.26 The second harmonic generation autocorrelation setup for measuring the width of mode-locked ultrashort pulses.

$$\begin{aligned}\mathscr{E}_2(t) &\propto [\mathscr{E}_1(t) + \mathscr{E}_1(t-\tau)e^{-i\omega\tau}]^2 \\ &= \mathscr{E}_1^2(t) + \mathscr{E}_1^2(t-\tau)e^{-2i\omega\tau} + 2\mathscr{E}_1(t)\mathscr{E}_1(t-\tau)e^{-i\omega\tau}\end{aligned} \tag{6.8-4}$$

The second harmonic field, $e_2(t) = \mathrm{Re}[\mathscr{E}_2(t)\exp(i2\omega t)]$, is incident next on the optical detector (photomultiplier, diode, etc.) whose output current i_d (see Section 11.1) is proportional to the incident intensity. Using Equation (6.8-4), we can obtain

$$\begin{aligned}i_d(t) \propto \mathscr{E}_2(t)\mathscr{E}_2^*(t) = [\mathscr{E}_1(t)\mathscr{E}_1^*(t)]^2 + [\mathscr{E}_1(t-\tau)\mathscr{E}_1^*(t-\tau)]^2 \\ + 4\mathscr{E}_1(t)\mathscr{E}_1^*(t)\mathscr{E}_1(t-\tau)\mathscr{E}_1^*(t-\tau) + s(\tau)\end{aligned} \tag{6.8-5}$$

where $s(\tau)$ is composed of terms with $\cos\omega\tau$ and $\cos 2\omega\tau$ dependence. Since these terms fluctuate with a delay period $\Delta\tau \sim 10^{-15}$ s, a small unintentional or deliberate integration (smearing) over the delay τ averages them out to near zero. The term $s(\tau)$ is consequently left out.

Since the temporal (t) variation of the first three terms in Equation (6.8-5) is on the scale of picoseconds (or less), the much slower optical detector inevitably integrates the current $i_d(t)$, with the result that the actual output from the optical detector is a function of the delay (τ) only,

$$i_d(\tau) \propto \langle I^2(t)\rangle + \langle I^2(t-\tau)\rangle + 4\langle I(t)I(t-\tau)\rangle \tag{6.8-6}$$

where the angle brackets signify time averaging and the *intensity* $I(t)$ is defined[4] as $I(t) = \mathscr{E}_1(t)\mathscr{E}_1^*(t)$. By dividing both sides of Equation (6.8-6) by $\langle I^2(t)\rangle$ and recognizing that $\langle I^2(t)\rangle = \langle I^2(t-\tau)\rangle$, the normalized detector output becomes

[4] A proportionality constant involved in this definition is left out since it cancels out in the subsequent division of Equation (6.8-9).

$$i_d(\tau) = 1 + 2G^{(2)}(\tau) \tag{6.8-7}$$

where $G^{(2)}(\tau)$, the second-order autocorrelation function of the intensity pulse, is defined by

$$G^{(2)}(\tau) \equiv \frac{\langle I(t)I(t-\tau)\rangle}{\langle I^2(t)\rangle} \tag{6.8-8}$$

In the case of a well-behaved ultrashort coherent light pulse of duration τ_0, we have

$$i_d(0) = 3, \quad i_d(\tau \gg \tau_0) = 1$$

since $G^{(2)}(0) = 1$ and $G^{(2)}(\tau \gg \tau_0) = 0$.

A plot of $i_d(\tau)$ versus τ will consist of a peak of (normalized) height of 3 atop a background of unity height. The central peak will have a width $\sim\tau_0$.

It is important in practice to be able to distinguish between the case just discussed and that of incoherent light (such as light due to a laser oscillating in a large number of independent modes). In this case, we have $i_d(0) = 3$ (since even incoherent light is correlated with itself at zero delay). For $\tau > 0$, or more precisely for τ longer than the coherence time of the light, we have

$$G^{(2)}(\tau) = \frac{\langle I(t)I(t-\tau)\rangle}{\langle I^2(t)\rangle} = \frac{\langle I(t)\rangle^2}{\langle I^2(t)\rangle} \tag{6.8-9}$$

since $I(t)$ and $I(t-\tau)$ are completely uncorrelated. For truly incoherent light of the type we are considering here, the time averaging indicated in Equation (6.8-9) can be replaced by ensemble averaging so that

$$\langle I^2(t)\rangle = \int_0^\infty p(I)I^2\,dI \tag{6.8-10}$$

where $p(I)$ is the intensity distribution function so that $p(I)\,dI$ is the probability that a measurement of I will result in a value between I and $I + dI$. For incoherent light[5]

$$p(I) = \frac{1}{\langle I\rangle} e^{-I/\langle I\rangle}$$

which when used in Equation (6.8-10) gives

$$\langle I^2(t)\rangle = 2\langle I\rangle^2$$

and, returning to Equation (6.8-7),

$$i_d(\tau > 0) \propto 1 + 2\frac{\langle I(t)\rangle^2}{\langle I^2(t)\rangle} = 2$$

A plot of $i_d(\tau)$ versus τ in the case of incoherent light should thus consist of a very narrow peak of height 3 on a background of height 2. The general features of the coherent mode-locked pulses and the incoherent light are depicted in Figure 6.27.

The determination of the original pulse width from the width of the second harmonic correlation trace is somewhat ambiguous. We can show by performing the integration indicated by Equation (6.8-8) that the width (at half-maximum) t_0 of $G^{(2)}(\tau)$ and τ_0 of $I(t)$ are related as in the case of the "popular" waveforms tabulated in the left column of Table 6.2.

---

[5] This follows directly from the fact that the optical field distribution $p(e)$ for the field of an incoherent beam function is Gaussian.

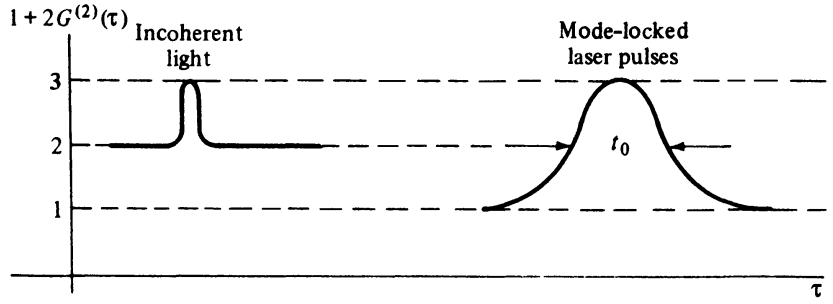


Figure 6.27 The second harmonic (averaged) integrated intensity due to two optical pulses as a function of time delay between them for the case (left) of an incoherent source and (right) coherent mode-locked optical pulses.

TABLE 6.2 Some Simple Pulse Widths

$I(t)$	t_0/τ_0
1 $(0 \le t \le t_0)$, zero otherwise	1
$\exp\left(-\frac{(4 \ln 2)t^2}{\tau_0^2}\right)$	$\sqrt{2}$
$\text{sech}^2\left(\frac{1.76t}{\tau_0^2}\right)$	1.55
$\exp\left(-\frac{(\ln 2)t}{\tau_0}\right)$ $(t \ge 0)$	2

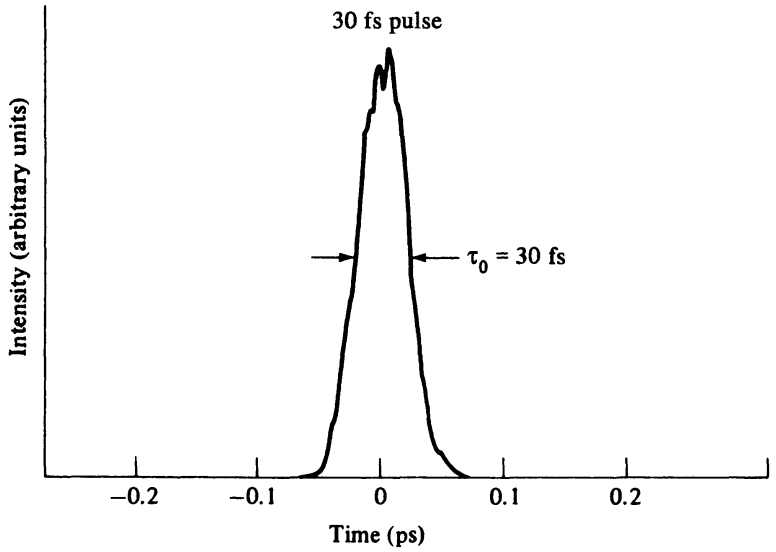


Figure 6.28 The autocorrelation trace of a mode-locked dye laser ($\lambda \sim 6100$ Å) pulse.

We conclude this section by showing in Figure 6.28 the autocorrelation trace of one of the shortest optical pulses ($\tau_0 \sim 30 \times 10^{-15}$ s) produced to date. It is interesting to note that within such a pulse the light rises and falls a mere 15(!) times.

With just one order of magnitude improvement we should thus be able to isolate a single optical cycle. There may, however, be some compelling reasons (and the student is encouraged to think of some) why this development may not take place anytime soon.

Pulse Narrowing by Chirping and Compression

Mode locking of a laser oscillator was shown in Section 6.6 to lead to a mode of oscillation consisting of a continuous train of ultrashort pulses. The width of each pulse is $\tau_p \sim 1/(\Delta\omega)_{\text{gain}}$, where $(\Delta\omega)_{\text{gain}}$ is the spectral width over which the gain is capable of sustaining oscillation. It is, however, possible to further reduce the pulse width beyond the limit $(\Delta\omega)_{\text{gain}}^{-1}$ by employing nonlinear optical techniques. To lend some plausibility to the new ideas that we will soon introduce, we remind ourselves of the exact formal analogy that exists between the diffractive propagation of an optical beam in free homogeneous space and that of an optical pulse propagation in a dispersive channel (fiber).

The propagation of an optical pulse, $f_1(t)\exp(i\omega_0 t)$, through a length L of a dispersive fiber can be described (see Equation (7.2-5)) by

$$\tilde{f}_2(\Omega) = \tilde{f}_1(\Omega)\exp\left(-i\frac{\beta''}{2}L\Omega^2\right) \tag{6.8-11}$$

where $\tilde{f}_1(\Omega)$ and $\tilde{f}_2(\Omega)$ are the Fourier transforms of the input and output pulse envelopes $f_1(t)$ and $f_2(t)$, respectively, and $\beta'' \equiv d^2\beta/d\omega^2$.

The propagation of a confined optical beam with a transverse profile $u_1(x)$ was found in Section 2.12 (see Equation (2.12-8)) to obey

$$\tilde{u}_2(K) = \tilde{u}_1(K)\exp\left(i\frac{LK^2}{2k}\right) \tag{6.8-12}$$

The exact formal analogy of Equations (6.8-11) and (6.8-12) suggests that is should be possible to construct a temporal analog of the familiar spatial optical demagnification by a lens as illustrated in Figure 6.29. Such a system can be used to narrow optical pulses. All that is needed is a time lens, a device that multiplies an incoming optical field by $\exp(iat^2)$, which is the exact analog of the factor $\exp(ikx^2/2f)$ by which the spatial lens multiplies an incoming beam.

Our temporal pulse narrower is illustrated in Figure 6.29a. The input pulse envelope, before narrowing, is taken as a Gaussian

$$f_1(t) = \exp(-t^2/\tau_p^2) \tag{6.8-13}$$

with a (FWHM) intensity width $\Delta\tau = \sqrt{2\ln 2}\,\tau_p$. Passage through the time lens multiplies $f_1(t)$ by a phase factor $\exp(iAt^2/\tau_p^2)$:

$$f_2(t) = \exp[-(1 - iA)(t/\tau_p)^2] \tag{6.8-14}$$

The total field at plane 2 is

$$E_2(t) = f_2(t)\exp(i\omega_0 t) = \exp[-(t/\tau_p)^2]\exp i\left[\omega_0 t + A\left(\frac{t}{\tau_p}\right)^2\right] \tag{6.8-15}$$

If we write $E_2(t) = \exp[-(t/\tau_p)^2 + i\phi(t)]$, the instantaneous frequency is $\omega(t) = d\phi/dt$

$$\omega(t) = \frac{d}{dt}\left[\omega_0 t + A\left(\frac{t}{\tau_p}\right)^2\right] = \omega_0 + 2A\frac{t}{\tau_p^2} \tag{6.8-16}$$

corresponding to a linear "chirp." The pulse width, however, is unaffected by the lens. The lens, however, modifies the spectrum of the pulse. To find the spectrum we take the Fourier transform of $f_2(t)$:

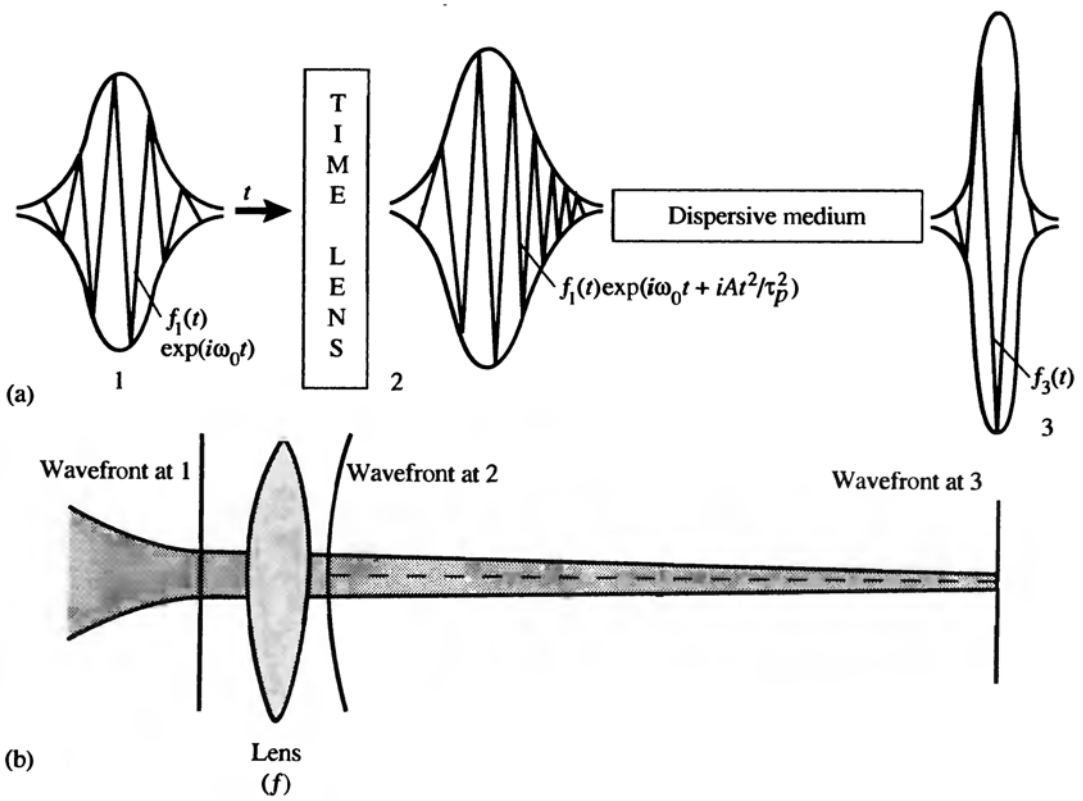


Figure 6.29 The equivalence between temporal focusing (pulse narrowing) in (a) and spatial focusing by a conventional lens in (b). Note that a chirped optical pulse corresponds to a Gaussian spatial beam with a curved wavefront.

$$\tilde{f}_2(\Omega) = \sqrt{\frac{\tau_p^2}{4\pi(1 - iA)}} \exp\left(-\frac{1 + iA}{4(1 + A^2)}(\Omega\tau_p)^2\right) \tag{6.8-17}$$

The effect of the chirp is thus to increase the spectral width (FWHM) by the factor $\sqrt{1 + A^2}$

TIME LENSES

The fact that the time lens has increased the spectral width while preserving the total pulse energy implies redistribution of power among the spectral components. This, by definition, cannot be accomplished by a linear passive device. We will, indeed, find that our practical candidates for time lenses all employ nonlinear techniques. One such method employs self phase modulation in a medium with a large Kerr coefficient, that is, a medium where the index of refraction depends on the light intensity, $n = n_0 + n_2 I$ (Kerr effect). I = intensity (W/m^2), and n_2 a constant of the material. Consider, for example, a Gaussian intensity pulse,

$$I = I_0 \exp(-2\alpha t^2)$$

entering a nonlinear medium characterized by n_2 so that the index of refraction is $n = n_0 + n_2 I(t)$

$$E_{\text{in}} = E_0 \exp(-\alpha t^2)$$

$$I_{\text{in}} = I_0 \exp(-2\alpha t^2)$$

The pulse will emerge at $z = L$ with a phase delay factor,

$$\exp\left[i\frac{\omega}{c}(n_0 + n_2 I(t))L\right]$$

If we expand the Gaussian pulse as $I(t) = I_0(1 - 2\alpha t^2 + \dots)$ and keep only the first two terms, we find that the delay factor becomes

$$E_{\text{out}} = E_{\text{in}} \times \text{delay factor}$$

$$= E_0 \exp(-\alpha t^2) \exp\left(i\frac{2\omega n_2 \alpha I_0 L}{c} t^2\right) \tag{6.8-18}$$

and has, according to Equation (6.8-15), the requisite chirp to serve as a time lens. We left out an inconsequential fixed phase and group delay.

The pulse $f_2(t)$ is thus no longer transform-limited since its time–bandwidth product

$$\Delta\nu\,\Delta\tau = \frac{2\ln 2}{\pi}(1 + A^2)^{1/2} = 0.4413(1 + A^2)^{1/2} \tag{6.8-19}$$

exceeds the minimum value for a Gaussian pulse by $(1 + A^2)^{1/2}$. This excess bandwidth should make it possible, in principle, to reduce the pulse width, employing passive means, by the factor $(1 + A^2)^{1/2}$, which will result, again, in a compressed transform-limited pulse.

In the spatial example of Figure 6.29a, the narrowing of the spatially chirped pulse is accomplished by propagating it through the appropriate distance L in space. The temporal equivalent is to propagate the chirped pulse through a dispersive device with transfer characteristics of the form (6.8-11). This results in

$$\tilde{f}_3(\Omega) = \tilde{f}_2(\Omega)\exp\left(-i\frac{\beta''}{2}L\Omega^2\right)$$

$$= \sqrt{\frac{\tau_p^2}{4\pi(1 - iA)}}\exp\left[-\frac{\Omega^2\tau_p^2}{4(1 + A^2)} - i\left(\frac{A\tau_p^2}{4(1 + A^2)} + \frac{\beta''L}{2}\right)\Omega^2\right] \tag{6.8-20}$$

where L is the length of the dispersive channel. If the condition

$$\beta''L = -\frac{A\tau_p^2}{2(1 + A^2)} \tag{6.8-21}$$

is satisfied, the output is given by

$$\tilde{f}_3(\Omega) = \sqrt{\frac{\tau_p^2}{4\pi(1 - iA)}}\exp\left(-\frac{(\Omega\tau_p)^2}{4(1 + A^2)}\right) \tag{6.8-22}$$

which corresponds in the time domain to

$$f_3(t) = F^{-1}\{\tilde{f}_3(\Omega)\} = \sqrt{1 + iA}\exp\left(-(1 + A^2)\frac{t^2}{\tau_p^2}\right) \tag{6.8-23}$$

The width of the output is thus compressed to

$$(\Delta\tau)_{\text{comp}} = \frac{\Delta\tau}{\sqrt{1 + A^2}} \tag{6.8-24}$$

For $A \gg 1$, the compression ratio is $\sim A$. This compression ratio is thus exactly the factor by which the spectral width is increased by the chirping. A simple integration shows that

$$\int_{-\infty}^{\infty} |f_1(t)|^2\, dt = \int_{-\infty}^{\infty} |f_2(t)|^2\, dt = \int_{-\infty}^{\infty} |f_3(t)|^2\, dt \tag{6.8-25}$$

so that the pulse energy is conserved.

The Grating Pair Compressor

In the above example, a fiber with group velocity dispersion ($\beta'' \neq 0$) was employed in order to compress the chirped pulse. The essential feature of the fiber was its transfer function, given by Equation (6.8-20) as

$$\text{Transfer function of fiber (length } L) = \exp\left(-\frac{i\beta''}{2} L\Omega^2\right)$$

Any other device with a transfer function of the form $\exp(ib\Omega^2)$, where b is some real constant, can thus serve as compressor provided b can be adjusted in magnitude as well as sign. A commonly used pulse-narrowing configuration is the dual grating telescope compressor [16, 28, 29] illustrated in Figure 6.30. It is based on the fact that different Fourier components (Ω) of the incident optical beam are diffracted by the gratings along different directions, thus following paths of different length between the input and output, accumulating in the process a differential phase $b\Omega^2$. (See Problem 6.19)

Figure 6.31 shows intensity profiles, obtained by second harmonic autocorrelation, of a chirped and a compressed pulse in the experiment depicted by Figure 6.30. The input pulse is compressed from an initial width of 5.2 ps to 0.32 ps. This corresponds, according to Equation (6.8-24), to a compression ratio $A \sim 16$. The $\Delta\nu(\Delta\tau)_{\text{comp}}$ product is ~0.45, close to the theoretical limit of 0.44, which indicates that the pulse is as narrow as allowed by its spectral content; that is, it is transform-limited.

A semiconductor laser excited by current pulses or pulsating due to mode locking is naturally chirped. This is due to carrier density—hence, index of refraction transients in the active region. This makes the output pulses of such lasers natural candidates for compression. Chirping in semiconductor lasers is treated in Chapter 15.

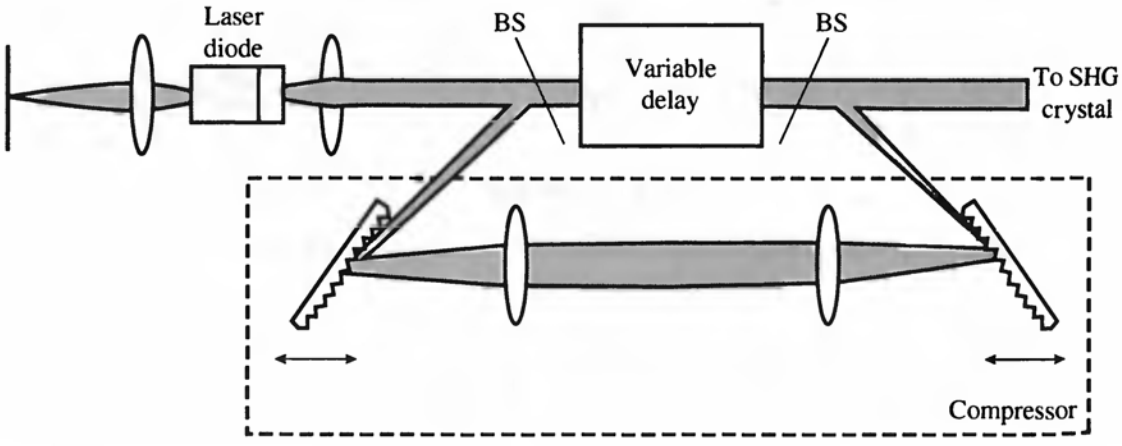


Figure 6.30 A dual grating telescope pulse compressor. (After Reference [43].)

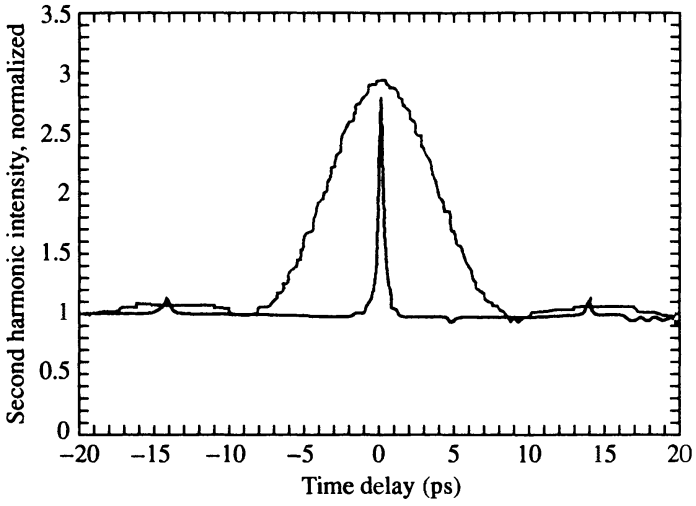


Figure 6.31 Intensity autocorrelation of uncompressed and compressed pulses. The compression ratio is ~$A = 16$. (After Reference [16].)

6.9 GIANT PULSE (*Q*-SWITCHED) LASERS

The "*Q*-switching" technique [30] is used to obtain intense and short bursts of oscillation from lasers; see References [31–33]. The quality factor Q of the optical resonator is degraded (lowered) by some means during the pumping so that the gain (i.e., inversion $N_2 - N_1$) can build up to a very high value without oscillation. (The spoiling of the Q raises the threshold inversion to a value higher than that obtained by pumping.) When the inversion reaches its peak, the Q is restored abruptly to its (ordinary) high value. The gain (per pass) in the laser medium is now well above threshold. This causes an extremely rapid buildup of the oscillation and a simultaneous exhaustion of the inversion by stimulated $2 \rightarrow 1$ transitions. This process converts most of the energy that was stored by atoms pumped into the upper laser level into photons, which are now inside the optical resonator. These proceed to bounce back and forth between the reflectors with a fraction $(1 - R)$ "escaping" from the resonator each time. This causes a decay of the pulse with a characteristic time constant (the "photon lifetime") given in Equation (4.7-3) as

$$t_c \simeq \frac{n_0 l}{c(1 - R)}$$

where n_0 is the index of refraction of the medium.

Both experiment and theory indicate that the total evolution of a giant laser pulse as described above is typically completed in $\sim 2 \times 10^{-8}$ s. We will consequently neglect the effects of population relaxation and pumping that take place during the pulse. We will also assume that the switching of the Q from the low to the high value is accomplished instantaneously.

The laser is characterized by the following variables: ϕ—the total number of photons in the optical resonator; $n \equiv (N_2 - N_1)V$—the total inversion; and t_c—the decay time constant for photons in the *passive* resonator. The exponential gain constant γ is proportional to n. The radiation intensity I thus grows with distance as $I(z) = I_0 \exp(\gamma z)$ and $dI/dz = \gamma I$. An observer traveling with the wave velocity will see it grow at a rate

$$\frac{dI}{dt} = \frac{dI}{dz}\frac{dz}{dt} = \gamma\left(\frac{c}{n_0}\right)I$$

and thus the temporal exponential growth constant is $\gamma(c/n_0)$. If the laser rod is of length L while the resonator length is l, then only a fraction L/l of the photons are undergoing amplification at any one time and the average growth constant is $\gamma c(L/n_0 l)$. We can thus write

$$\frac{d\phi}{dt} = \phi\left(\frac{\gamma cL}{n_0 l} - \frac{1}{t_c}\right) \tag{6.9-1}$$

where $-\phi/t_c$ is the decrease in the number of resonator photons per unit time due to incidental resonator losses and to the output coupling. Defining a dimensionless time by $\tau = t/t_c$, we obtain, upon multiplying Equation (6.9-1) by t_c,

$$\frac{d\phi}{d\tau} = \phi\left[\left(\frac{\gamma}{n_0 l/cLt_c}\right) - 1\right] = \phi\left(\frac{\gamma}{\gamma_t} - 1\right)$$

where $\gamma_t = (n_0 l/cLt_c)$ is the minimum value of the gain constant at which oscillation (i.e., $d\phi/d\tau = 0$) can be sustained. Since, according to Equation (5.6-19), γ is proportional to the inversion n, the last equation can also be written

$$\frac{d\phi}{d\tau} = \phi\left(\frac{n}{n_t} - 1\right) \tag{6.9-2}$$

where $n_t = N_t V$ is the total inversion at threshold as given by Equation (6.1-15).

The term $\phi(n/n_t)$ in Equation (6.9-2) gives the number of photons generated by induced emission per unit of normalized time. Since each generated photon results from a single transition, it corresponds to a decrease of $\Delta n = -2$ in the total inversion. We can thus write directly

$$\frac{dn}{d\tau} = -2\phi\frac{n}{n_t} \tag{6.9-3}$$

The coupled pair of equations, (6.9-2) and (6.9-3), describes the evolution of ϕ and n. It can be solved easily by numerical techniques. Before we proceed to give the results of such calculation, we will consider some of the consequences that can be deduced analytically.

Dividing Equation (6.9-2) by (6.9-3) results in

$$\frac{d\phi}{dn} = \frac{n_t}{2n} - \frac{1}{2}$$

and, by integration,

$$\phi - \phi_i = \frac{1}{2}\left[n_t \ln\frac{n}{n_i} - (n - n_i)\right]$$

where n_i is the initial population inversion and ϕ_i is the initial photon number. Assuming that ϕ_i, the initial number of photons in the cavity, is negligible, we obtain

$$\phi = \frac{1}{2}\left(n_t \ln\frac{n}{n_i} - (n - n_i)\right) \tag{6.9-4}$$

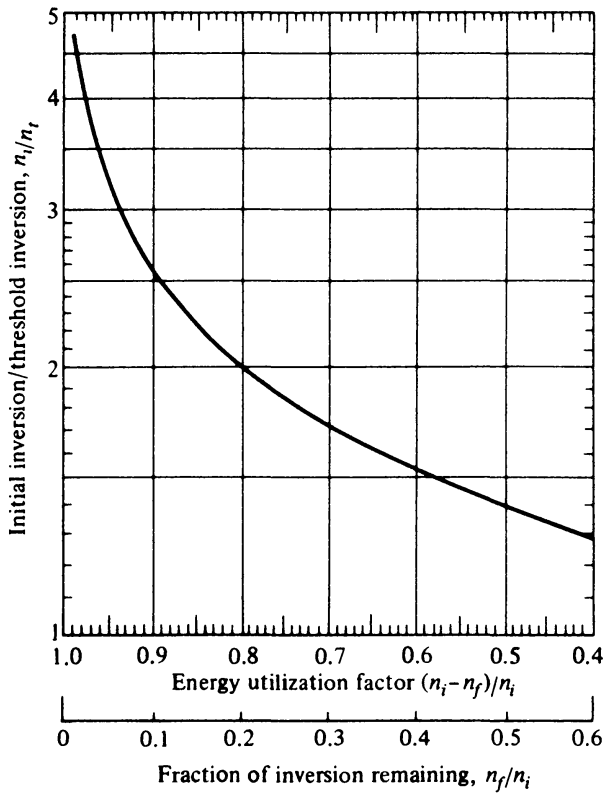


Figure 6.32 Energy utilization factor $(n_i - n_f)/n_i$ and inversion remaining after the giant pulse. (After Reference [30].)

for the relation between the number of photons ϕ and the inversion n at any moment. At $t \gg t_c$, the photon density ϕ will be zero so that setting $\phi = 0$ in Equation (6.9-4) results in the following expression for the final inversion n_f:

$$\frac{n_f}{n_i} = \exp\left(\frac{n_f - n_i}{n_t}\right) \tag{6.9-5}$$

This equation is of the form $(x/a) = \exp(x - a)$, where $x = n_f/n_t$ and $a = n_i/n_t$, so that it can be solved graphically (or numerically) for n_f/n_i as a function of n_i/n_t.[6] The result is shown in Figure 6.32. We notice that the fraction of the energy originally stored in the inversion that is converted into laser oscillation energy is $(n_i - n_f)/n_i$ and that it tends to unity as n_i/n_t increases.

The instantaneous power output of the laser is given by $P = \phi h\nu/t_c$ or, using Equation (6.9-4), by

$$P = \frac{h\nu}{2t_c}\left(n_t \ln\frac{n}{n_i} - (n - n_i)\right) \tag{6.9-6}$$

Of special interest to us is the peak power output. Setting $\partial P/\partial n = 0$, we find that maximum power occurs when $n = n_t$. Putting $n = n_t$ in Equation (6.9-6) gives

$$P_p = \frac{h\nu}{2t_c}\left(n_t \ln\frac{n_t}{n_i} - (n_t - n_i)\right) \tag{6.9-7}$$

[6] This can be done by assuming a value of a and finding the corresponding x at which the plots of x/a and $\exp(x - a)$ intersect.

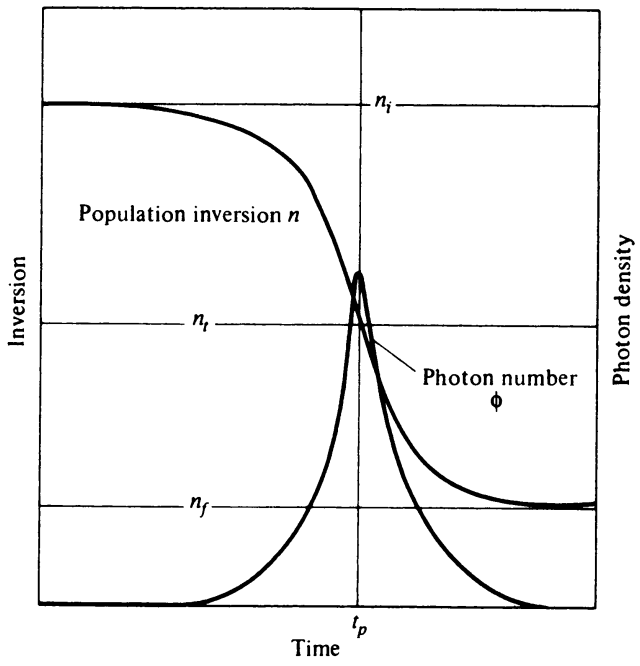


Figure 6.33 Inversion and photon density during a giant pulse. (After Reference [30].)

for the peak power. If the initial inversion is well in excess of the (high-Q) threshold value (i.e., $n_i \gg n_t$), we obtain from Equation (6.9-7)

$$(P_p)_{n_i \gg n_t} \simeq \frac{n_i h\nu}{2t_c} \tag{6.9-8}$$

Since the power P at any moment is related to the number of photons ϕ by $P = \phi h\nu/t_c$, it follows from Equation (6.9-8) that the maximum number of stored photons inside the resonator is $n_i/2$. This can be explained by the fact that if $n_i \gg n_t$, the buildup of the pulse to its peak value occurs in a time short compared to t_c so that at the peak of the pulse, when $n = n_t$, most of the photons that were generated by stimulated emission are still present in the resonator. Moreover, since $n_i \gg n_t$, the number of these photons $(n_i - n_t)/2$ is very nearly $n_i/2$.

A typical numerical solution of Equations (6.9-2) and (6.9-3) is given in Figure 6.33.

To initiate the pulse we need, according to Equations (6.9-2) and (6.9-3), to have $\phi_i \neq 0$. Otherwise the solution is trivial ($\phi = 0$, $n = n_i$). The appropriate value of ϕ_i is usually estimated on the basis of the number of spontaneously emitted photons within the acceptance solid angle of the laser mode at $t = 0$. We also notice, as discussed earlier, that the photon density—hence the power—reaches a peak when $n = n_t$. The energy stored in the cavity ($\propto \phi$) at this point is maximum, so stimulated transitions from the upper to the lower laser levels continue to reduce the inversion to a final value $n_f < n_t$.

Numerical solutions of Equations (6.9-2) and (6.9-3) corresponding to different initial inversions n_i/n_t are shown in Figure 6.34. We notice that for $n_i \gg n_t$ the rise time becomes short compared to t_c but the fall time approaches a value nearly equal to t_c. The reason is that the process of stimulated emission is essentially over at the peak of the pulse ($\tau = 0$) and the observed output is due to the free decay of the photons in the resonator.

In Figure 6.35, we show an actual oscilloscope trace of a giant pulse. Giant laser pulses are used extensively in applications that require high peak powers and short duration. These applications include experiments in nonlinear optics, ranging, material machining and drilling, initiation of chemical reactions, and plasma diagnostics.

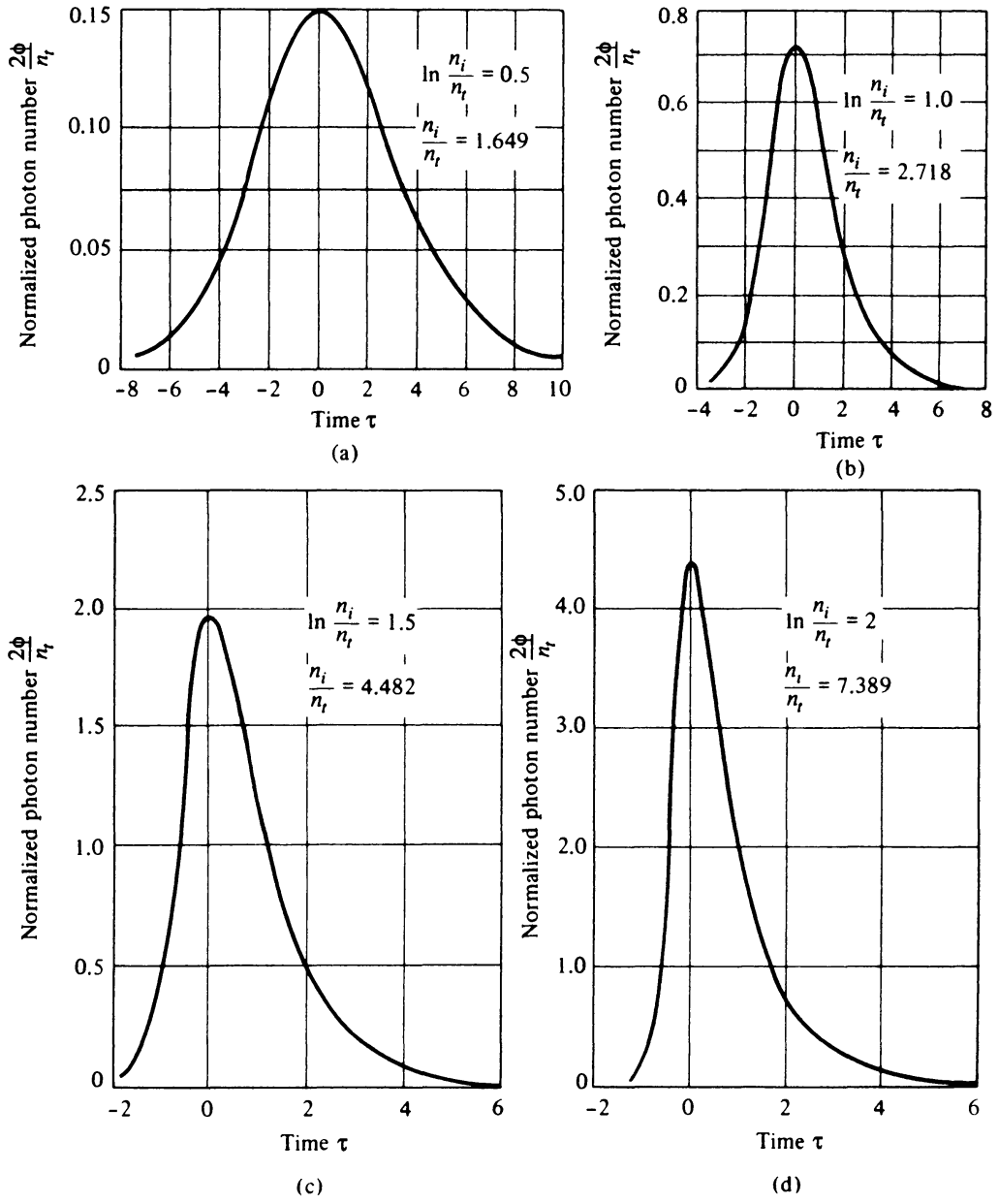


Figure 6.34 Photon number versus time in central region of giant pulse. Time is measured in units of photon lifetime. (After Reference [30].)

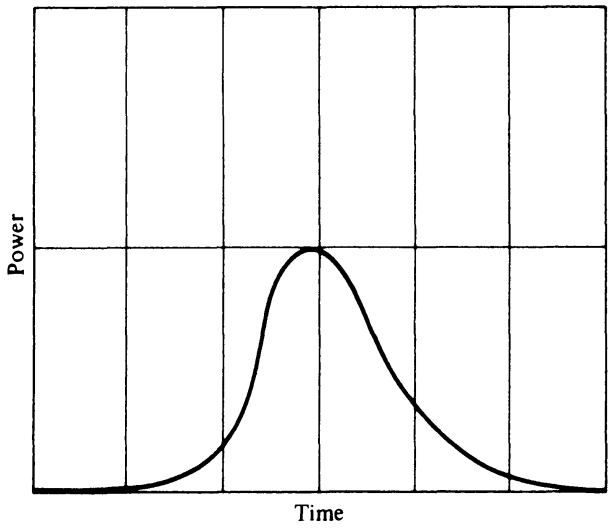


Figure 6.35 An oscilloscope trace of the intensity of a giant pulse. Time scale is 20 ns per division.

EXAMPLE: GIANT PULSE RUBY LASER

Consider the case of pink ruby with a chromium ion density of $N = 1.58 \times 10^{19}$ cm^{-3}. Its absorption coefficient is taken from Figure 6.42, where it corresponds to that of the R_1 line at 6943 Å, and is $\alpha \simeq 0.2$ cm^{-1} (at 300 K). Other assumed characteristics are:

$$l = \text{length of ruby rod} = 10 \text{ cm}$$

$$A = \text{cross-sectional area of mode} = 1 \text{ cm}^2$$

$$(1 - R) = \text{fractional intensity loss per pass} = 20\%$$

$$n_0 = 1.78$$

Since, according to Equation (5.6-19), the exponential loss coefficient is proportional to $N_1 - N_2$, we have

$$\alpha(\text{cm}^{-1}) = 0.2 \frac{N_1 - N_2}{1.58 \times 10^{19}} \tag{6.9-9}$$

Thus, at room temperature, where $N_2 \ll N_1$ when $N_1 - N_2 \cong 1.58 \times 10^{19}$ cm^{-3}, Equation (6.9-9) yields $\alpha = 0.2$ cm^{-1} as observed. The expression for the gain coefficient follows directly from Equation (6.9-9):

$$\gamma(\text{cm}^{-1}) = 0.2 \frac{N_2 - N_1}{1.58 \times 10^9}$$

$$= 0.2 \frac{n}{1.58 \times 10^9 V} \tag{6.9-10}$$

where n is the total inversion and $V = AL$ is the crystal volume in cm^3.

Threshold is achieved when the net gain per pass is unity. This happens when

$$e^{\gamma_t l} R = 1 \quad \text{or} \quad \gamma_t = -\frac{1}{l} \ln R \tag{6.9-11}$$

where the subscript t indicates the threshold condition.

Using Equation (6.9-10) in the threshold condition (6.9-11) plus the appropriate data from above gives

$$n_t = 1.8 \times 10^{19} \tag{6.9-12}$$

Assuming that the initial inversion is $n_i = 5n_t = 9 \times 10^{19}$, we find from Equation (6.9-8) that the peak power is approximately

$$P_p = \frac{n_i h\nu}{2t_c} = 5.1 \times 10^9 \text{ W} \tag{6.9-13}$$

where $t_c = n_0 l/c(1 - R) \simeq 2.5 \times 10^{-9}$ s.

The total pulse energy is

$$\mathscr{E} \sim \frac{n_i h\nu}{2} \sim 13 \text{ J}$$

while the pulse duration (see Figure 6.34) $\simeq 3t_c \simeq 7.5 \times 10^{-9}$ s.

Methods of *Q*-Switching

Some of the schemes used in Q-switching are:

1. Mounting one of the two end reflectors on a rotating shaft so that the optical losses are extremely high except for the brief interval in each rotation cycle in which the mirrors are nearly parallel.
2. The inclusion of a saturable absorber (bleachable dye) in the optical resonator; see References [10, 11, 34]. The absorber whose opacity decreases (saturates) with increasing optical intensity prevents rapid inversion depletion due to buildup of oscillation by presenting a high loss to the early stages of oscillation during which the slowly increasing intensity is not high enough to saturate the absorption. As the intensity increases the loss decreases, and the effect is similar, but not as abrupt, as that of a sudden increase of Q.
3. The use of an electro-optic crystal (or liquid Kerr cell) as a voltage-controlled gate inside the optical resonator. It provides a more precise control over the losses (Q) than schemes 1 and 2. Its operation is illustrated by Figure 6.36 and is discussed in some detail in the following. The control of the phase delay in the electro-optic crystal by the applied voltage is discussed in detail in Chapter 9.

During the pumping of the laser by the light from a flashlamp, a voltage is applied to the electro-optic crystal of such magnitude as to introduce a $\pi/2$ relative phase shift (retardation) between the two mutually orthogonal components (x' and y') that make up the linearly polarized (x) laser field. On exiting from the electro-optic crystal at point f, the light traveling to the right is circularly polarized. After reflection from the right mirror, the light passes once more through the crystal. The additional retardation of $\pi/2$ adds to the earlier one to give a total retardation of π, thus causing the emerging beam at d to be linearly polarized a long y and consequently to be blocked by the polarizer.

It follows that with the voltage on, the losses are high, so oscillation is prevented. The Q-switching is timed to coincide with the point at which the inversion reaches its peak and is achieved by a removal of the voltage applied to the electro-optic crystal. This reduces the retardation to zero so that the state of polarization of the wave passing through the crystal is unaffected and the Q regains its high value associated with the ordinary losses of the system.

6.10 HOLE BURNING AND THE LAMB DIP IN DOPPLER-BROADENED GAS LASERS

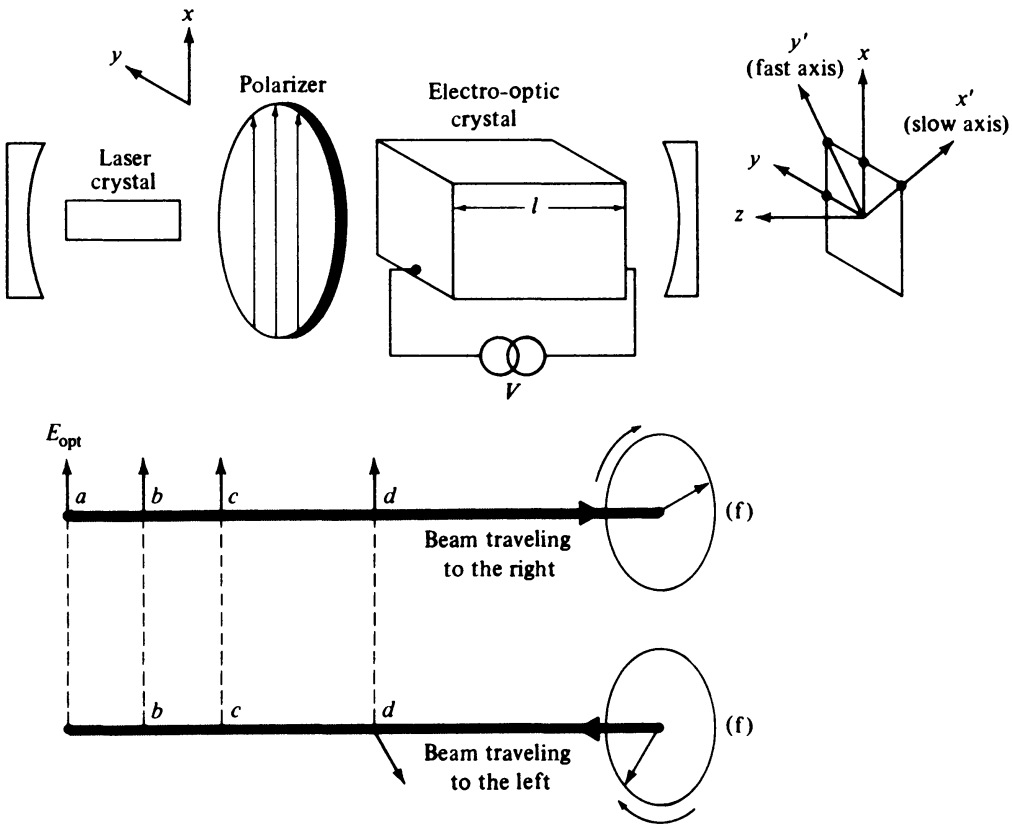
In this section we concern ourselves with some of the consequences of Doppler broadening in low-pressure gas lasers.

Consider an atom with a transition frequency $\nu_0 = (E_2 - E_1)/h$, where 2 and 1 refer to the upper and lower laser levels, respectively. Let the component of the velocity of the a tom parallel to the wave propagation direction be v. This component, thus, has the value

$$v = \frac{\mathbf{v}_{\text{atom}} \cdot \mathbf{k}}{k} \tag{6.10-1}$$

where the electromagnetic wave is described by

$$\mathbf{E} = \mathbf{E}e^{i(2\pi\nu t - \mathbf{k}\cdot\mathbf{r})} \tag{6.10-2}$$



For beam traveling to right:

At point *d*,

$$E_{x'} = \frac{E}{\sqrt{2}} \cos \omega t$$
$$E_{y'} = \frac{E}{\sqrt{2}} \cos \omega t$$

The optical field is linearly polarized with its electric field vector parallel to x

At point *f*,

$$E_{x'} = \frac{E}{\sqrt{2}} \cos\left(\omega t - kl - \frac{\pi}{2}\right)$$
$$E_{y'} = \frac{E}{\sqrt{2}} \cos(\omega t - kl)$$

Circularly polarized

For beam traveling to left:

At point *f*,

$$E_{x'} = -\frac{E}{\sqrt{2}} \cos\left(\omega t - kl - \frac{\pi}{2}\right)$$
$$E_{y'} = -\frac{E}{\sqrt{2}} \cos(\omega t - kl)$$

Circularly polarized

At point *d*,

$$E_{x'} = -\frac{E}{\sqrt{2}} \cos(\omega t - 2kl - \pi)$$
$$E_{y'} = -\frac{E}{\sqrt{2}} \cos(\omega t - 2kl)$$

Linearly polarized along y

Figure 6.36 Electro-optic crystal used as voltage-controlled gate in Q-switching a laser.

An atom moving with a constant velocity $\mathbf{v}$, so that $\mathbf{r} = \mathbf{v}t + \mathbf{r}_0$, will experience a field

$$\begin{aligned} \mathbf{E}_{\text{atom}} &= \mathbf{E} e^{i[2\pi\nu t - \mathbf{k}\cdot(\mathbf{r}_0 + \mathbf{v}t)]} \\ &= \mathbf{E} e^{i[(2\pi\nu - \mathbf{v}\cdot\mathbf{k})t - \mathbf{k}\cdot\mathbf{r}_0]} \end{aligned} \tag{6.10-3}$$

and will thus "see" a Doppler-shifted frequency

$$\nu_D = \nu - \frac{\mathbf{v}\cdot\mathbf{k}}{2\pi} = \nu - \frac{v}{c}\nu \tag{6.10-4}$$

where in the second equality we took $n = 1$ so that $k = 2\pi\nu/c$ and used Equation (6.10-1).

The condition for the maximum strength of interaction (i.e., emission or absorption) between the moving atom and the wave is that the apparent (Doppler) frequency ν_D "seen" by the atom be equal to the atomic resonant frequency ν_0

$$\nu_0 = \nu - \frac{v}{c}\nu \tag{6.10-5}$$

or reversing the argument, a wave of frequency ν moving through an ensemble of atoms will "seek out" and interact most strongly with those atoms whose velocity component v satisfies

$$\nu = \frac{\nu_0}{1 - \frac{v}{c}} \approx \nu_0\left(1 + \frac{v}{c}\right) \tag{6.10-6}$$

where the approximation is valid for $v \ll c$.

Now consider a gas laser oscillating at a single frequency ν, where, for the sake of definiteness, we take $\nu > \nu_0$. The standing-wave electromagnetic field at ν inside the laser resonator consists of two waves traveling in opposite directions. Consider, first, the wave traveling in the positive x direction (the resonator axis is taken parallel to the axis). Since $\nu > \nu_0$, the wave interacts, according to Equation (6.10-6), with atoms having $v > 0$, that is, atoms with

$$v_x = +\frac{c}{\nu}(\nu - \nu_0) \tag{6.10-7}$$

The wave traveling in the opposite direction ($-x$) must also interact with atoms moving in the same direction so that the Doppler-shifted frequency is reduced from ν to ν_0. These are atoms with

$$v_x = -\frac{c}{\nu}(\nu - \nu_0) \tag{6.10-8}$$

We conclude that due to the standing-wave nature of the field inside a conventional two-mirror laser oscillator, a given frequency of oscillation interacts with two velocity classes of atoms.

Consider, next, a four-level gas laser oscillating at a frequency $\nu > \nu_0$. At negligibly low levels of oscillation and at low gas pressure, the velocity distribution function of atoms in the upper laser level is given, according to Equation (5.5-11), by

$$f(v_x) \propto e^{-Mv_x^2/2kT} \tag{6.10-9}$$

where $f(v_x)\, dv_x$ is proportional to the number of atoms (in the upper laser level) with the x component of velocity between v_x and $v_x + dv_x$. As the oscillation level is increased, say, by reducing the laser losses, we expect the number of atoms in the upper laser level, with x velocities near $v_x = \pm(c/\nu)(\nu - \nu_0)$, to decrease from their equilibrium value as given by Equation (6.10-9). This is due to the fact that these atoms undergo stimulated downward transitions from level 2 to 1, thus reducing the number of atoms in level 2. The velocity distribution function under conditions of oscillation consequently has two depressions as shown schematically in Figure 6.37.

If the oscillation frequency ν is equal to ν_0, only a single "hole" exists in the velocity distribution function of the inverted atoms. This hole is centered on $v_x = 0$. We may thus expect the power output of a laser oscillating at $\nu = \nu_0$ to be less than that of a laser in which ν is tuned slightly to one side or the other of ν_0 (this tuning can be achieved by moving one of the laser mirrors). This power dip first predicted by Lamb [35] is indeed observed in gas lasers [36]. An experimental plot of the power versus frequency in a He–Ne 1.15 μm laser

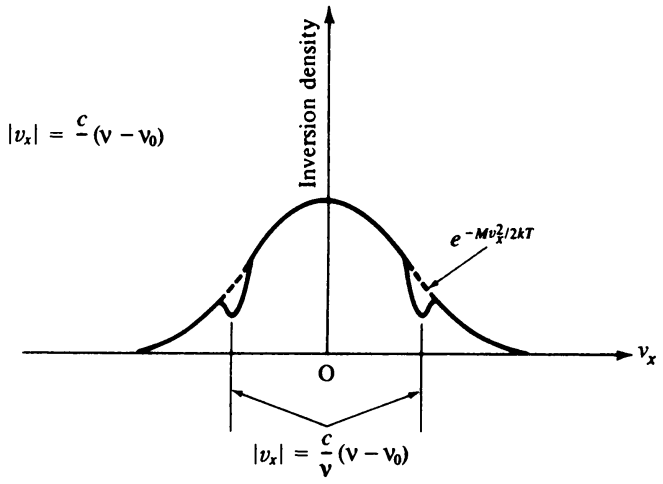


Figure 6.37 The distribution of inverted atoms as a function of v_x. The dashed curve that is proportional to $\exp(-Mv_x^2/2kT)$ corresponds to the case of zero field intensity. The solid curve corresponds to a standing-wave field at $\nu = \nu_0/(1 - v_x/c)$ or one at $\nu = \nu_0/(1 + v_x/c)$.

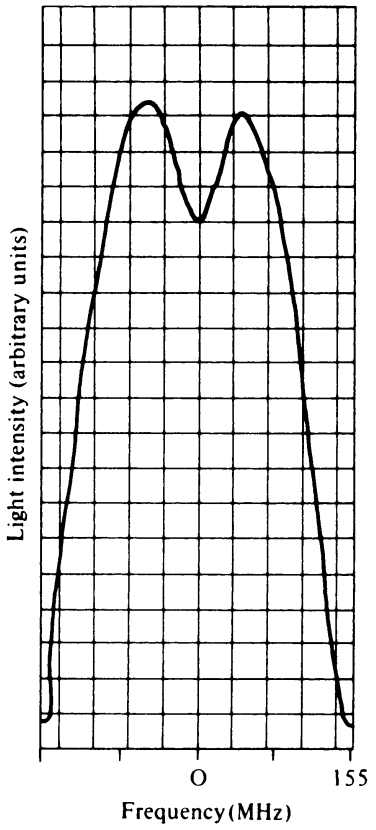


Figure 6.38 The power output as a function of the frequency of a single-mode 1.15 μm He–Ne laser using the ^{20}Ne isotope. (After Reference [36].)

is shown in Figure 6.38. The phenomenon is referred to as the "Lamb dip" and is used in frequency stabilization schemes of gas lasers [37].

6.11 SOME SPECIFIC LASER SYSTEMS

In this section, we discuss some specific laser systems. The pumping of the atoms into the upper laser level is accomplished in a variety of ways, depending on the type of laser. Here we will review some of the more common laser systems and in the process describe their pumping mechanisms. The laser systems described include Cr^{3+}:Al_2O_3 (ruby), Nd^{3+}:YAG, Nd^{3+}:glass, and Er^{3+}:silica lasers. The semiconductor current-pumped (or carrier injection) laser, because

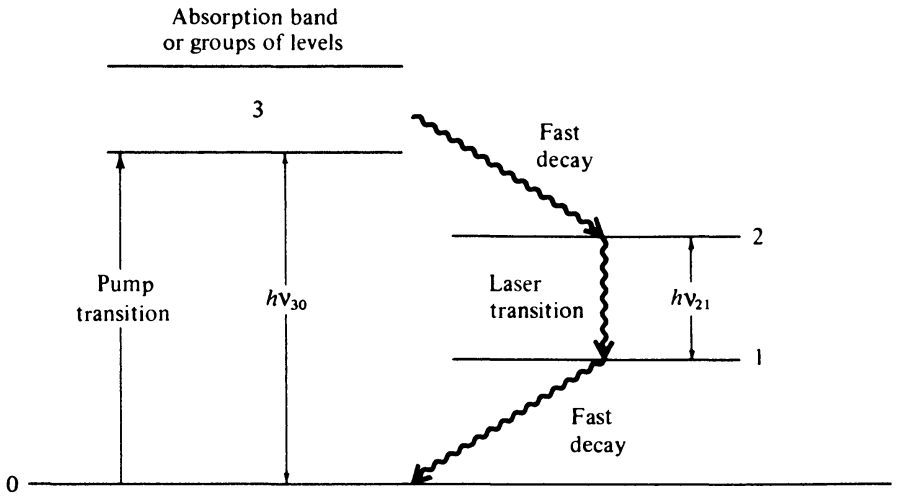


Figure 6.39 Pumping–oscillation cycle of a typical laser.

of its unique technological importance, receives special billing in Chapters 15 and 16. Figure 6.39 shows the pumping–oscillation cycle of a (hypothetical) representative laser. The pumping agent elevates the atoms into some excited state 3 from which they relax into the upper laser level 2. The stimulated laser transition takes place between levels 2 and 1 and results in the emission of a photon of frequency ν_{21}. It is evident from this figure that the minimum energy input per output photon is $h\nu_{30}$, so the power efficiency of the laser cannot exceed

$$\eta_{\text{atomic}} = \frac{\nu_{21}}{\nu_{30}} \tag{6.11-1}$$

referred to as the "atomic quantum efficiency." The overall laser efficiency depends on the fraction of the total pump power that is effective in transferring atoms into level 3 and on the pumping quantum efficiency defined as the fraction of the atoms that once in level 3, make a transition to level 2. The product of the last two factors, which constitutes an upper limit on the efficiency of optically pumped lasers, ranges from about 1% for solid-state lasers such as Nd^{3+}:YAG, to about 30% in the CO_2 laser, and to near unity in the GaAs junction laser. We shall discuss these factors when we get down to some specific laser systems. We may note, however, that according to Equation (6.11-1), in an efficient laser system ν_{21} and ν_{30} must be of the same order of magnitude, so the laser transition should involve low-lying levels.

The laser examples discussed in this section consist of an optically transparent host material such as glass (SiO_2) or sapphire crystal (Al_2O_3) doped with a small amount of Cr, Nd, or Er ions. In a glass or crystal host, the quantum states of these active atoms (ionized) are modified by the local electric field that causes Stark splitting and shifts. For the purpose of discussion, we briefly review the spectroscopic notation used for the quantum states. For atoms with many electrons, the quantum state is characterized by a set L, S, J of quantum numbers. They are total orbital angular momentum, total spin angular momentum, and total angular momentum of the electrons, respectively. The state is designated as

$$^{2S+1}L_J$$

where $2S + 1$, the multiplicity, denotes the total number of different J-values associated with a given set of (L, S), assuming $S \leq L$, which is normally the case. The total angular momentum is the vector sum of the total orbital angular momentum and the total spin angular momentum. Thus the quantum number J is often limited within $L - S$ and $L + S$. The letter

TABLE 6.3 Electronic State Designations

Letter	L =
S	0
P	1
D	2
F	3
G	4
H	5
I	6
⋮	⋮

symbol L, denoting the orbital angular momentum, is designated as shown in Table 6.3. For each state $^{2S+1}L_J$, there are $2J + 1$ mutually orthogonal sublevels with the same energy. An external magnetic field can cause the splitting of the state $^{2S+1}L_J$ into $2J + 1$ energy levels. For example, the ground state of Er^{3+} ion is designated as $^4I_{15/2}$, which means $S = 3/2$, $L = 6$, and $J = 15/2$. The multiplicity of $2S + 1 = 4$ indicates that the multiplet consists of four states with the same L and S. These four states are $^4I_{9/2}$, $^4I_{11/2}$, $^4I_{13/2}$, and $^4I_{15/2}$. These states within the same multiplet have different total angular momentum J and have different energies due to the so-called spin–orbit interaction.

Ruby Laser

The first material in which laser action was demonstrated [17] and still one of the most useful laser materials is ruby, whose output is at $\lambda_0 = 0.6943$ μm. The active laser particles are Cr^{3+} ions present as impurities in Al_2O_3 crystal. Typical Cr^{3+} concentrations are ~0.05% by weight. The triply ionized Cr ion has 3 electrons in the 3d state. These unpaired d electrons form a high-spin ground state referred to as ^4F ($S = 3/2$). The first excited state of the free ion is a ^2G state ($S = 1/2$) about 2 eV higher in energy than the ground state. Both of the F and G levels have high orbital degeneracy. There are seven orbital sublevels of the F level ($L = 3$) and nine sublevels of the G level ($L = 4$). Under the action of crystalline electric fields in the sapphire crystal (Al_2O_3), the ^4F state splits into three states ($^4A_2$, $^4T_2$, $^4T_1$), and the ^2G state splits into four levels (^2E, $^2T_1$, $^2T_2$, $^2A_1$). As a result, the ground state becomes $^4A_2$. In these notations, the A level is singly degenerate, the E level is doubly degenerate, and the T level is triply degenerate. The pertinent energy level diagram is shown in Figure 6.40.

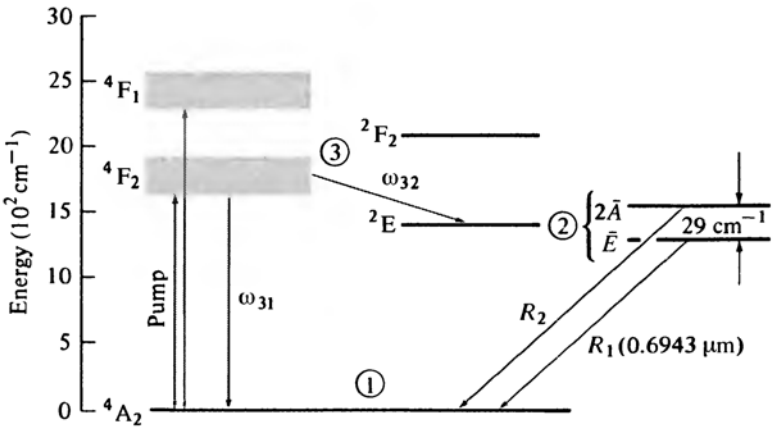


Figure 6.40 Energy levels pertinent to the operation of a ruby laser. The unit 1 cm^{-1} corresponds to $\nu = 30$ GHz, or $h\nu = 1.24 \times 10^{-4}$ eV. (After Reference [38].)

The pumping of ruby is usually performed by subjecting it to the light of intense flashlamps (quite similar to the types used in flash photography). A portion of this light that corresponds in frequency to the two absorption bands $^4F_2$ ($^4T_2$) and $^4F_1$ ($^4T_1$) is absorbed, thereby causing Cr^{3+} ions to be transferred into these levels. The ions proceed to decay, within an average time of $\omega_{32}^{-1} \cong 5 \times 10^{-8}$ s [38], into the upper laser level ^2E. Level ^2E is composed of two separate levels $2\bar{A}$ and $\bar{E}$ separated by 29 cm^{-1} (due to spin–orbit interaction, etc.) The unit 1 cm^{-1} (one wavenumber) is equivalent to $\nu = 30$ GHz, or $h\nu = 1.24 \times 10^{-4}$ eV. It is also used as a measure of energy where 1 cm^{-1} corresponds to the energy $h\nu$ of a photon with $\nu = 30$ GHz. The lower of these two, $\bar{E}$, is the upper laser level. The lower laser level is the ground state and thus, according to the discussion earlier in this section, a ruby laser is a three-level laser. The lifetime of atoms in the upper laser level $\bar{E}$ is $t_2 = 3 \times 10^{-3}$ s. Each decay results in the (spontaneous) emission of a photon, so $t_2 = t_{spont}$.

An absorption spectrum of a typical ruby with two orientations of the optical field relative to the c (optic) axis is shown in Figure 6.41. The two main peaks correspond to absorption into the useful $^4F_1$ and $^4F_2$ bands, which are responsible for the characteristic (ruby) color. The ordinate is labeled in terms of the absorption coefficient and in terms of the transition cross section σ, which may be defined as the absorption coefficient per unit inversion per unit volume and has consequently the dimension of area. According to this definition, the linear absorption coefficient $\alpha(\nu)$ is given by

$$\alpha(\nu) = (N_1 - N_2)\sigma(\nu) \tag{6.11-2}$$

A more detailed plot of the absorption near the laser emission wavelength is shown in Figure 6.42. The width $\Delta\nu$ of the laser transition as a function of temperature is shown in Figure 6.43. At room temperature, $\Delta\nu = 11$ cm^{-1} (or 3.3×10^{11} Hz). We can use ruby to illustrate some of the considerations involved in optical pumping of solid-state lasers. Figure 6.44 shows a typical setup of an optically pumped laser, such as ruby. The helical flashlamp

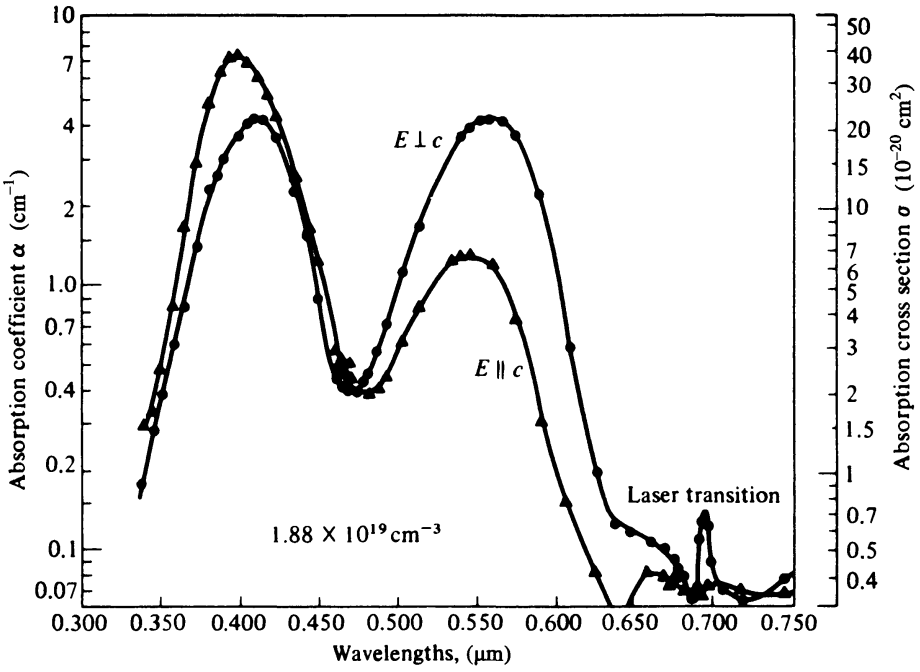


Figure 6.41 Absorption coefficient and absorption cross section as functions of wavelength for $E \parallel c$ and $E \perp c$. The 300 K data were derived from transmittance measurements on pink ruby with an average Cr ion concentration of 1.88×10^{19} cm^{-3}. (After Reference [39].)

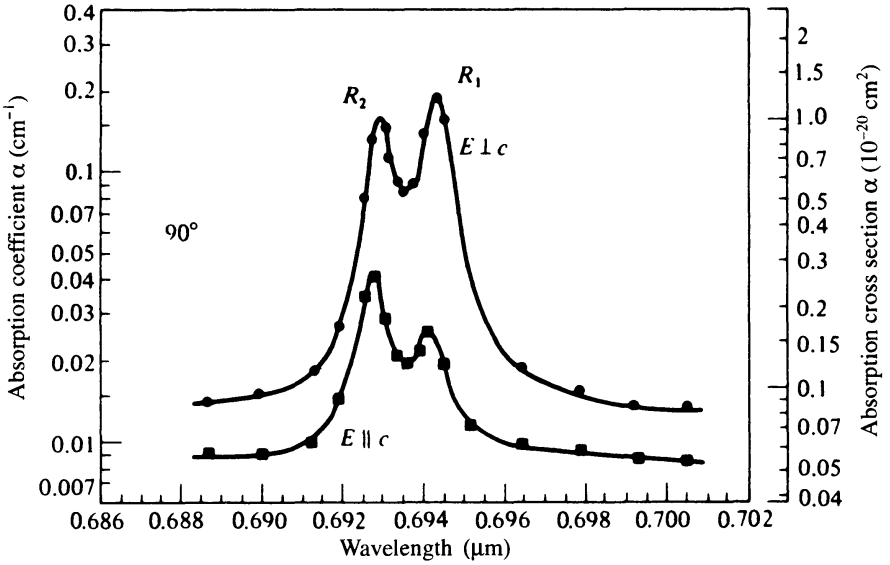


Figure 6.42 Absorption coefficient and absorption cross section as functions of wavelength for $E \parallel c$ and $E \perp c$. Sample was a pink ruby laser rod having a 90° c-axis orientation with respect to the rod axis and a Cr ion concentration of 1.58×10^{19} cm^{-3}. (After Reference [39].)

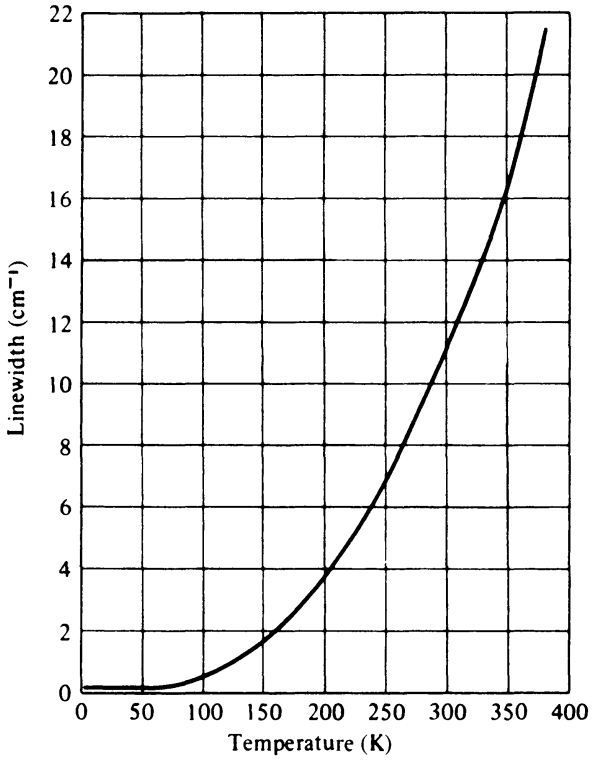


Figure 6.43 Linewidth of the R_1 line of ruby as a function of temperature. (After Reference [40].)

surrounds the ruby rod. The flash excitation is provided by the discharge of the charge stored in a capacitor bank across the lamp.

The typical flash output consists of a pulse of light of duration $t_{\text{flash}} = 5 \times 10^{-4}$ s. Let us, for the sake of simplicity, assume that the flash pulse is rectangular in time and of duration t_{flash} and that it results in an optical flux at the crystal surface having $s(\nu)$ watts per unit area per unit frequency at the frequency ν. If the absorption coefficient of the crystal is $\alpha(\nu)$, then the amount of energy absorbed by the crystal per unit volume is

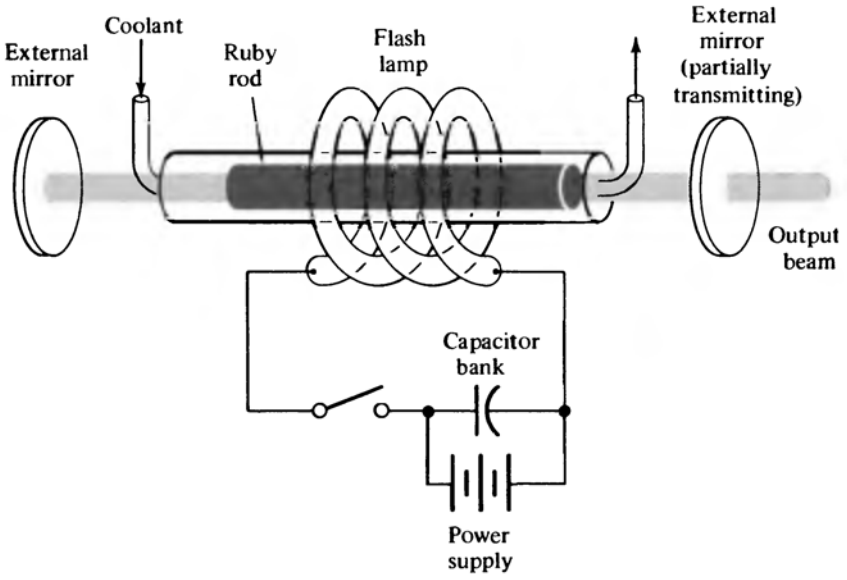


Figure 6.44 Typical setup of a pulsed ruby laser using flashlamp pumping and external mirrors.

$$t_{\text{flash}} \int_0^\infty S(\nu)\alpha(\nu)\, d\nu$$

where we assume that the total absorption in passing the crystal is small, so $s(\nu)$ is taken to be independent of the distance through the crystal.

If the absorption quantum efficiency (the probability that the absorption of a pump photon at ν results in transferring one atom into the upper laser level) is $\eta(\nu)$, the number of atoms pumped into level 2 per unit volume is

$$N_2 = t_{\text{flash}} \int_0^\infty \frac{s(\nu)\alpha(\nu)\eta(\nu)}{h\nu}\, d\nu \tag{6.11-3}$$

Since the lifetime $t_2 = 3 \times 10^{-3}$ s of atoms in level 2 is considerably longer than the flash duration ($\sim 5 \times 10^{-4}$ s), we may neglect the spontaneous decay out of level 2 during the time of the flash pulse, so N_2 represents the population of level 2 after the flash.

EXAMPLE: FLASH PUMPING OF A PULSED RUBY LASER

Consider the case of a ruby laser with the following parameters:

$$N_0 = 2 \times 10^{19} \text{ atoms/cm}^3 \ (\text{Cr}^{3+} \text{ concentration})$$

$$t_2 = t_{\text{spont}} = 3 \times 10^{-3} \text{ s}$$

$$t_{\text{flash}} = 5 \times 10^{-4} \text{ s}$$

If the useful absorption is limited to relatively narrow spectral regions, we may approximate Equation (6.11-3) by

$$N_2 = \frac{t_{\text{flash}}\, \overline{s(\nu)}\ \overline{\alpha(\nu)}\ \overline{\eta(\nu)}}{h\bar{\nu}}\, \overline{\Delta\nu} \tag{6.11-4}$$

where the bars represent average values over the useful absorption region whose width is $\overline{\Delta\nu}$.

From Figure 6.41, we deduce an average absorption coefficient of $\overline{\alpha(\nu)} \cong 2\ \text{cm}^{-1}$ over the two central peaks. Since ruby is a three-level laser, the upper level population is, according to Equation (6.3-2), $N_2 \sim N_0/2 = 10^{19}\ \text{cm}^{-3}$. Using $\bar{\nu} \cong 5 \times 10^{14}$ Hz and $\overline{\eta(\nu)} \cong 1$, Equation (6.11-4) yields

$$\bar{s}\,\overline{\Delta\nu}\,t_{\text{flash}} \cong 1.5\ \text{J/cm}^2$$

for the pump energy in the useful absorption region that must fall on each square centimeter of crystal surface in order to obtain threshold inversion. To calculate the total lamp energy that is incident on the crystal, we need to know the spectral characteristics of the lamp output. Typical data of this sort are shown in Figure 6.45. The mercury-discharge lamp is seen to contain considerable output in the useful absorption regions (near 4000 and 5500 Å) of ruby. If we estimate the useful fraction of the lamp output at 10%, the fraction of the lamp light actually incident on the crystal as 20%, and the conversion of electrical-to-optical energy as 50%, we find that the threshold electric energy input to the flashlamp per square centimeter of laser surface is

$$\frac{1.5}{0.1 \times 0.2 \times 0.5} = 150\ \text{J/cm}^2$$

These are, admittedly, extremely crude calculations. They are included not only to illustrate the order of magnitude numbers involved in laser pumping, but also as an example of the quick and rough estimates needed to discriminate between feasible ideas and "pie-in-the-sky" schemes.

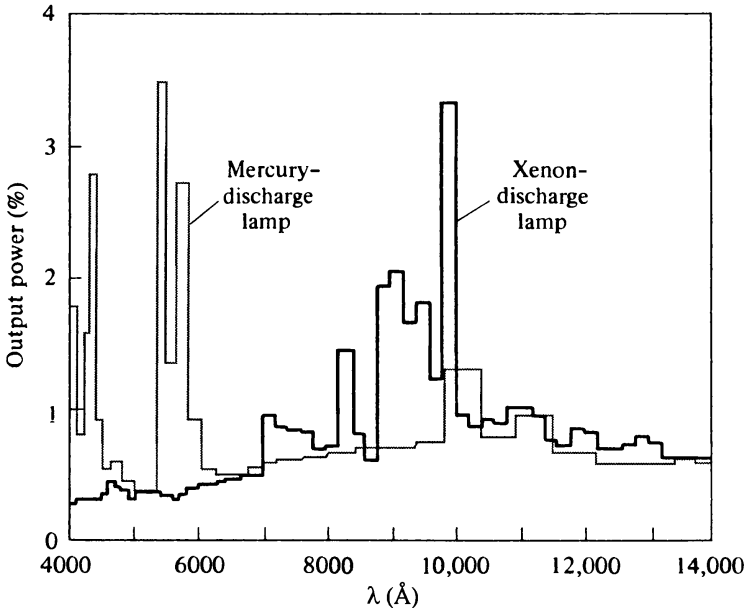


Figure 6.45 Spectral output characteristics of two commercial high-pressure lamps. Output is plotted as a fraction of electrical input to lamp over certain wavelength intervals (mostly 200 Å) between 0.4 and 1.4 μm. (After Reference [41].)

Nd^{3+}:YAG Laser

One of the most important laser systems is that using trivalent neodymium ions (Nd^{3+}), which are present as impurities in yttrium aluminum garnet (YAG = $Y_3Al_5O_{12}$); see References [42, 43]. The laser emission occurs at $\lambda_0 = 1.0641$ μm at room temperature. Figure 6.46 illustrates the relevant energy levels of the neodymium ions (Nd^{3+}) in the host material. The lower

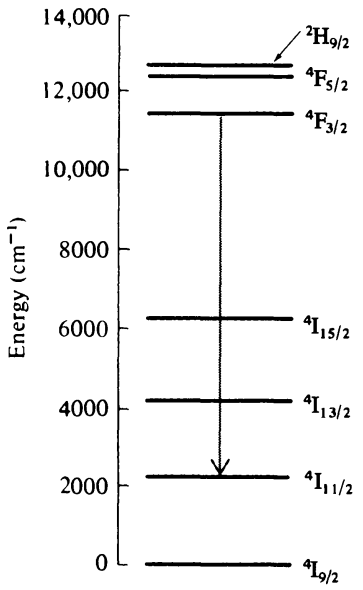


Figure 6.46 Energy-level diagram of Nd^{3+} in YAG. (After Reference [42].) The unit 1 cm^{-1} corresponds to $\nu = 30$ GHz, or $h\nu = 1.24 \times 10^{-4}$ eV.

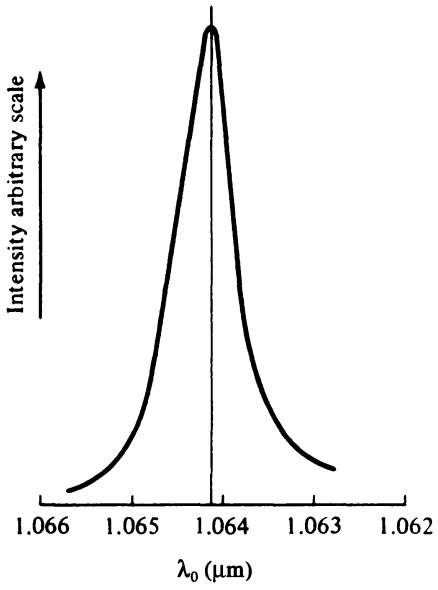


Figure 6.47 Spontaneous emission spectrum of Nd^{3+} in YAG near the laser transition at $\lambda_0 = 1.064$ μm. (After Reference [43].)

laser level is at $E_2 = 2111$ cm^{-1} (0.26 eV) from the ground state so that at room temperature its population is down by a factor of $\exp(-E_2/kT) = e^{-10}$ from that of the ground state and can be neglected. The Nd^{3+}:YAG thus fits our definition of a four-level laser.

The spontaneous emission spectrum of the laser transition is shown in Figure 6.47. The width of the gain linewidth at room temperature is $\Delta\nu = 6$ cm^{-1} (or 180 GHz). The spontaneous lifetime for the laser transition has been measured [43] as $t_{spont} = 5.5 \times 10^{-4}$ s. The room-temperature cross section at the center of the laser transition is $\sigma = 9 \times 10^{-19}$ cm^2. If we compare this number to $\sigma = 1.22 \times 10^{-20}$ cm^2 in ruby (see Figure 6.42), we expect that at a given inversion the optical gain constant γ in Nd^{3+}:YAG is approximately 75 times that of ruby. This causes the oscillation threshold to be very low and explains the easy continuous (CW) operation of this laser compared to the ruby laser.

The absorption responsible for populating the upper laser level takes place in a number of bands between 13,000 and 25,000 cm^{-1}. Here, we recall the unit 1 cm^{-1} corresponds to $\nu = 30$ GHz, or $h\nu = 1.24 \times 10^{-4}$ eV.

EXAMPLE: THRESHOLD OF AN Nd^{3+}:YAG LASER

Pulsed Threshold

First we estimate the energy needed to excite a typical Nd^{3+}:YAG laser on a pulse basis so that we can compare it with that of the ruby laser. We use the following data:

$$\left.\begin{array}{l} l = 20 \text{ cm (length optical resonator)} \\ \text{L} = 4\% \\ n = 1.8 \end{array}\right\} t_c = \frac{nl}{Lc} = 2.5 \times 10^{-8} \text{ s}$$

$\Delta\nu = 6 \text{ cm}^{-1}$ ($= 6 \times 3 \times 10^{10}$ Hz) (Note: 1 cm^{-1} = 30 GHz)

$t_{\text{spont}} = 5.5 \times 10^{-4}$ s

$\lambda = 1.06$ μm

Using the foregoing data in Equation (6.1-18) gives

$$N_t = \frac{8\pi n^3 t_{\text{spont}} \Delta\nu}{c t_c \lambda^2} \cong 1.7 \times 10^{15} \text{ cm}^{-3}$$

Assuming that 5% of the exciting light energy falls within the useful absorption bands, that 5% of this light is actually absorbed by the crystal, that the average ratio of laser frequency to the pump frequency is 0.5, and that the lamp efficiency (optical output/electrical input) is 0.5, we obtain

$$E_{\text{lamp}} = \frac{N_t h \nu_{\text{laser}}}{5 \times 10^{-2} \times 5 \times 10^{-2} \times 0.5 \times 0.5} \cong 0.3 \text{ J/cm}^3$$

for the energy input to the lamp at threshold. It is interesting to compare this last number to the figure of 150 joules per square centimeter of surface area obtained in the ruby example. For reasonable crystal dimensions (say, length = 5 cm, r = 2 mm), we obtain $E_{\text{lamp}} = 0.19$ J. We expect the ruby threshold to exceed that of Nd^{3+}:YAG by three orders of magnitude, which is indeed the case.

Continuous Operation

The critical fluorescence power—that is, the actual power given off by spontaneous emission just below threshold—is given by Equation (6.3-4) as

$$\left(\frac{P_s}{V}\right) = \frac{N_t h\nu}{t_{\text{spont}}} \cong 0.34 \text{ W/cm}^3$$

Taking the crystal diameter as 0.25 cm and its length as 3 cm and using the same efficiency factors assumed in the first part of this example, we can estimate the power input to the lamp at threshold as

$$P_{\text{lamp}} = \frac{0.34 \times (\pi/4) \times (0.25)^2 \times 3}{5 \times 10^{-2} \times 5 \times 10^{-2}\ 0.5 \times 0.5} \cong 81 \text{ W}$$

which is in reasonable agreement with experimental values [42]. A typical arrangement used in continuous solid-state lasers is shown in Figure 6.48. The highly polished elliptic cylinder is used to concentrate the light from the lamp, which is placed along one focal axis, onto the laser rod, which occupies the other axis. This configuration guarantees that most of the light emitted by the lamp passes through the laser rod. The reflecting mirrors (for the laser resonator) are placed outside the cylinder.

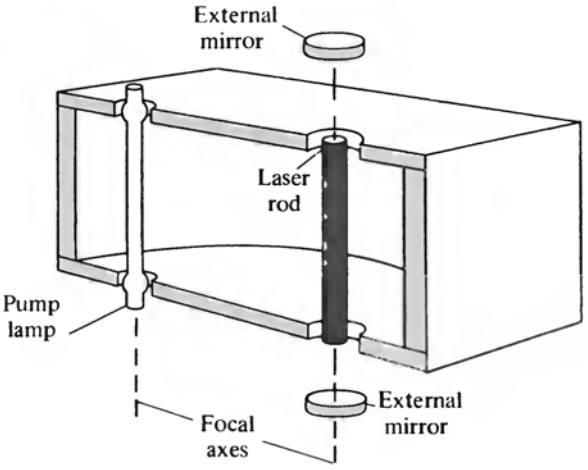


Figure 6.48 Typical continuous solid-state laser arrangement employing an elliptic cylinder housing for concentrating lamplight onto laser.

Neodymium–Glass Laser

One of the most useful laser systems is that which results when the Nd^{3+} ion is present as an impurity atom in glass [44]. The energy levels involved in the laser transition in a typical glass are shown in Figure 6.49. The laser emission wavelength is at $\lambda = 1.059\ \mu m$ and the lower level is approximately 1950 cm^{-1} (0.24 eV) above the ground state. As in the case of Nd^{3+}:YAG described above, we have here a four-level laser, since the thermal population of the lower laser level is negligible. The fluorescent emission near $\lambda_0 = 1.06\ \mu m$ is shown in Figure 6.50. The fluorescent linewidth can be measured off directly and ranges, for the glasses shown, around 300 cm^{-1} (9 THz). This width is approximately a factor of 50 larger than that of Nd^{3+} in YAG. This is due to the amorphous structure of glass, which causes different Nd^{3+} ions to "see" slightly different surroundings. This causes their energy splittings to vary slightly. Different ions consequently radiate at slightly different frequencies, causing a broadening of the spontaneous emission spectrum. This is a good example of inhomogeneous broadening. The absorption bands responsible for pumping the laser level are shown in Figure 6.51. The probability that the absorption of a photon in any of these bands will result in pumping an atom to the upper laser level (i.e., the absorption quantum efficiency) has been estimated [44] at about 0.4.

The lifetime t_2 of the upper laser level depends on the host glass and on the Nd^{3+} concentration. This variation in two glass series is shown in Figure 6.52.

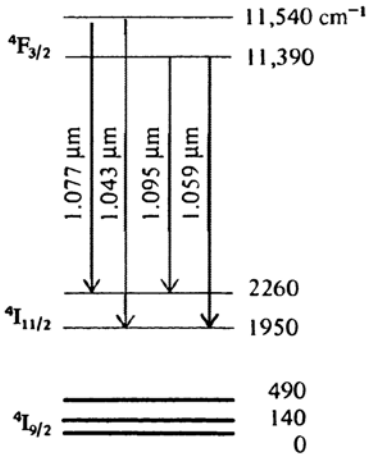


Figure 6.49 Energy-level diagram for the ground state and the states involved in laser emission at 1.059 μm for Nd^{3+} in a rubidium potassium barium silicate glass. The unit 1 cm^{-1} corresponds to $\nu = 30$ GHz, or $h\nu = 1.24 \times 10^{-4}$ eV. (After Reference [44].)

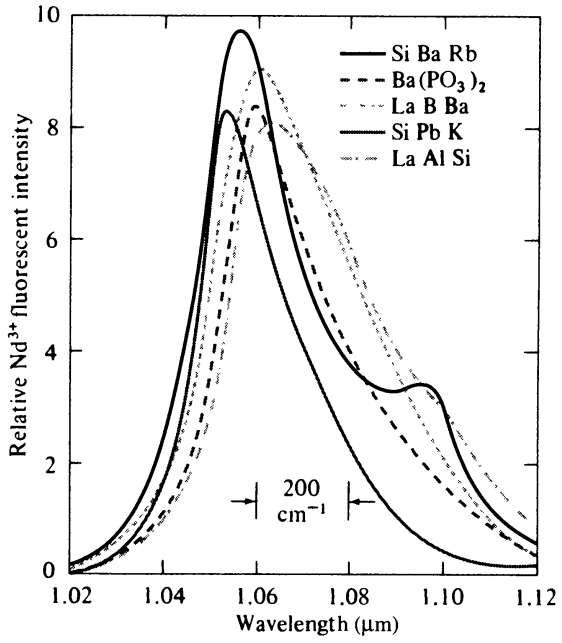


Figure 6.50 Fluorescent emission of the 1.06 μm line of Nd^{3+} at 300 K in various glass bases. (After Reference [44].)

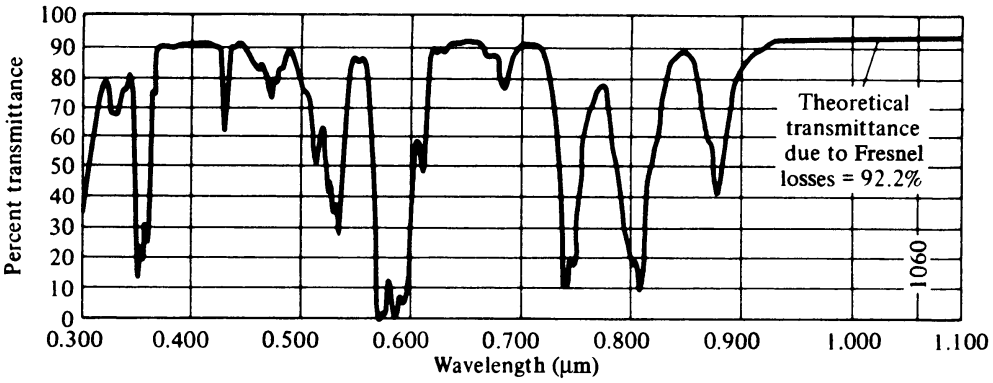


Figure 6.51 Nd^{3+} absorption spectrum for a sample of glass 6.4 mm thick with the composition 66 wt.% SiO_2, 5 wt.% Nd_2O_3, 16 wt.% Na_2O, 5 wt.% BaO, 2 wt.% Al_2O_3, and 1 wt.% Sb_2O_3. (After Reference [44].)

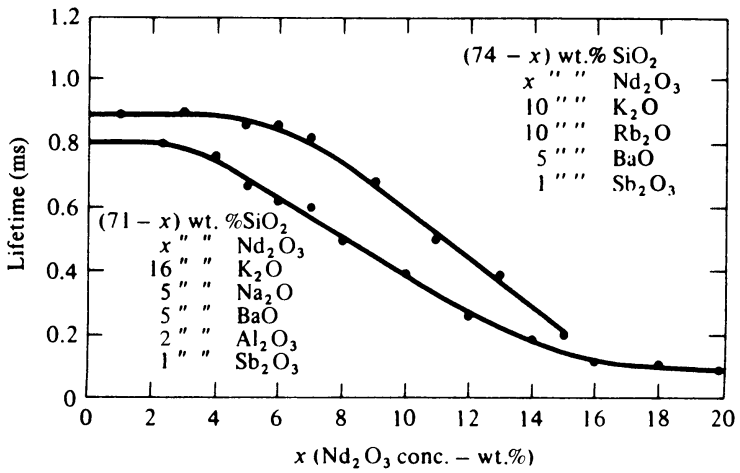


Figure 6.52 Lifetime as a function of concentration for two glass series. (After Reference [44].)

EXAMPLE: THRESHOLDS FOR CW AND PULSED OPERATION OF Nd^{3+}:GLASS LASERS

Let us estimate first the threshold for continuous (CW) laser action in a Nd^{3+}:glass laser using the following data:

$$\Delta\nu = 200 \text{ cm}^{-1} \text{ (or } \Delta\nu = 6 \text{ THz, see Figure 6.50)}$$

$$n = 1.5$$

$$t_{spont} \sim t_2 = 3 \times 10^{-4} \text{ s}$$

$$\left.\begin{array}{l} l = \text{length optical resonator} = 20 \text{ cm} \\ \text{L} = \text{loss per pass} = 4\% \end{array}\right\} t_c \cong \frac{nl}{Lc} = 5 \times 10^{-8} \text{ s}$$

Using Equation (6.1-18), we obtain

$$N_t = \frac{8\pi t_{spont} n^3 \Delta\nu}{c t_c \lambda^2} = 9.05 \times 10^{15} \text{ atoms/cm}^3$$

for the critical inversion. The fluorescence power at threshold P_s is thus (see Equation (6.3-5))

$$P_s = \frac{N_t h\nu V}{t_{spont}} = 5.65 \text{ watts}$$

in a crystal volume $V = 1 \text{ cm}^3$.

We assume (a) that only 10% of the pump light lies within the useful absorption bands, (b) that because of the optical coupling inefficiency and the relative transparency of the crystal only 10% of the energy leaving the lamp within the absorption bands is actually absorbed, (c) that the absorption quantum efficiency is 40%, and (d) that the average pumping frequency is twice that of the emitted radiation. The lamp output at threshold is thus

$$\frac{2 \times 5.65}{0.1 \times 0.1 \times 0.4} = 2825 \text{ W}$$

If the efficiency of the lamp in converting electrical to optical energy is about 50%, we find that continuous operation of the laser requires about 5 kW of power. This number is to be contrasted with a threshold of approximately 100 watts for the Nd:YAG laser, which helps explain why Nd:glass lasers are not operated continuously.

If we consider the pulsed operation of a Nd:glass laser by flash excitation, we have to estimate the minimum energy needed to pump the laser at threshold. Let us assume here that the losses (attributable mostly to the output mirror transmittance) are $L = 20\%$. (Because of the higher pumping rate available with flash pumping, optimum coupling in this case calls for larger mirror transmittances compared to the CW case.) A recalculation of N_t gives

$$N_t = 9.05 \times 10^{16} \text{ atoms/cm}^3$$

The minimum energy needed to pump N_t atoms into level 2 is then

$$\frac{E_{min}}{V} = N_t(h\nu) = 1.7 \times 10^{-2} \text{ J/cm}^3$$

Assuming a crystal volume $V = 10 \text{ cm}^3$ and the same efficiency factors used in the previous CW example, we find that the input energy to the flashlamp at thresholds $\sim 2 \times 1.7 \times 10^{-2} \times$

$10/(0.1 \times 0.1 \times 0.4) = 85$ J. Typical Nd^{3+}:glass lasers with characteristics similar to those used in this example are found to require an input of about 150–300 joules at threshold.

Er^{3+}:Silica Laser

In addition to neodymium atoms, many other rare-earth atoms can be employed as dopants in glass or crystals. These active elements can be pumped with external light sources to reach population inversion. As a result of their intrinsic atomic energy states, these rare-earth atoms cover various spectral regimes of interest. In the area of fiberoptic communications, we are particularly interested in the spectral regime around $\lambda = 1550$ nm. One of the most important laser systems is that of *Er-doped* silica fibers [45–50] at $\lambda = 1.54$–1.55 μm. Such fibers pumped at $\lambda = 0.98$ μm or $\lambda = 1.48$ μm are used as inline optical amplifiers in optical communications fiber systems and have major system implications [47]. Oscillation can be obtained if the Er:silica fiber is placed inside an optical resonator. This is often achieved by using fiber Bragg gratings (FBGs) at both ends of a section of these fibers. The fiber Bragg gratings in this case act as the mirrors of the laser resonator. Optical properties of fiber Bragg gratings will be discussed in Chapter 12.

The energy levels involved in the laser transition in a typical Er:silica fiber laser are shown in Figure 6.53. Er^{3+} is a rare-earth ion from the lanthanide family. The erbium ion Er^{3+} has 11 electrons in the 4f orbital. It is an element of choice for lasing and amplification in the 1.5 μm spectral regime, due to its extremely wide gain bandwidth resulting from the transition between the states $^4I_{15/2}$ and $^4I_{13/2}$. In the host medium, all $2J + 1$ components of both multiplets of the erbium ion are split and homogeneously broadened by the environmental perturbations (Stark effect, etc.) due to the neighboring oxygen ions in the silica glass (SiO_2). As a result of the random environments for each of the erbium ions, all these sublevels are further broadened due to inhomogeneous broadening. This leads to the extremely wide gain bandwidth, which is desirable for broadband optical communications applications. There are two strong peaks in absorption cross section of the erbium-doped fiber, at 1480 nm and 980 nm corresponding to the transitions $^4I_{15/2} \rightarrow {}^4I_{11/2}$ and $^4I_{15/2} \rightarrow {}^4I_{13/2}$, respectively. The lifetime of atoms in the upper laser level $^4I_{13/2}$ is $t_2 = 10.2 \times 10^{-3}$ s. Each decay results in the (spontaneous) emission of a photon, so $t_2 = t_{spont}$. Since the laser emission involves the ground state as the lower level, an Er:silica fiber laser is a three-level laser (similar to the ruby laser).

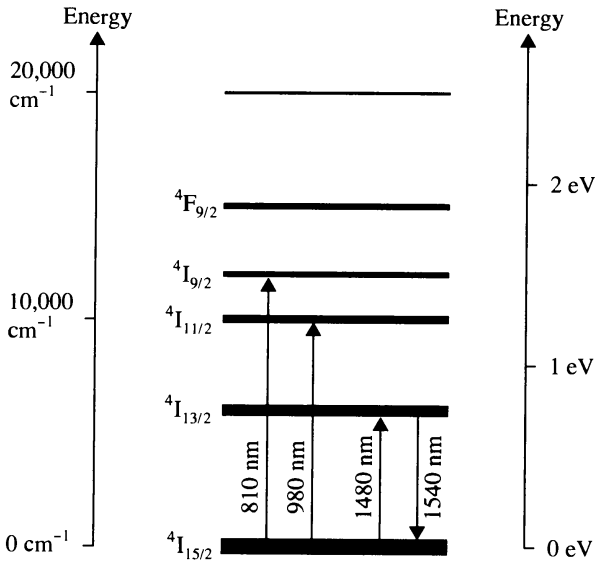


Figure 6.53 Energy levels of Er^{3+} ion and some dominant transitions. The unit 1 cm^{-1} corresponds to $\nu = 30$ GHz, or $h\nu = 1.24 \times 10^{-4}$ eV.

As a result, the fiber must be pumped relatively hard to achieve gain. This problem is eased by the fiber geometry, where the pumped beam can be copropagating with the lasing beam inside the fiber core.

An Er:silica fiber without the fiber Bragg gratings can be employed as an inline optical amplifier in telecommunications networks. More discussion on erbium-doped fiber amplifiers (EDFAs) can be found in Chapter 17.

6.12 FREQUENCY COMB AND OPTICAL FREQUENCY METROLOGY

The mode-locked lasers discussed in Section 6.6 provide a train of ultrashort optical pulses. These optical pulses are identical in shape and equally spaced in time (or space). The periodic pulses are a result of the coupling of the modes of oscillation of the laser resonator via either active or passive mode locking. We showed that the output beam of a mode-locked laser can be written

$$E(t) = \sum_m C_m e^{i[(\Omega_0 + m\Omega)t + \phi_m]} \tag{6.12-1}$$

where Ω is the longitudinal mode spacing in angular frequency, C_m is the amplitude of the mth mode, and ϕ_m is the phase of the mth mode (m = integer). The summation is extended over all the oscillating modes and Ω_0 is a reference frequency. In Equation (6.12-1), we assume that the oscillation frequencies of the resonator are

$$\omega_m = \Omega_0 + m\Omega \tag{6.12-2}$$

where m is an integer and the longitudinal mode spacing Ω is given by

$$\Omega = 2\pi \frac{c}{2n_g l} \tag{6.12-3}$$

where l is the length of the resonator and n_g is the group index. In active mode locking, the modulation frequency is chosen to be exactly Ω. The modulation at this frequency leads to coupling among the modes of oscillation.

In cases where the group velocity dispersion (n_g depends on ω) is present, the mode spacing varies with ω slightly over the spectrum. In these cases the modulation frequency is chosen to be the average of the mode spacing over the entire gain spectrum. When mode locking occurs, coupling among the modes of oscillation causes the "actual frequency" of oscillation to be locked with the frequency given by Equation (6.12-2). In other words, they are exactly equally spaced, even though the modes of the cold resonator may not be exactly equally spaced. As a result, the field envelope is exactly periodic in time with a period of $\tau \equiv 2\pi/\Omega = 2l/v_g = 2n_g l/c$, which is the average round-trip transit time of the mode-locked pulses inside the resonator. According to Equation (6.6-5), we can write

$$E(t+\tau) = \sum_m C_m \exp\left\{i\left[(\Omega_0 + m\Omega)\left(t + \frac{2\pi}{\Omega}\right) + \phi_m\right]\right\}$$

$$= E(t)\exp(i2\pi\Omega_0/\Omega) \tag{6.12-4}$$

Note that $E(t+\tau)$ is identical to $E(t)$, except for a constant phase factor $\Delta\phi = 2\pi\Omega_0/\Omega$. Since Ω_0 is, in general, not an integral multiple of Ω, this phase shift at $t = \tau$ is not an integral multiple of 2π. As a result, the peak of the actual electric field may not coincide with

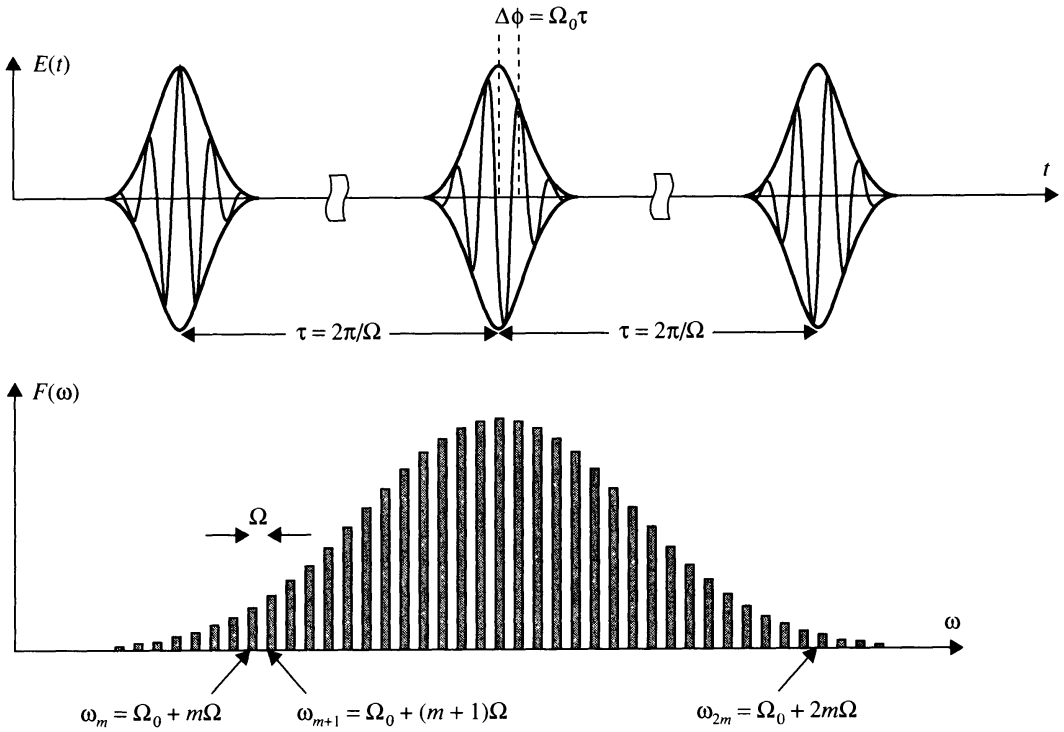


Figure 6.54 A periodic pulse train and its frequency spectrum. The frequency spectrum consists of a discrete set of frequencies that are exactly equally spaced, provided the pulse train is exactly periodic in time.

the peak of the envelope. Figure 6.54 shows the periodic pulse train and the corresponding frequency comb.

As discussed earlier in this chapter, the pulse width is inversely proportional to the gain bandwidth. In the optical spectral regime, the carrier frequency is in the range of a few hundred terahertz (e.g., green light at $\lambda = 500$ nm corresponds to a frequency of $\nu = 600$ THz). A pulse with a temporal duration of few femtoseconds would require a gain bandwidth of approximately a few hundred terahertz. These ultrashort pulses consist of a few optical cycles within the pulse envelope (see Figure 6.54). Mode-locked pulses in the femtosecond regime can be employed for application in optical frequency metrology and, equivalently, the standard of time unit. The discussion here follows closely that of Reference [51]. In the microwave regime, we are able to count the exact number of cycles of the electromagnetic waves. One second in SI units is defined as the time duration that is exactly 9,192,631,770 cycles of oscillation of the cesium clock. The cesium clock is tuned to resonance with the ground-state hyperfine splitting of the cesium atoms, which are kept at an extremely low temperature. In the optical regime, most spectroscopic approaches rely on the measurement of optical wavelengths rather than frequencies. The spectral resolution is limited by several practical issues, including the transverse aperture of the optical beam as well as the physical size of the instrument. For example, the spectral resolution of a grating is of the order of $1/N$, where N is the number of periods. A resolution of $\Delta\lambda/\lambda \approx 10^{-12}$ would require a grating with 10^{12} periods, which is virtually impossible. If we are able to count the number of cycles of oscillation of an optical beam, then the spectral resolution can be at least $\Delta\nu/\nu \approx 10^{-15}$.

By virtue of its periodic nature, the frequency comb provided by a mode-locked laser,

$$\omega_m = \Omega_0 + m\Omega \tag{6.12-5}$$

can be employed as an "optical ruler" for the measurement of optical frequencies, provided both Ω_0 and Ω are kept constant. The integer m (in the range of 10^5–10^6) in the above equation can be defined in such a way that the offset frequency Ω_0 is less than Ω. In this case, Ω_0 may not be the frequency of a mode of oscillation. Equation (6.12-5) maps two radio frequencies Ω_0 and Ω onto the optical frequency ω. The modulation frequency Ω can be kept constant, for example, by using a modulator that is locked to the cesium clock, operating at 9,192,631,770 cycles of oscillation per second. In the case of passive mode locking (e.g., via saturable absorbers), the repetition frequency Ω, which lies between a few tens of megahertz and a few gigahertz, can easily be measured with a photodetector. The offset frequency Ω_0, however, is not easily measurable unless the frequency comb contains more than an optical octave. The measurement of the offset frequency in this case involves the beating of a selected frequency from the comb and the second harmonic of a different frequency of the same comb. This is described as follows.

Referring to the frequency comb in Figure 6.54, we consider a mode of oscillation in the low-frequency portion of the comb whose frequency is given according to Equation (6.12-5) by $\omega_m = \Omega_0 + m\Omega$. Using second-harmonic generation in a nonlinear crystal, we obtain a beam of light at a frequency of $2\omega_m = 2\Omega_0 + 2m\Omega$. If the frequency comb covers a full optical octave, then a mode with a mode number $2m$ should oscillate simultaneously at $\omega_{2m} = \Omega_0 + 2m\Omega$. The mixing of the frequency-doubled mode (m) and the mode at $2m$ yields a beating (difference) frequency of

$$2\omega_m - \omega_{2m} = (2\Omega_0 + 2m\Omega) - (\Omega_0 + 2m\Omega) = \Omega_0 \tag{6.12-6}$$

which is exactly the offset frequency. With this approach, both the repetition frequency Ω and the offset frequency Ω_0 are determined. As discussed in Chapter 4, both frequencies Ω and ω_0 depend on the cavity length. Frequency stabilization is needed to ensure a stable comb. This can be achieved by phase locking both Ω and Ω_0 to a precise radio frequency reference (e.g., Cs clock). Once Ω and Ω_0 are precisely determined, each frequency of the comb is precisely determined. It is even possible to lock the offset frequency at exactly zero ($\Omega_0 = 0$). In this case, the frequency comb is an exact integral multiple of Ω and the electric field is a periodic function with a period of $2\pi/\Omega$.

For the purpose of optical frequency metrology, a mode-locked laser with a gain spectrum spanning a full optical octave is needed. In the following we discuss the Ti:sapphire laser, which can provide the broad bandwidth needed for optical frequency metrology. The discussion here follows closely that of Reference [52]. The active laser particles in the Ti:sapphire laser are Ti^{3+} ions present in Al_2O_3 crystals. Typical concentrations are ~10^{19}–10^{20} cm^{-3} [53]. The energy level diagram of Ti^{3+} ions in sapphire crystal is shown in Figure 6.55.

The electronic structure of the Ti^{3+} ion is a closed shell (argon) plus a single 3d electron. The energy state (3d) of the free Ti^{3+} ion has a degeneracy of 5 due to the spherical symmetry of free space. In the host crystal (Al_2O_3), the Ti^{3+} ion is surrounded by six oxygen atoms situated at the corners of a slightly distorted octahedron. These oxygen atoms produce a nonspherical electric potential at the Ti^{3+} ion. The lower symmetry of the crystal field leads to removal of the degeneracy. As a result, the energy levels are split by the crystal field, which can be viewed as a combination of a cubic field and a trigonal field (sapphire crystal has a point group symmetry of $\bar{3}$ m). The cubic field dominates and splits the energy state into a triply degenerate $^2T_{2g}$ ground state and a doubly degenerate excited state ^2E_g [54]. The trigonal field further splits the ground state $^2T_{2g}$ into two levels, ^2E and $^2A_1$, and the lower one is further split into two levels by the spin–orbit interaction. Figure 6.56 shows the absorption cross section of Ti^{3+} ion due to the transition $^2T_2 \rightarrow {}^2E$. Pumping of the Ti:sapphire laser can

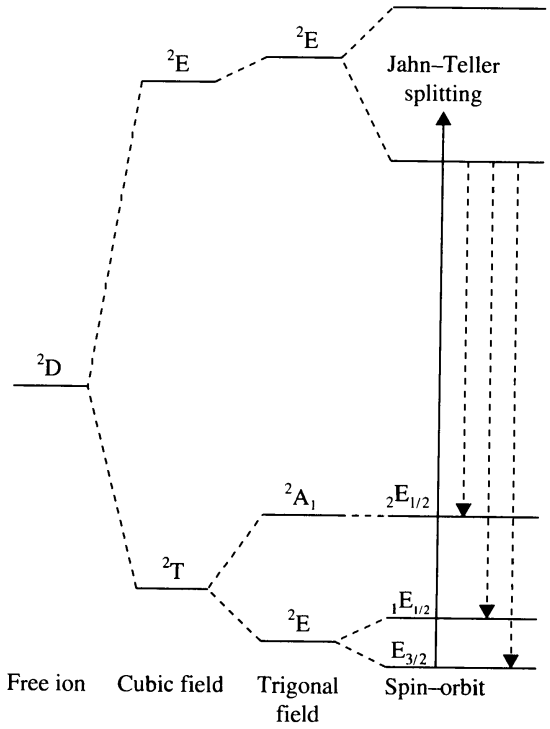


Figure 6.55 Energy levels of Ti^{3+} ions in sapphire crystal showing the effect due to crystal–field, spin–orbit, and Jahn–Teller splitting.

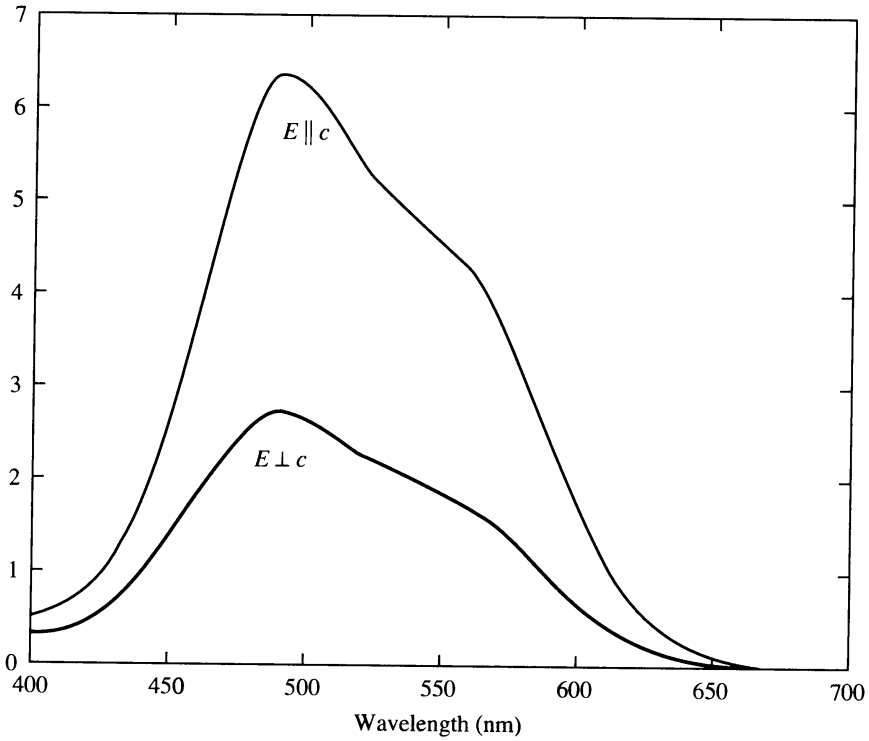


Figure 6.56 Absorption cross section (in units of 10^{-20} cm^2) for the transition $^2T_2 \rightarrow {}^2E$ in the Ti:sapphire laser as a function of wavelength (in units of nm). E = polarization of electric field; c = axis of symmetry of crystal. (Data from Reference [52].)

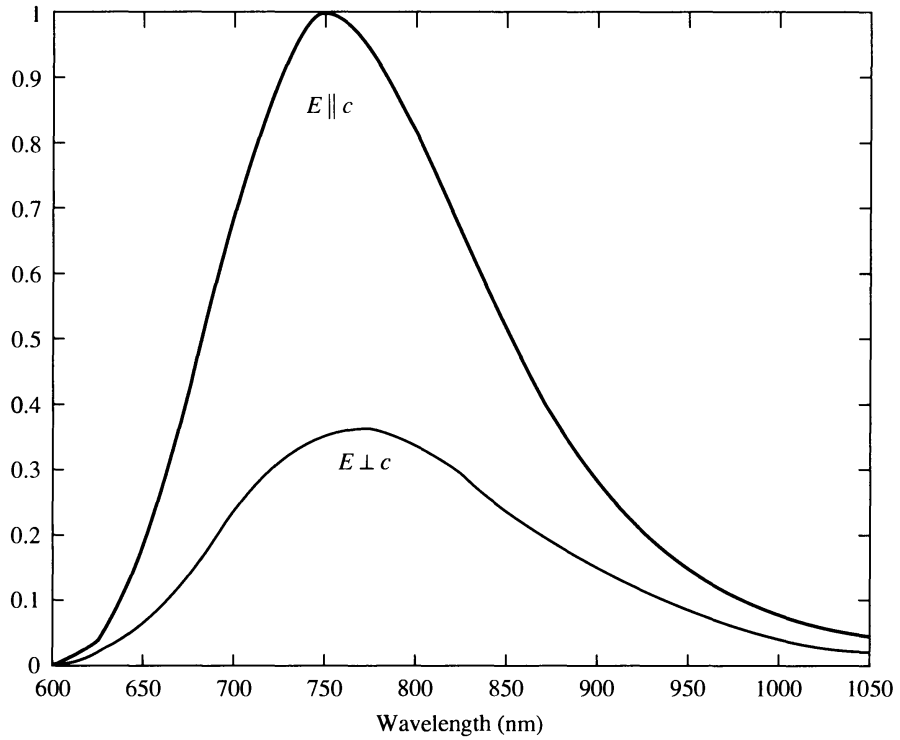


Figure 6.57 Emission spectra of the Ti:sapphire laser. E = polarization of electric field; c = axis of symmetry of crystal. (Data from Reference [52].)

be performed by subjecting it to the light of intense flashlamps (see Figure 6.44). The main peak of absorption spectrum occurs at around λ = 500 nm. The broad absorption peak at around 500 nm is a good match to the 514.5 nm line of the argon ion laser or the doubled YAG wavelength at 532 nm for the purpose of laser optical pumping. When the Ti^{3+} ion is excited to the ^2E_g manifold, it quickly ($\sim 10^{-12}$ s) decays to the minimum energy states within the manifold via the emission of phonons. The system can then decay to the ground state with the emission of a photon and one or more phonons, with a decay lifetime of 3.15 μs at room temperature. Phonons and the nature of the electronic transition are responsible for the broad gain bandwidth of the Ti:sapphire laser.

A close examination of the absorption spectrum reveals actually a dual peak, which is due to the Jahn–Teller splitting of the ^2E state [54]. In addition to the effect of the crystal field, the electronic states are further affected by the interaction between the electron and the vibration of the crystal lattice (phonon). As a result, the doubly degenerate excited state is split into two levels (known as Jahn–Teller splitting [55]). At finite temperatures, the vibration of the Ti ion and the vibration of the crystal lattice lead to a multitude of vibrational states. As a result of homogeneous broadening and inhomogeneous broadening, these states overlap and form a continuum. Depending on the temperature of the crystal and the phonon density in the crystal, these two upper levels can span an energy range of several thousand cm^{-1}. (The unit 1 cm^{-1} corresponds to ν = 30 GHz, or an energy of $h\nu = 1.24 \times 10^{-4}$ eV). As a result, the Ti:sapphire laser exhibits a broad emission spectrum that spans from 600 to 1100 nm (see Figure 6.57). As discussed earlier, a broad gain bandwidth is a prerequisite for ultrashort pulses. To ensure the whole frequency comb spans an optical octave, pulse compression schemes described earlier in this chapter can be employed. This can lead to a frequency comb that is broader than that provided by the gain bandwidth of the laser medium.

PROBLEMS

6.1 Show that the effect of frequency pulling by the atomic medium is to reduce the intermode frequency separation from $c/2l$ to

$$\frac{c}{2l}\left(1 - \frac{\gamma c}{2\pi\Delta\nu}\right)$$

where the symbols are defined in Section 6.2. Calculate the reduction for the case of a laser with $\Delta\nu = 10^9$ Hz, $\gamma = 4 \times 10^{-2}$ m^{-1}, and $l = 100$ cm.

6.2 Derive Equation (6.4-3).

6.3 Derive the optimum coupling condition (Equation (6.5-11)).

6.4 Calculate the critical fluorescence power P_s of the He–Ne laser operating at 6328 Å. Assume $V = 2$ cm^3, $L = 1\%$ per pass, $l = 30$ cm, and $\Delta\nu = 1.5 \times 10^9$ Hz.

6.5 Calculate the critical inversion density N_t of the He–Ne laser described in Problem 6.4.

6.6 Derive an expression for the finesse of a Fabry–Perot etalon containing an inverted population medium. Assume that $r_1^2 = r_2^2 \cong 1$ and that the inversion is insufficient to result in oscillation. Compare the finesse to that of a passive Fabry–Perot etalon.

6.7 Derive an expression for the maximum gain–bandwidth product of a Fabry–Perot regenerative amplifier, that is, with an amplifying medium. Define the bandwidth as the frequency region in which the intensity gain $(E_t E_t^*)/(E_i E_i^*)$ exceeds half its peak value. Assume that $\nu_0 = \nu_m$.

6.8

(a) Derive Equation (6.6-7) and show that the intensity is periodic in time.

(b) Derive Equation (6.6-26)

6.9

(a) Describe qualitatively what one may expect to see in parts A, B, C, and D of the mode-locking experiment sketched in Figure 6.14. (The reader may find it useful to read first the section on photomultipliers in Chapter 11.)

(b) What is the effect of mode locking on the intensity of the beat signal (at $\Omega = \pi c/l$) displayed by the RF spectrum analyzer in B? Assume N equal amplitude modes spaced by Ω whose phases before mode locking are random. [Answer: Mode locking increases the beat signal power by N.]

(c) Show that a standing wave at $\nu_0 + \delta$ (ν_0 is the center frequency of the Doppler-broadened lineshape function) in a gas laser will burn the same two holes in the velocity distribution function (see Figure 6.37) as a field at $\nu_0 - \delta$.

(d) Can two traveling waves, one at $\nu_0 + \delta$ the other at $\nu_0 - \delta$, interact with the same class of atoms? If the answer is yes, under what conditions?

6.10 Design a frequency stabilization scheme for gas lasers based on the Lamb dip (see Figure 6.38). [Hint: You may invent a new scheme, but, failing that, consider what happens to the phase of the modulation in the power output when the cavity length is modulated sinusoidally near the bottom of the Lamb dip. Can you derive an error correction signal from this phase that will control the cavity length?]

6.11 Verify the relations of Table 6.2.

6.12 A helium–neon laser ($\lambda = 0.63$ μm) operating in the fundamental transverse mode has mirrors separated by $l = 30$ cm. The Doppler width is $\Delta\nu_D = 1.5$ GHz, and the effective refractive index is $n = 1$. The output mirror is flat, and the other mirror is spherical with a radius of curvature 16 m.

(a) What is the frequency difference between longitudinal modes in the resonator?

(b) Show that the resonator is stable.

(c) What would the Doppler width become if the temperature of the laser medium were doubled?

(d) What is the spot size at the flat mirror?

(e) If the output is taken from the flat mirror, what is the spot size 16 km away?

(f) Given that the internal cavity loss is $L_i = 10^{-1}$, the small signal gain coefficient is $\gamma_0 = 10^{-3}$ cm^{-1}, and the reflection coefficient of the spherical mirror is 1.0, what is the reflection coefficient (R) of the output mirror that will give the maximum output power?

(g) A thin lens of focal length f is placed against the output mirror. Find the radius of curvature of the Gaussian beam at a distance $d = f$ from the lens.

6.13 Derive the expression relating the absorption cross section at ν in a given $a \rightarrow b$ transition to the spontaneous $b \rightarrow a$ lifetime.

6.14

$$E(t) = \sum_{n=1}^{N} C_n \exp(i\omega_0 t + in\Omega t)$$

For the following, take $N = 5$ and plot in the range Ωt $(-4\pi, 4\pi)$.

(a) Generate a set of N random phases (ϕ_n), and then plot $|E(t)|$ versus Ωt, with $C_n = \exp(i\phi_n)/\sqrt{N}$.
(b) Let $C_n = 1/\sqrt{N}$. Plot $|E(t)|$ versus Ωt.
(c) Let $C_n = \sin\{[s\pi/(N+1)]n\}/\sqrt{(N+1)/2}$. Plot $|E(t)|$ versus Ωt for $s = 1, 2, 3, 4, 5$.
(d) For (b) and (c), derive analytic expressions for the electric field. [Hint: For (c), use $\sin x = (e^{ix} - e^{-ix})/2i$, and sum up the geometric series.]

6.15 Consider an asymmetric Fabry–Perot etalon, with mirror reflectivities R_1 and R_2. The linear attenuation coefficient of the cavity medium is α.

(a) Find the etalon transmission as a function of frequency for normal incidence.
(b) Find the full width at half-maximum $(\Delta\nu_{1/2})_e$ of the transmission peaks.
(c) An external laser beam is in resonance with the etalon. Find the exponential decay constant of the energy stored in the etalon when the laser beam is turned off. This is the photon lifetime t_c in the etalon.
(d) Write down an expression for the E field of electromagnetic radiation emitted from the etalon after the laser beam is turned off. Find the Fourier transform of the E field. Find the full width at half-maximum $(\Delta\nu_{1/2})_s$ of the Fourier spectrum.
(e) Assume $|1 - \sqrt{R_1R_2}\exp(-\alpha l)| \ll 1$. Compare $(\Delta\nu_{1/2})_e$ and $(\Delta\nu_{1/2})_s$ and find their relationship with t_c.
(f) Assume $|1 - \sqrt{R_1R_2}\exp(-\alpha l)| \ll 1$. Find the relationship between cavity Q-factor and the finesse F.

6.16

(a) Substitute Equation (6.7-18) into Equation (6.7-17) and show that

$$\left(1 + \frac{2g_0}{\Omega_g^2\tau^2}\right)\tau^4 = \frac{g_0}{4\sigma_1(\Omega\Omega_g)^2} \quad \text{and}$$

$$g_0 - L = 2\sigma_1\Omega^2\tau^2\left(1 + \frac{g_0}{\Omega_g^2\tau^2}\right)$$

(b) For mode-locked lasers with $g_0 \ll \Omega_g^2\tau^2$, the above equations become

$$\tau^4 = \frac{g_0}{4\sigma_1(\Omega\Omega_g)^2} \quad \text{and} \quad g_0 - L = 2\sigma_1\Omega^2\tau^2$$

Show that the condition $g_0 \ll \Omega_g^2\tau^2$ is equivalent to $4\sigma_1\Omega^2\tau^2 \ll 1$.

6.17 Substitute $E(t) = E_0\,\mathrm{sech}(t/\tau)$ into Equation (6.7-31). Show that $\Delta E(t) = 0$ requires

$$\tau^2 = \frac{2g_0}{\gamma E_0^2\Omega_g^2} \quad \text{and} \quad L - g_0 = \frac{g_0}{\Omega_g^2\tau^2}$$

6.18

(a) Show that for large N $(N \to \infty)$, the full width at half-maximum (FWHM) of the pulse given by Equation (6.6-8) is

$$\Delta t_{1/2} = 0.886\frac{\tau}{N}$$

where $\tau = 2\pi/\Omega$.

(b) Show that for large $N(N \to \infty)$, the FWHM of the intensity of the pulse given by Equation (6.6-26) is

$$\Delta t_{1/2} = 1.19\frac{\tau}{N}$$

We note that the pulse width is about 34% broader.

6.19 Consider a pair of identical diffraction gratings with a separation d and a period Λ. The two gratings are parallel. For simplicity, we assume that they are transmission gratings. Let τ be the group delay, find the group velocity dispersion $d\tau/d\lambda$ as a function of the angle of incidence. Group delay dispersion in double gratings can be employed for pulse compression.

REFERENCES

1. Schawlow, A. L., and C. H. Townes, Infrared and optical masers. *Phys. Rev.* **112**:1940 (1958).
2. Yariv, A., *Quantum Electronics*, 2nd ed. Wiley, New York, 1975.
3. Laures, P., Variation of the 6328 Å gas laser output power with mirror transmission. *Phys. Lett.* **10**:61 (1964).
4. Bennett, W. R., Jr., Gaseous optical masers. *Appl. Opt. Suppl. 1 Optical Masers*, 24 (1962).
5. Fork, R. L., D. R. Herriott, and H. Kogelnik, A scanning spherical mirror interferometer for spectral analysis of laser radiation. *Appl. Opt.* **3**:1471 (1964).

6. DiDomenico, M., Jr., Small signal analysis of internal modulation of lasers. *J. Appl. Phys.* **35**:2870 (1964).
7. Yariv, A., Internal modulation in multimode laser oscillators. *J. Appl. Phys.* **36**:388 (1965).
8. Hargrove, L. E., R. L. Fork, and M. A. Pollack, Locking of He–Ne laser modes induced by synchronous intracavity modulation. *Appl. Phys. Lett.* **5**:4 (1964).
9. DiDomenico, M., Jr., J. E. Geusic, H. M. Marcos, and R. G. Smith, Generation of ultrashort optical pulses by mode locking the Nd^{3+}:YAG laser. *Appl. Phys. Lett.* **8**:180 (1966).
10. Mocker, H., and R. J. Collins, Mode competition and self-locking effects in a Q-switched ruby laser. *Appl. Phys. Lett.* **7**:270 (1965).
11. DeMaria, A. J., Mode locking. *Electronics* Sept. 16:112 (1968).
12. Valdmanis, J. A., R. L. Ford, and J. P. Gordon, *Opt. Lett.* **10**:131 (1985).
13. Shapiro, S. L. (ed.), *Ultrashort Light Pulses* (Topics in Applied Physics, Vol. 18). Springer-Verlag, New York, 1977.
14. Hochstrasser, R. M., W. Kaiser, and C. V. Shank (eds.), *Picosecond Phenomena II* (Series in Chemical Physics, Vol. 14). Springer-Verlag, New York, 1980.
15. Eisenthal, K. B., R. M. Hochstrasser, W. Kaiser, and A. Laubereau (eds.), *Picosecond Phenomena III* (Series in Chemical Physics, Vol. 23). Springer-Verlag, New York, 1982.
16. Schrans, T., *Part I: Longitudinal Static and Dynamic Effects in Semiconductor Lasers; Part II: Spectral Characteristics of Passively Mode-Locked Quantum Well Lasers.* Ph.D. thesis, California Institute of Technology, 1994.
17. Maiman, T. H., Stimulated optical radiation in ruby masers. *Nature* **187**:493 (1960).
18. Sanders, S., L. Eng, J. Paslaski, and A. Yariv, 108 GHz passive mode locking of a multiple quantum well semiconductor laser with an intracavity absorber. *Appl. Phys. Lett.* **56**:310 (1990).
19. Koch, T. L., *Gigawatt Picosecond Dye Lasers and Ultrafast Processes in Semiconductor Lasers.* Ph.D. thesis, California Institute of Technology, 1982.
20. Yariv, A., *Quantum Electronics*, 3rd ed. Wiley, New York, 1989, p. 550.
21. Siegman, A. E., and D. J. Kuizenga, Simple analytic expressions for AM and FM mode locked pulses in homogeneous lasers. *Appl. Phys. Lett.* **14**:181 (1969).
22. Kuizenga, D. J., and A. E. Siegman, FM and AM mode locking of the homogeneous laser: Part I, Theory; Part II, Experiment. *J. Quantum Electron.* **QE- 6**:694 (1970).
23. Siegman, A. E., *Lasers.* University Science Books, Mill Valley, CA, 1986.
24. Tracy, E. B., *IEEE J. Quantum Electron.* **5**:454 (1969).
25. Haus, H. A., Mode-locking of lasers. *IEEE J. Selected Topics Quantum Electron.* **6**(6):1173 (2000).
26. Kuizenga, D. I., and A. E. Siegman, Modulator frequency detuning effects in the FM mode-locked lasers. *IEEE J. Quantum Electron.* **6**:803 (1970).
27. Smith, P. J., T. J. Bridges, and E. J. Burkhardt, Mode locked high pressure CO_2 laser. *Appl. Phys. Lett.* **21**:470 (1972).
28. Martinez, O. E., J. P. Gordon, and R. L. Fork, *J. Opt. Soc. Am. A* **1**:1003 (1984).
29. Martinez, O. E., *IEEE J. Quantum Electron.* **23**:59 (1987).
30. Wagner, W. G., and B. A. Lengyel, Evolution of the giant pulse in a laser. *J. Appl. Phys.* **34**:2042 (1963).
31. Hellwarth, R. W., Control of fluorescent pulsations. In: *Advances in Quantum Electronics*, J. R. Singer (ed.). Columbia University Press, New York, 1961, p. 334.
32. McClung, F. J., and R. W. Hellwarth, Giant Optical Pulsations from Ruby. *J. Appl. Phys.* **33**:828 (1962).

33. Hellwarth, R. W., *Q* modulation of lasers. In: *Lasers*, Vol. 1, A. K. Levine (ed.). Marcel Dekker, New York, 1966, p. 253.
34. DeMaria, A. J., Picosecond laser pulses. *Proc. IEEE* **57**:3 (1969).
35. Lamb, W. E., Jr., Theory of an optical maser. *Phys. Rev. A* **134**:1429 (1964).
36. Szöke, A., and A. Javan, Isotope shift and saturation behavior of the *1.15jL* transition of neon. *Phys. Rev. Lett.* **10**:512 (1963).
37. Bloom, A., *Gas Lasers*. Wiley, New York, 1963, p. 93.
38. Maiman, T. H., Optical and microwave-optical experiments in ruby. *Phys. Rev. Lett.* **4**:564 (1960).
39. Cronemeyer, D. C., Optical absorption characteristics of pink ruby. *J. Opt. Soc. Am.* **56**:1703 (1966).
40. Schawlow, A. L., Fine structure and properties of chromium fluorescence. In: *Advances in Quantum Electronics*, J. R. Singer (ed.). Columbia University Press, New York, 1961, p. 53.
41. Yariv, A., Energy and power considerations in injection and optically pumped lasers. *Proc. IEEE* **51**:1723 (1963).
42. Geusic, J. E., H. M. Marcos, and L. G. Van Uitert, Laser oscillations in Nd-doped yttrium aluminum, yttrium gallium and gadolinium garnets. *Appl. Phys. Lett.* **4**:182 (1964).
43. Kushida, T., H. M. Marcos, and J. E. Geusic, Laser transition cross section and fluorescence branching ratio for Nd^{3+} in yttrium aluminum garnet. *Phys. Rev.* **167**:1289 (1968).
44. Snitzer, E., and C. G. Young, Glass lasers. In: *Lasers*, Vol. 2, A. K. Levine (ed.). Marcel Dekker, New York, 1968, p. 191.
45. Simon, J. C., Semiconductor laser amplifier for single mode optical fiber communications. *J. Opt. Commun.* **4**:51 (1983).
46. Mears, R. J., L. Reekie, I. M. Jauncey, and D. N. Payne, Low noise erbium-doped fiber amplifier operating at 1.54 mm. *Electron. Lett.* **23**:1026 (1987).
47. Hagimoto, K., et al., A 212 km non-repeatered transmission experiment at 1.8 Gb/s using LD pumped Er^{+}-doped fiber amplifiers in an 1m/direct-detection repeater system. In: *Proceedings of the Optical Fiber Conference, Houston, IX*. Postdeadline Paper PD15, 1989.
48. Olshansky, R., Noise figure for Er-doped optical fibre amplifiers. *Electron. Lett.* **24**:1363 (1988).
49. Payne, D. N., Tutorial session abstracts. In: *Optical Fiber Communication (OFC 1990) Conference*, San Francisco, 1990.
50. See, for example, G. Eisenstein, U. Koren, G. Raybon, T. L. Koch, M. Wiesenfeld, M. Wegener, R. S. Tucker, and B. I. Miller, Large-signal and small-signal gain characteristics of 1.5 μm quantum well optical amplifiers. *Appl. Phys. Lett.* **56**:201 (1990).
51. Udem, Th., R. Holzwarth, and T. W. Hänsch, Optical frequency metrology. *Nature* **416**:233 (14 Mar. 2002).
52. Moulton, P. F., Spectroscopic and laser characteristics of Ti:Al_2O_3. *J. Opt. Soc. Am. B* **3**(1):125 (1986).
53. McKinnie, I. T., A. L. Oien, D. M. Warrington, P. N. Tonga, L. A. W. Gloster, and T. A. King, Ti ion concentration and Ti:sapphire laser performance. *IEEE J. Quantum Electron.* **33**(7):1221 (1997).
54. In group theory, A_1 and A_2 stand for one-dimensional irreducible representations, E stands for a two-dimensional irreducible representation, and T_1 and T_2 stand for three-dimensional irreducible representation of point group O. The subscripts *g* and *u*, for gerade and ungerade, indicate even and odd representations under inversion. See, for example, M. Tinkham, *Group Theory and Quantum Mechanics*. McGraw-Hill, New York, 1964.
55. Jahn, H. A., and E. Teller, *Proc. Roy. Soc. London A* **161**:220 (1937).

ADDITIONAL READING

Becker, P. C., N. A. Olsson, and J. R. Simpson, *Erbium-Doped Fiber Amplifiers: Fundamentals and Technology*, Academic Press, San Diego, 1999.

Desurvire, E., *Erbium-Doped Fiber Amplifiers: Principles and Applications*. Wiley, New York, 2002.

Digonnet, M. J. F., *Rare-Earth Doped Fiber Lasers and Amplifiers*. Marcel Dekker, New York, 2001.

CHAPTER 7

CHROMATIC DISPERSION AND POLARIZATION MODE DISPERSION IN FIBERS

7.0 INTRODUCTION

Chromatic dispersion is an intrinsic property of virtually all optical materials. The phenomenon is manifested as a dependence of the phase velocity (and group velocity) of a beam of light in a transparent material on the frequency (or wavelength) of light. Similarly, chromatic dispersion occurs in optical fibers that are made of transparent materials. As a result of chromatic dispersion, an optical pulse spreads as it propagates along the fiber. The spreading leads to signal fading when neighboring pulses overlap. The problem becomes progressively more severe as the propagation distance and the data rate increase. In addition to chromatic dispersion, optical fibers also exhibit birefringent properties in which the phase and (group) velocity depend on the beam polarization. The birefringence is a result of fiber imperfection and/or external perturbations. As a result, the two LP modes (e.g., LP_{01x} and LP_{01y}) are no longer degenerate. In other words, a time delay occurs between the two principal modes of propagation—the slow mode and the fast mode. The time delay leads to a broadening of the optical pulse and to signal fading. In this chapter we investigate quantitatively the effect of chromatic dispersion and birefringence on the propagation of pulses in optical fibers.

7.1 CHROMATIC DISPERSION IN OPTICAL TRANSMISSION SYSTEMS

In Chapter 1 we describe chromatic dispersion in terms of the dependence of the index of refraction on the frequency (or wavelength) of light. This is important for studying the propagation of optical pulses in a homogeneous medium. As a result of chromatic dispersion, the group velocity of light depends on the frequency (or wavelength). This is known as the group velocity dispersion (GVD). In optical electronics, we often deal with the transmission of optical waves in various optical systems, including fibers, modulators, and amplifiers. The group velocity dispersion in a general optical system can be described by examining the phase shift as a function of the frequency.

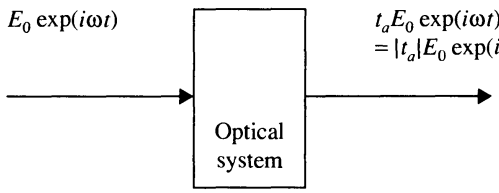


Figure 7.1 Schematic drawing of a general optical transmission system, where t_a is the amplitude transmission coefficient. The box can be a spool of fiber, a Fabry–Perot etalon, a spectral filter, and so on.

Referring to Figure 7.1, we consider the transmission of a monochromatic beam of light through a general optical system.

The amplitude transmission coefficient t_a can be written

$$t_a = \sqrt{T}\ \exp(-i\psi) \tag{7.1-1}$$

where T is the intensity transmission coefficient and ψ is the phase shift. In a homogeneous medium, the phase shift is simply

$$\psi = kL = \frac{\omega}{c} nL \tag{7.1-2}$$

where ω is the angular frequency, n is the index of refraction, and L is the length of propagation.

We consider the transmission of an optical pulse. Let the input optical pulse be written as a product of the optical carrier $\exp(i\omega_0 t)$ and an envelope function $s(t)$:

$$E_{\text{in}}(t) = s(t)e^{i\omega_0 t} = e^{i\omega_0 t} \int A(\Omega)e^{i\Omega t} d\Omega \tag{7.1-3}$$

where ω_0 is the carrier frequency of the pulse and $A(\Omega)$ is the Fourier spectrum of the envelope pulse. The envelope function $s(t)$ can be viewed as a time-varying amplitude. The output pulse after transmitting through the optical system in Figure 7.1 can be written

$$E_{\text{out}}(t) = s'(t)e^{i\omega_0 t} = e^{i\omega_0 t} \int A(\Omega)e^{i\Omega t}\sqrt{T}e^{-i\psi} d\Omega \tag{7.1-4}$$

We consider the situation where the intensity transmission is uniform over the optical spectral regime of the input pulse. In this case, the output pulse can be written

$$E_{\text{out}}(t) = s'(t)e^{i\omega_0 t} = e^{i\omega_0 t}\sqrt{T} \int A(\Omega)e^{i\Omega t}e^{-i\psi} d\Omega \tag{7.1-5}$$

Generally, the output pulse shape $s'(t)$ is different from the input pulse shape $s(t)$. This is a consequence of the dependence of the phase shift $\psi(\Omega)$ on frequency.

In optical communications, especially when the carrier is a laser beam with a well-defined carrier frequency ω_0, the Fourier spectrum of the envelope pulse $A(\Omega)$ is often sharply peaked around $\Omega = 0$. To study the pulse shape of the output pulse, we expand $\psi(\Omega)$ around $\Omega = 0$ in terms of a Taylor series:

$$\psi(\Omega) = \psi_0 + \left(\frac{d\psi}{d\Omega}\right)_0 \Omega + \cdots \tag{7.1-6}$$

which when substituted in Equation (7.1-5) results in

$$E_{\text{out}}(t) = s'(t)e^{i\omega_0 t} = e^{i\omega_0 t}\sqrt{T}e^{-i\psi_0} \int A(\Omega)e^{i\Omega t}e^{-i(d\psi/d\Omega)_0 \Omega} d\Omega \tag{7.1-7}$$

where we have neglected the higher-order terms in Ω. Comparing the last equation with Equation (7.1-3), we can rewrite it as

$$E_{\text{out}}(t) = s'(t)e^{i\omega_0 t} = e^{i\omega_0 t}\sqrt{T}\,e^{-i\psi_0}s(t-\tau_0) \tag{7.1-8}$$

where

$$\tau_0 = \left(\frac{d\psi}{d\Omega}\right)_0 \tag{7.1-9}$$

is the time of flight (or propagation time) of the pulse envelope.

This shows that apart from an overall phase factor and an amplitude factor, the output pulse remains identical to that of the input pulse, with a time delay of $(d\psi/d\Omega)_0$. This approximation is legitimate provided the spectral distribution $A(\Omega)$ is sharply peaked around $\Omega = 0$, and the phase shift is a smoothly varying function of Ω. Thus the propagation time delay, referred to as the group delay, due to propagation through the optical system described in Figure 7.1 is given by

$$\tau = \frac{d\psi}{d\omega} \tag{7.1-10}$$

which is, in general, a function of frequency. In a homogeneous medium, $\psi = kL$, Equation (7.1-10) becomes

$$\tau = \frac{dk}{d\omega}L = \frac{L}{v_g} \tag{7.1-11}$$

where v_g is the group velocity.

If the flight time τ is independent of the frequency, then the group delay has no dispersion. This is the situation when all frequency components of an optical pulse travel at exactly the same group velocity, and thus the same flight time τ. In this case, the pulse shape remains the same. In the event when the group delay τ is a function of frequency, then the output pulse may spread. The pulse spread can be estimated as follows, provided the pulse is not severely distorted:

$$\Delta\tau = \left(\frac{d\tau}{d\omega}\right)_0 \Delta\omega = \left(\frac{d^2\psi}{d\omega^2}\right)_0 \Delta\omega \tag{7.1-12}$$

where $\Delta\omega$ is the spectral spread of the input pulse. We note that the pulse spread is proportional to the product of the second-order derivative of the phase shift with respect to the frequency, and the spectral bandwidth of the input pulse.

In optical communications, the pulse spread is often written in terms of the wavelength spread $\Delta\lambda$:

$$\Delta\tau = \left(\frac{d\tau}{d\lambda}\right)_0 \Delta\lambda \tag{7.1-13}$$

where we can define

$$\tau_\lambda = \left(\frac{d\tau}{d\lambda}\right)_0 \quad \text{(ps/nm)} \tag{7.1-14}$$

as the group delay dispersion, which is often expressed in units of ps/nm. For propagation in a single-mode fiber, the phase shift is simply the product of the propagation constant and the length of the fiber,

$$\psi = \beta L \tag{7.1-15}$$

where β is the propagation constant of the mode. Using Equations (7.1-12), (7.1-13), (7.1-14), and (7.1-15), we obtain

$$\tau_\lambda = -\frac{2\pi c}{\lambda^2}\left(\frac{d^2\psi}{d\omega^2}\right)_0 = -\frac{2\pi c}{\lambda^2}\left(\frac{d^2\beta}{d\omega^2}\right)_0 L \tag{7.1-16}$$

We note that τ_λ is proportional to the length of propagation. For optical engineering, it is convenient to define the group velocity dispersion D in units of ps/km-nm as

$$D = -\frac{2\pi c}{\lambda^2}\left(\frac{d^2\beta}{d\omega^2}\right)_0 \tag{7.1-17}$$

so that τ_λ is given by

$$\tau_\lambda = DL \tag{7.1-18}$$

where L is the length of the fiber. For example, consider a single-mode fiber with a length of $L = 100$ km and a dispersion of $D = 17$ ps/nm-km; then the group delay dispersion is $\tau_\lambda = 1700$ ps/nm.

In the event when the group delay τ is dependent on the frequency, the output shape can be obtained by using the following expression of the phase shift and Equation (7.1-5):

$$\begin{aligned}\psi(\Omega) &= \psi_0 + \tau_0\Omega + \frac{1}{2}\left(\frac{d^2\psi}{d\omega^2}\right)_0 \Omega^2 + \cdots \\ &= \psi_0 + \tau_0\Omega - \frac{\pi c}{\omega_0^2}\tau_\lambda\Omega^2 + \cdots\end{aligned} \tag{7.1-19}$$

In the next section, we consider the transmission of an optical pulse with a Gaussian pulse shape through a dispersive system.

In optical communications networks, signal pulses may have to travel through various optical elements or subsystems. The total group delay is the sum of the group delays in each transmission system. Consider, for example, the transmission through two optical systems in series. The total phase shift can be written

$$\psi = \psi_1 + \psi_2 \tag{7.1-20}$$

where ψ_1 and ψ_2 are the phase shifts of the optical systems, respectively. The total group delay is given by

$$\tau = \frac{d\psi}{d\omega} = \frac{d\psi_1}{d\omega} + \frac{d\psi_2}{d\omega} = \tau_1 + \tau_2 \tag{7.1-21}$$

where τ_1 and τ_2 are the individual group delays. We note that τ_1 and τ_2 are both positive. The pulse broadening due to group velocity dispersion (GVD) is given by

$$\Delta\tau = \left(\frac{d\tau_1}{d\lambda}\right)_0 \Delta\lambda + \left(\frac{d\tau_2}{d\lambda}\right)_0 \Delta\lambda \tag{7.1-22}$$

where $\Delta\lambda$ is the spectral bandwidth (in units of nm) of the pulse.

If the two optical systems have group velocity dispersion of opposite signs, then the pulse broadening can vanish provided

$$\left(\frac{d\tau_1}{d\lambda}\right)_0 + \left(\frac{d\tau_2}{d\lambda}\right)_0 = 0 \tag{7.1-23}$$

This is the basic principle of dispersion compensation. In the case of transmission through two optical fibers, Equation (7.1-23) is written, according to Equation (7.1-18),

$$D_1 L_1 + D_2 L_2 = 0 \tag{7.1-24}$$

where D_1 and D_2 are the dispersions of the fibers, and L_1 and L_2 are the fiber lengths. Most single-mode fibers for optical communications exhibit a positive dispersion ($D_1 > 0$). Thus a specially designed optical fiber with negative dispersion ($D_2 < 0$) can be employed for the restoration of pulse shapes.

7.2 OPTICAL PULSE SPREADING IN DISPERSIVE MEDIA

Virtually all the traffic carried by optical fibers in modern communications is in the form of digital pulses, each representing one bit of information. It thus follows that the narrower the pulses, the more of them can be packed into a given (transmission) time slot, and thus more data (bits) can be transmitted during that time. As a matter of fact, modern communications systems are designed with pulse widths as narrow as 3×10^{-11} s and with data rates exceeding 10^{10} bits/s. In a 10 Gb/s system, there are 10 billion bits per second. The trend to ever narrower pulses and higher rates continues unabated. What limits our ability to reduce the pulse width even further is the basic phenomenon of pulse broadening due to the dependence of its group velocity on the frequency. This phenomenon is termed *group velocity dispersion* (*GVD*). To be specific, we consider the realistic scenario of a fiber supporting a single spatial mode, usually the lowest-order fundamental mode, which is excited at $z = 0$ by a temporal Gaussian pulse

$$E(x, y, z = 0, t) = u_0(x, y) \exp(-\alpha t^2 + i\omega_0 t) \tag{7.2-1}$$

where $u_0(x, y)$ is the wavefunction of a confined mode, α is a constant, and ω_0 is the optical carrier frequency. We consider the case of a slowly varying envelope so that there are many optical oscillations (cycles) within the envelope. This corresponds to the situation where $\alpha^{1/2} \ll \omega_0$. We may express the input pulse $E(x, y, 0, t)$ as a Fourier integral

$$E(z = 0, t) = \exp(i\omega_0 t) \int F(\Omega) e^{i\Omega t}\, d\Omega \tag{7.2-2}$$

where $F(\Omega)$ is the Fourier transform of the Gaussian envelope $\exp(-\alpha t^2)$

$$F(\Omega) = \sqrt{\frac{1}{4\pi\alpha}} \exp\left(-\frac{\Omega^2}{4\alpha}\right) \tag{7.2-3}$$

In the above equations, we ignore the wavefunction $u_0(x, y)$. This is legitimate provided the wavefunction remains the same within the spectral band of the signal. Note that the frequency spectrum of a Gaussian pulse is also a Gaussian. We may view Equation (7.2-2) as an assembly of harmonic fields, each with its unique frequency ($\omega_0 + \Omega$) and amplitude $F(\Omega)\, d\Omega$. To obtain the field at an output plane z, we need to multiply each frequency component $F(\Omega)\, d\Omega \exp[i(\omega_0 + \Omega)t]$ in Equation (7.2-2) by its propagation phase delay factor,

$$\exp[-i\beta(\omega_0 + \Omega)z] \tag{7.2-4}$$

The result is

$$E(z, t) = \int F(\Omega) \exp\{i[(\omega_0 + \Omega)t - \beta(\omega_0 + \Omega)z]\}\, d\Omega \tag{7.2-5}$$

We can expand $\beta(\omega_0 + \Omega)$ near the center (optical) frequency ω_0 in a Taylor series

$$\beta(\omega_0 + \Omega) = \beta(\omega_0) + \left.\frac{d\beta}{d\omega}\right|_{\omega_0} \Omega + \frac{1}{2}\left.\frac{d^2\beta}{d\omega^2}\right|_{\omega_0} \Omega^2 + \cdots \tag{7.2-6}$$

and obtain

$$\begin{aligned} E(z, t) &= \exp[i(\omega_0 t - \beta_0 z)] \int_{-\infty}^{\infty} d\Omega\, F(\Omega) \exp\left\{i\left[\Omega t - \frac{\Omega z}{v_g} - \frac{1}{2}\frac{d}{d\omega}\left(\frac{1}{v_g}\right)\Omega^2 z\right]\right\} \\ &\equiv \exp[i(\omega_0 t - \beta_0 z)]\mathscr{E}(z, t) \end{aligned} \tag{7.2-7}$$

where

$$\beta_0 \equiv \beta(\omega_0), \quad \left.\frac{d\beta}{d\omega}\right|_{\omega_0} = \frac{1}{v_g} = \frac{1}{\text{Group velocity}} \tag{7.2-8}$$

The field envelope is given by the integral (7.2-7)

$$\begin{aligned} \mathscr{E}(z, t) &= \int_{-\infty}^{\infty} d\Omega\, F(\Omega) \exp\left\{i\Omega\left[\left(t - \frac{z}{v_g}\right) - \frac{1}{2}\frac{d}{d\omega}\left(\frac{1}{v_g}\right)\Omega z\right]\right\} \\ &= \int_{-\infty}^{\infty} d\Omega\, F(\Omega) \exp\left\{i\Omega\left[\left(t - \frac{z}{v_g}\right) - a\Omega z\right]\right\} \end{aligned} \tag{7.2-9}$$

where a is a constant defined as $a = (d^2\beta/d\omega^2)/2$. Here we neglect the high-order terms (Ω^3 and up) in the Taylor's series expansion in Equation (7.2-6). We note that in the absence of group velocity dispersion when $a = 0$, the envelope can be written

$$\mathscr{E}(z, t) = \int_{-\infty}^{\infty} d\Omega\, F(\Omega) \exp\left\{i\Omega\left(t - \frac{z}{v_g}\right)\right\} = \mathscr{E}\left[0, \left(t - \frac{z}{v_g}\right)\right] \tag{7.2-10}$$

In other words, the pulse envelope remains unchanged and propagates at the group velocity, v_g. This is in agreement with the general discussion in the last section (Equation (7.1-8)). The pulse spreading is caused by the group velocity dispersion characterized by the parameter

$$a \equiv \frac{1}{2}\left.\frac{d^2\beta}{d\omega^2}\right|_{\omega=\omega_0} = \frac{1}{2}\frac{d}{d\omega}\left(\frac{1}{v_g}\right) = -\frac{1}{2v_g^2}\frac{dv_g}{d\omega} \tag{7.2-11}$$

After substituting $F(\Omega)$ from (7.2-3), Equation (7.2-9) becomes

$$\mathscr{E}(z, t) = \sqrt{\frac{1}{4\pi\alpha}} \int_{-\infty}^{\infty} \exp\left\{-\left[\Omega^2\left(\frac{1}{4\alpha} + iaz\right) - i\left(t - \frac{z}{v_g}\right)\Omega\right]\right\} d\Omega \tag{7.2-12}$$

Carrying out the integration, we obtain

$$\mathcal{E}(z, t) = \frac{1}{\sqrt{1 + i4a\alpha z}} \exp\left(-\frac{(t - z/v_g)^2}{1/\alpha + 16a^2z^2\alpha}\right) \exp\left(i\frac{4az(t - z/v_g)^2}{1/\alpha^2 + 16a^2z^2}\right) \tag{7.2-13}$$

The pulse duration τ at z can be taken as the separation between the two times when the pulse envelope squared is reduced by a factor of $\frac{1}{2}$ from its peak value (so-called FWHM), that is,

$$\tau(z) = \sqrt{2 \ln 2}\sqrt{\frac{1}{\alpha} + 16a^2z^2\alpha} \tag{7.2-14}$$

The initial pulse width is

$$\tau_0 = \left(\frac{2 \ln 2}{\alpha}\right)^{1/2} \tag{7.2-15}$$

The pulse width after propagating a distance L can thus be expressed as

$$\tau(L) = \tau_0\sqrt{1 + \left(\frac{8aL \ln 2}{\tau_0^2}\right)^2} \tag{7.2-16}$$

At large distances such that $|aL| \gg \tau_0^2$, we obtain

$$\tau(L) \sim \frac{(8 \ln 2)aL}{\tau_0} \tag{7.2-17}$$

If we use the definition (7.2-11) of the factor a, the last expression becomes

$$\tau(L) = \frac{4 \ln 2}{v_g^2}\left|\frac{dv_g}{d\omega}\right|\frac{L}{\tau_0} \tag{7.2-18}$$

The group velocity dispersion is often characterized by $D \equiv L^{-1}(dT/d\lambda)$, where T is the pulse transmission time (group delay) through length L of the fiber. This definition is related to the second-order derivative of β with respect to ω as, according to Equation (7.1-17),

$$D = -\frac{2\pi c}{\lambda^2}\left(\frac{d^2\beta}{d\omega^2}\right) \tag{7.2-19}$$

and is related to the parameter a used above by

$$D = -\frac{4\pi c}{\lambda^2}a \tag{7.2-20}$$

With this new definition, the pulse width expression (7.2-9) can be written

$$\tau(L) = \tau_0\sqrt{1 + \left(\frac{2 \ln 2}{\pi c}\frac{DL\lambda^2}{\tau_0^2}\right)^2} \tag{7.2-21}$$

If DL is in units of picoseconds per nanometer, λ is in units of micrometers, and τ_0 is in units of picoseconds, the pulse width in units of picoseconds can be written

$$\tau(L) = \tau_0\sqrt{1 + \left(\frac{1.47DL\lambda^2}{\tau_0^2}\right)^2} \tag{7.2-22}$$

At large distances such that $|DL\lambda^2| \gg \tau_0^2$, Equation (7.2-21) becomes

$$\tau(L) = \frac{2 \ln 2}{\pi c} \frac{DL\lambda^2}{\tau_0} \tag{7.2-23a}$$

or equivalently (when DL is in units of ps/nm, λ is in units of μm, and τ_0 is in units of ps), according to Equation (7.2-22),

$$\tau(L) = \frac{1.47 DL\lambda^2}{\tau_0} \tag{7.2-23b}$$

The expression (7.2-23a) is consistent with

$$\Delta\tau = DL\,\Delta\lambda \tag{7.2-24}$$

provided we take the bandwidth of the pulse as

$$\Delta\lambda = \frac{2 \ln 2}{\pi c} \frac{\lambda^2}{\tau_0} \tag{7.2-25}$$

This bandwidth $\Delta\lambda$ is consistent with the FWHM of the square of the Fourier spectrum (7.2-3) (see Problem 7.2).

The group velocity dispersion, that is, the dependence of v_g on ω, in optical fibers is due to the sum of material dispersion and waveguide dispersion (see Section 3.5). As we discussed in the last section, it is possible to design optical fibers such that the group velocity dispersion is either at a minimum or at a desired value in the spectral regime of interest.

Frequency Chirp

In addition to pulse broadening, the effect of pulse propagation in a dispersive fiber ($d^2\beta/d\omega^2 \neq 0$) is to modify the optical frequency. In the case of a Gaussian pulse after propagating a distance z, the result, according to Equations (7.2-7) and (7.2-13), is

$$E(z, t) = \frac{1}{\sqrt{1 + i4a\alpha z}} \exp\left(-\frac{(t - z/v_g)^2}{1/\alpha + 16a^2z^2\alpha}\right) \exp\left(i(\omega_0 t - \beta_0 z) + i\frac{4az(t - z/v_g)^2}{1/\alpha^2 + 16a^2z^2}\right)$$
$$a = \frac{1}{2}\frac{d^2\beta}{d\omega^2} = -\frac{1}{2v_g^2}\frac{dv_g}{d\omega} \tag{7.2-26}$$

The overall phase of the optical field is

$$\phi(z, t) = \omega_0 t - \beta_0 z + \frac{4az(t - z/v_g)^2}{1/\alpha^2 + 16a^2z^2} \tag{7.2-27}$$

The optical frequency $\omega(z, t)$ is fundamentally the number of (optical) oscillations per second as measured by a stationary observer at z. It is thus given by

$$\omega(z, t) = \frac{\partial}{\partial t}\phi(z, t) = \omega_0 + 8\frac{az}{(1/\alpha^2 + 16a^2z^2)}(t - z/v_g) \tag{7.2-28}$$

The frequency is not a constant but is "chirped." The linear chirp is a consequence of the group velocity dispersion causing different "groups" of frequencies to travel at different velocities and thus spread themselves along the pulse.

A chirped optical pulse resulting from propagation in a fiber with $d^2\beta/d\omega^2 < 0$ ($dv_g/d\omega > 0$) is shown in Figure 7.2, where the horizontal axis is the position. The "blue" (high-frequency)

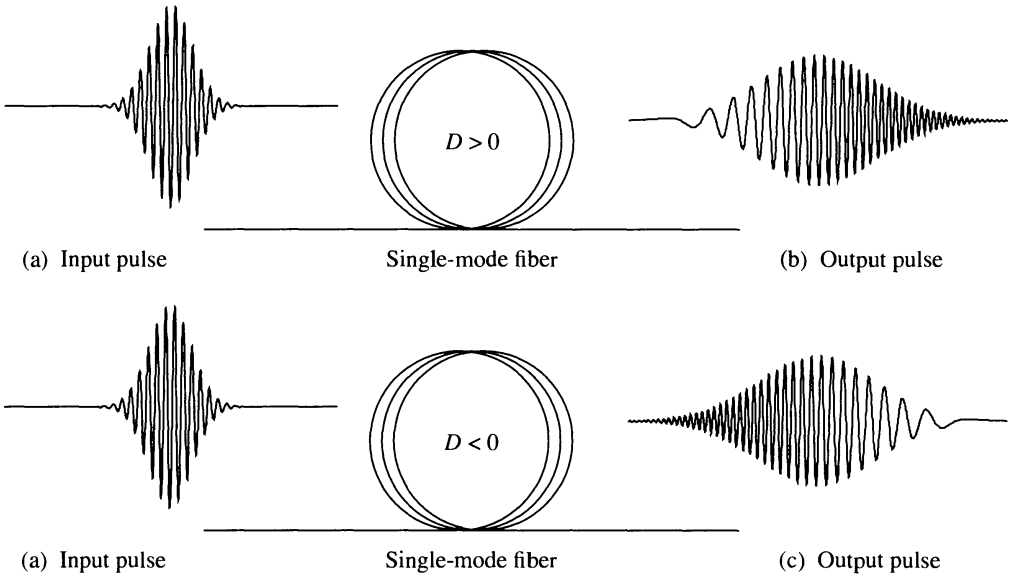


Figure 7.2 (a) An unchirped optical pulse with a Gaussian envelope (spatial profile). (b) The spatial profile of the same pulse after propagation in a single-mode fiber with positive dispersion ($D > 0$). The pulse is broadened and chirped. Note that the leading edge of the pulse consists of high-frequency components that travel faster in positive dispersion fibers. (c) The spatial profile of the same pulse after propagation in a single-mode fiber with negative dispersion ($D < 0$). The pulse is broadened and chirped. Note that the leading edge of the pulse consists of low-frequency components that travel faster in negative dispersion fibers.

portion of the pulse spectrum travels faster and arrives at z before the "red" portion, so that the frequency decreases (linearly) with time.

Pulses with Other Profiles

For optical pulses with a Gaussian profile, the Fourier transforms are also Gaussian. In other words, the spectrum follows a Gaussian distribution around the carrier frequency. In practice, the pulses in optical communications are not exactly Gaussian in shape. The variation in shape can lead to a change of the spectral distribution and hence can affect the broadening of the pulse after propagation through a dispersive medium. Figure 7.3 shows the pulse broadening of three different pulses. They are a trapezoid pulse, a Gaussian pulse, and a cosine pulse. Note that they have different spectral distributions and different pulse broadenings. The pulse with a trapezoid shape has the widest spectral bandwidth. This is a result of the sharp edges (steep rise and fall). The broader spectral distribution leads to a larger broadening and a smaller "eye" (see Figure 7.3). Generally, a Gaussian pulse shape has the best pulse width–spectral width product. This is consistent with the concept of minimum wave packet in physics.

The "eye" diagram is a method used by transmission and network system engineers to depict the degradation of digital binary pulses due to pulse spreading, distortion, and noise. The stream of received pulses is displayed on a storage oscilloscope whose horizontal scan is triggered in synchronism with the bit rate. The storage oscilloscope thus records multiple sequences of bits of zeros and/or ones (with spreading and possible distortion), adding up to an "eye" diagram. An eye diagram with large open eyes indicates a clear transmission with a low bit error rate.

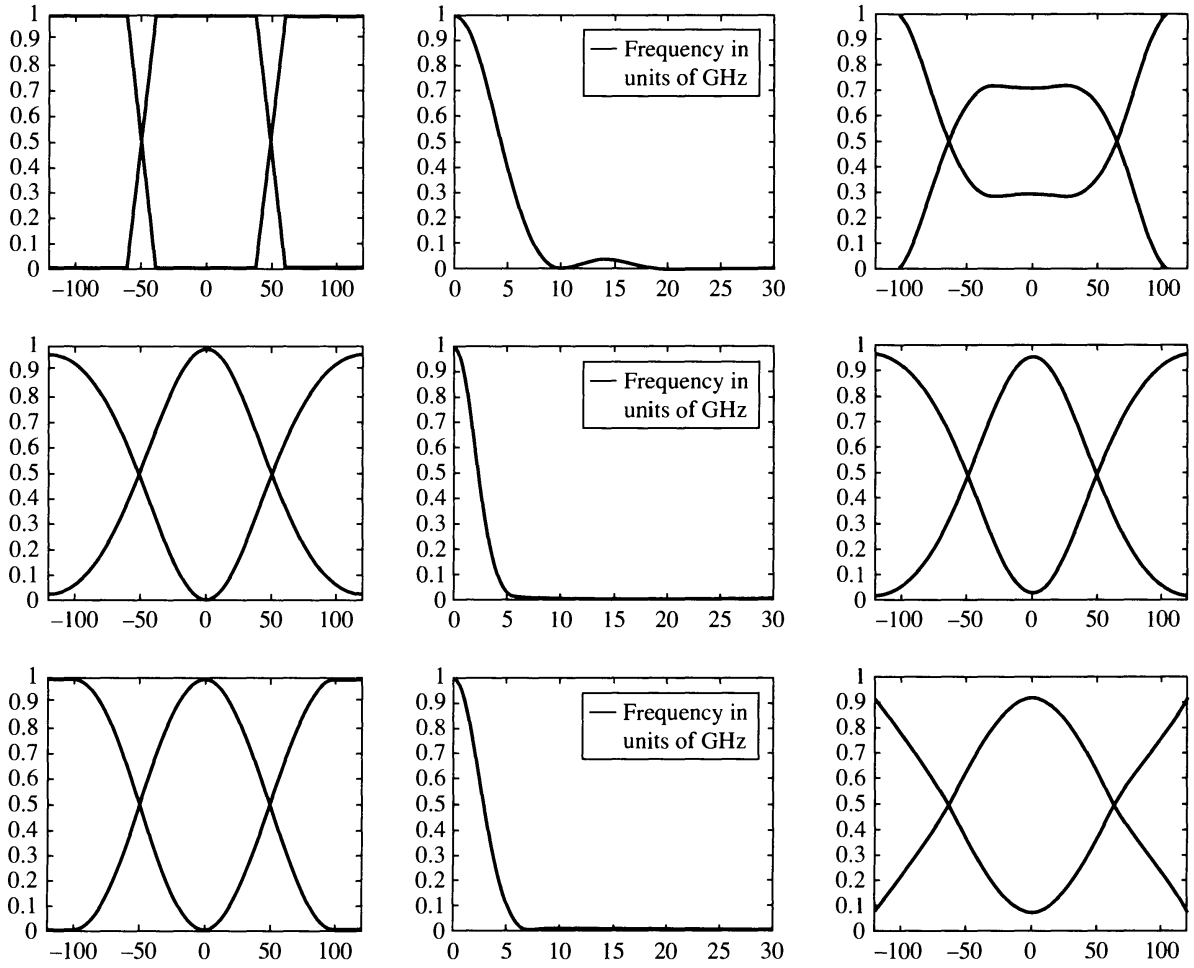


Figure 7.3 Eye diagrams of three different pulses. Initial pulse shape and eye diagram, spectral distributions, and final pulse shape and eye diagram after propagating through a dispersive medium of 800 ps/nm. Column 1—initial pulse shape; column 2—spectral distribution; column 3—final pulse shape after propagation. Row 1 is for an initial pulse shape of a trapezoid; row 2 is for an initial pulse of Gaussian shape; and row 3 is for an initial pulse of cosine shape. The time scale in columns 1 and 3 is in units of picosecond.

7.3 POLARIZATION EFFECTS IN OPTICAL FIBERS

We have demonstrated that chromatic dispersion can lead to the spreading of optical pulses. In this section, we show that polarization effects in fibers can also lead to the spreading of optical pulses. In Chapter 3, we discussed the modes of propagation in circular dielectric fibers. Specifically, we discussed linearly polarized modes in circular dielectric single-mode fibers. Despite their name, these "single-mode" fibers actually support two modes of propagation distinguished by their polarization states (LP_{01x} and LP_{01y}). As a result of the circular symmetry, these modes of propagation (LP_{01x} and LP_{01y}) are degenerate. In other words, at any given frequency they have exactly the same propagation constant and thus travel at exactly the same phase and group velocities.

In practice, all circular dielectric fibers are not exactly circular. The fiber cross section can be slightly flattened in one direction and slightly elongated in the other direction. This leads to a slightly elliptical cross section of the core. Figure 7.4 shows the polarization states of LP

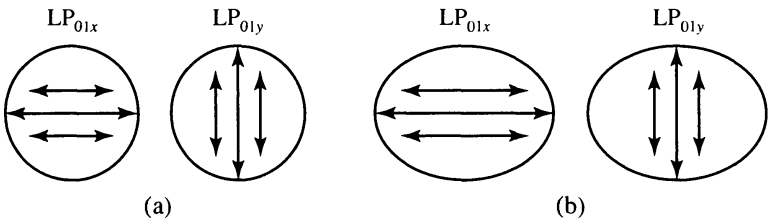


Figure 7.4 Polarization states of LP modes in optical fiber with an elliptical core. (a) In a circular fiber, the two modes are degenerate, that is, $\beta_{LP_{01x}} = \beta_{LP_{01y}}$ (b) In an elliptical fiber, $\beta_{LP_{01x}} \neq \beta_{LP_{01y}}$.

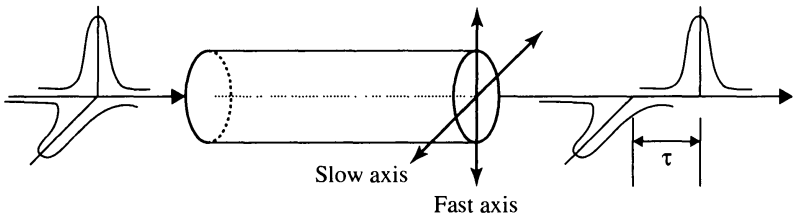


Figure 7.5 Pulse propagation through a uniform birefringent fiber. A time delay occurs between the slow and fast components of the pulse.

modes in an optical fiber with an elliptical core. The deviation from circular symmetry leads to a small difference in the propagation constant of the modes of propagation. In other words, the elliptical core leads to an optical birefringence in single-mode fibers. If the x and y axes coincide with the long and short axes of the elliptical core, then these two modes (LP_{01x} and LP_{01y}) travel with different group velocities. In practical fibers, the orientation of the long axis of the elliptical core may change with position along the fiber, as the fiberoptic cable twists and bends. Furthermore, the photoelastic effect due to stress along the fiber can cause a coupling between the modes of propagation. So, as a result of manufacturing imperfections and environmental perturbations (bending, twisting, etc.), an optical fiber actually behaves like a birefringent medium with the orientation of the slow axis that changes with position along the fiber. We will show that such a birefringence can lead to pulse distortion, broadening, and system impairments that limit the transmission capacity of the fiber.

Referring to Figure 7.5, we consider the simplest case of a single segment of fiber with a uniform birefringence. For a beam of unpolarized light, 50% of the energy is propagating along the fiber with polarization parallel to the slow axis of the fiber, and the other 50% is parallel to the fast axis of the fiber. As a result of the different speeds of propagation, a time delay occurs between these two components. Let n_s and n_f be the effective indices of the slow mode and the fast mode, respectively. It can easily be shown that the (group) time delay τ between these two components can be written (see Problem 7.6)

$$\tau = \left(\frac{d}{c} \Delta n + \frac{\omega d}{c} \frac{\partial \Delta n}{\partial \omega} \right) \tag{7.3-1}$$

where d is the length of the fiber segment and Δn is the birefringence given by

$$\Delta n = n_s - n_f \tag{7.3-2}$$

The spreading of optical pulses due to birefringence in fibers is known as polarization mode dispersion (PMD).

Mean Signal Delay and Broadening Due to Birefringence

Let the normalized input signal intensity be $I(t)$; then the output intensity can be written

$$I_{out}(t) = \gamma I(t - \tau_1) + (1 - \gamma)I(t - \tau_2) \tag{7.3-3}$$

where γ is the fraction of energy of the input wave along the direction of the slow axis and times τ_1 and τ_2 are the flight times of these modes, respectively. For simplicity, we assume $I(t)$ has a centroid at $t = 0$. In other words,

$$\langle tI(t)\rangle = 0 \tag{7.3-4}$$

In terms of the normalized Jones vector of the polarization state of the input wave $\mathbf{J}$, the fraction γ is given by

$$\gamma = |\mathbf{J}^* \cdot \mathbf{J}_{slow}|^2 \tag{7.3-5}$$

where $\mathbf{J}_{slow}$ is the normalized Jones vector of the polarization state of the slow mode. In investigating PMD, we often use the Poincaré sphere to represent the polarization states of a beam of polarized light. For polarized light, we can ignore the S_0 component. Thus the Stokes vector of a beam of polarized light is often written $\mathbf{s} = (S_1, S_2, S_3)$ as a real unit vector in Poincaré space. In terms of the Stokes vector of the input wave $\mathbf{s}$, the fraction γ is given by, according to Equation (1.6-41),

$$\gamma = \tfrac{1}{2}(1 + \mathbf{p} \cdot \mathbf{s}) \tag{7.3-6}$$

where $\mathbf{s}$ is the three-component Stokes vector of the polarization state of the input wave, and $\mathbf{p}$ is the three-component Stokes vector of the polarization state of the slow mode.

The mean signal delay due to birefringence is given by

$$\tau_g = \langle t\rangle = \int tI_{out}(t)\, dt = \gamma\tau_1 + (1 - \gamma)\tau_2 = (\gamma - \tfrac{1}{2})\tau \tag{7.3-7}$$

where we have taken $\tau_1 = \tau/2$ (slow mode) and $\tau_2 = -\tau/2$ (fast mode) with τ being the time delay between these two modes. For a beam of unpolarized light, $\gamma = \frac{1}{2}$, the mean delay is zero.

The mean delay can also be written, according to Equations (7.3-6) and (7.3-7),

$$\tau_g = \tfrac{1}{2}\boldsymbol{\tau} \cdot \mathbf{s} \tag{7.3-8}$$

where $\boldsymbol{\tau} = \tau\mathbf{p}$ is the so-called PMD vector. We will discuss the PMD vector in more detail in Section 7.5.

The pulse broadening can be evaluated by calculating the variance of the output intensity profile. Using Equation (7.3-3), we obtain

$$(\Delta\tau_b)^2 = \langle (t - \tau_g)^2\rangle - b^2 = \int (t - \tau_g)^2 I_{out}(t)\, dt - b^2 = \gamma(1 - \gamma)(\tau)^2 \tag{7.3-9}$$

where b is a measure of the original pulse width of the input signal $I(t)$,

$$b^2 = \langle t^2 I(t)\rangle = \int t^2 I(t)\, dt \tag{7.3-10}$$

In terms of the PMD vector $\boldsymbol{\tau} = \tau\mathbf{p}$ and the Stokes vector $\mathbf{s}$ of the input wave, the PMD broadening can be written, according to Equations (7.3-6) and (7.3-9),

$$(\Delta\tau_b)^2 = \tfrac{1}{4}[\tau^2 - (\boldsymbol{\tau} \cdot \mathbf{s})^2] \tag{7.3-11}$$

where $\mathbf{s}$ is the Stokes vector of the input wave. We note that when the input Stokes vector $\mathbf{s}$ is aligned along the PMD vector $\boldsymbol{\tau} = \tau\mathbf{p}$, the PMD broadening is zero. For an unpolarized wave $\gamma = 0.5$ and $\boldsymbol{\tau} \cdot \mathbf{s} = 0$, the PMD broadening is, according to Equation (7.3-9) or (7.3-11),

$$(\Delta\tau_b)^2 = \tfrac{1}{4}(\tau)^2 \tag{7.3-12}$$

We note that the pulse broadening due to birefringence in a short segment of fiber with uniform birefringence is proportional to the length of the fiber d. Although the result (7.3-11) is obtained by assuming a linear birefringence of the medium, it is actually valid for a general birefringent medium with elliptically polarized modes of propagation.

In real fiberoptic telecommunications systems, the fibers are much more complicated than a single segment of uniformly birefringent medium. We can try to understand the pulse broadening by dividing a long segment of fiber into a large number of small segments of uniformly birefringent fibers. Each segment is assumed to be uniformly birefringent with a fixed polarization state of the slow and fast modes. Based on the analysis above, each segment will introduce a broadening of the pulse. The total of the broadening is a result of the contributions from all the segments. In the simplest case, when the birefringence is completely random, the total broadening is a random walk process and can be written

$$(\Delta\tau_b)^2 = \sum_{i=1}^{N} \frac{1}{4}(\tau)_i^2 \tag{7.3-13}$$

$$= \sum_{i=1}^{N} \frac{1}{4}\left(\frac{\Delta n}{c} + \frac{\omega}{c}\frac{\partial \Delta n}{\partial \omega}\right)_i^2 d_i^2 \tag{7.3-14}$$

where d_i is the length of the ith segment. If we assume that all segments are of equal broadening (i.e., $\tau_i = \tau_1$ for all i), then it is clear from Equation (7.3-14), that the total broadening $\Delta\tau_b$ is proportional to the square root of the length of the fiber (see Problem 7.7). It is important to note that Equation (7.3-14) is valid only in the oversimplified case when the birefringence is completely random (see Problem 7.8). In practical fibers, although the birefringence may be random, the relative orientations of the slow modes (or fast modes) between neighboring elements are not completely random. A more rigorous analysis of PMD in multisegment uniform birefringent fibers will be discussed in Section 7.5. In the next section, we introduce the important concept of the principal states of polarization (PSP).

7.4 PRINCIPAL STATES OF POLARIZATION

For the purpose of analyzing the evolution of the polarization state, we may divide a long segment of fiber into a large number of small segments of uniformly birefringent fibers. Each segment is assumed to be uniaxially birefringent with a fixed orientation of the slow and fast axes. In such a birefringent network, the input and output polarization states can be obtained by using the Jones matrix method, which was discussed in detail in Chapter 1. Figure 7.6 shows

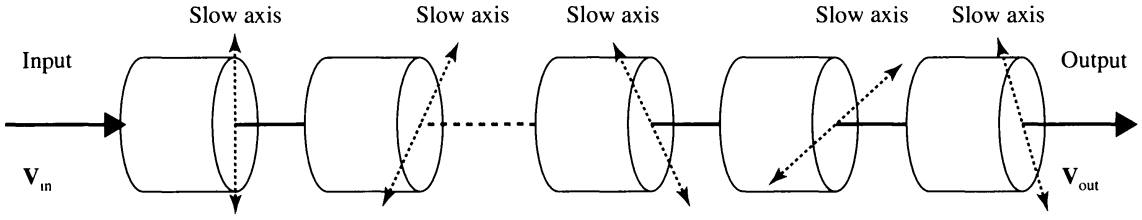


Figure 7.6 A long segment of fiber can be represented by an equivalent circuit that consists of a series of birefringent elements. The slow axis of each of the birefringent elements can be in an arbitrary orientation. The fast axis is perpendicular to the slow axis in each birefringent element.

a schematic drawing of the birefringent optical circuit representing a long segment of real fibers. In real fibers, the loss of circular symmetry can be a result of an elliptical core or cladding of the fibers, or a result of stress-induced anisotropy due to bending, twisting, and so on.

Given a fixed input state of polarization, the output state of polarization is in general a function of the frequency of the light. In what follows, we describe a set of particular states of input polarization, known as the principal states of polarization (PSP), such that the output states of polarization are insensitive to the variation of the frequency of the light. The concept of PSP was first introduced by Poole and Wagner in 1986 [1]. For the purpose of describing the concept of PSP, we assume that the fiber is lossless. Let $\mathbf{V}_{in}$ be the Jones vector of the input optical wave. The output polarization state $\mathbf{V}_{out}$ is given by

$$\mathbf{V}_{out} = M_N M_{N-1} \cdots M_3 M_2 M_1 \mathbf{V}_{in} = U\mathbf{V}_{in} \tag{7.4-1}$$

where M_m (m = 1, 2, 3, . . . , N) is the Jones matrix of the mth birefringent element. The overall Jones matrix U is the product of all the individual Jones matrices in sequence. For lossless fibers, all these Jones matrices are unitary. Thus we can write

$$\mathbf{V}_{out} = \begin{bmatrix} a & b \\ -b^* & a^* \end{bmatrix} \mathbf{V}_{in} \tag{7.4-2}$$

where a and b are elements of the overall Jones matrix U of the whole birefringent network. In practice, these matrix elements are obtained by multiplying all the individual Jones matrices in sequence. These elements (a and b) depend on the frequency of light, as well as the birefringence distribution of the fiber. The Jones matrix U in Equation (7.4-2) is unitary and thus

$$|a|^2 + |b|^2 = 1 \tag{7.4-3}$$

Generally, if the input polarization state $\mathbf{V}_{in}$ is fixed, the output polarization state $\mathbf{V}_{out}$ may depend on the frequency of light. The dependence can be investigated by examining the derivative of $\mathbf{V}_{out}$ in Equation (7.4-2) with respect to ω; in other words,

$$\frac{\partial \mathbf{V}_{out}}{\partial \omega} = \begin{bmatrix} a' & b' \\ -b'^* & a'^* \end{bmatrix} \mathbf{V}_{in} \tag{7.4-4}$$

where a prime indicates differentiation with respective to ω. In arriving at Equation (7.4-4), we assume a fixed input polarization state $\mathbf{V}_{in}$ over the spectral regime of interest.

If the output polarization state satisfies the following equation,

$$\frac{\partial \mathbf{V}_{out}}{\partial \omega} = -i\delta \mathbf{V}_{out} \tag{7.4-5}$$

where δ is a constant, then the Jones vector of the output beam $\mathbf{V}_{out}$ represents a polarization state that is independent of the frequency (to the first order in frequency). This can be seen by integrating the above equation with respect to the frequency ω. We obtain

$$\mathbf{V}_{out}(\omega) = \mathbf{V}_{out}(\omega_0) \exp\left(-i \int_{\omega_0}^{\omega} \delta \, d\omega\right) \tag{7.4-6}$$

where ω_0 is a constant, usually the carrier frequency. Note that the exponential multiplicative factor of a Jones vector does not change the polarization state. By examining Equation (7.4-6), we note that δ is actually the propagation time delay.

We now find the input polarization states $\mathbf{V}_{in}$ such that the output polarization state satisfies Equation (7.4-5). Substituting Equations (7.4-2) and (7.4-4) into Equation (7.4-5), we obtain after few steps of algebra

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \mathbf{V}_{\text{in}} = -i\delta \mathbf{V}_{\text{in}} \tag{7.4-7}$$

where

$$\begin{aligned} A &= a^* a' + b b'^* \\ B &= a^* b' - a'^* b \\ C &= a' b^* - a b'^* = -B^* \\ D &= a a'^* + b' b^* = A^* \end{aligned} \tag{7.4-8}$$

In other words, the input principal polarization states must be eigenvectors of the above matrix equation (7.4-7). When input polarization states are chosen so that Equation (7.4-7) is satisfied, the output polarization states will be independent of frequency to the first order in ω. Using Equations (7.4-3) and (7.4-8), it can be shown that $A + D = 0$. In other words, the trace of the matrix in Equation (7.4-7) is zero. This is a direct consequence of the unitary property of the Jones matrix and can easily be obtained by differentiating Equation (7.4-3) with respect to ω. Thus the matrix element A in Equation (7.4-7) is purely imaginary.

The eigenvectors and eigenvalues of Equation (7.4-7) are

$$\mathbf{V}_{\text{in}} = \begin{bmatrix} -B \\ A + i\delta \end{bmatrix} \tag{7.4-9}$$

and

$$\delta = \pm\sqrt{AD - BC} = \pm\sqrt{|a'|^2 + |b'|^2} \tag{7.4-10}$$

respectively. Let these two input principal states be designated as $\mathbf{V}_1$ and $\mathbf{V}_2$ with eigenvalues $\delta = \tau/2, -\tau/2$. Let τ_0 be the time of flight in the absence of the birefringence; then the time of flight of these two modes are, according to Equation (7.4-6),

$$\tau_0 + \tau/2 \quad \text{(slow mode)}$$

$$\tau_0 - \tau/2 \quad \text{(fast mode)}$$

respectively. These two eigenvectors are mutually orthogonal (i.e., $\mathbf{V}_1^* \cdot \mathbf{V}_2 = 0$). Using these two eigenvectors as the basis, a general input polarization state can be written

$$\mathbf{V}_{\text{in}} = c_1 \mathbf{V}_1 + c_2 \mathbf{V}_2 \tag{7.4-11}$$

where c_1 and c_2 are constants. The output polarization state can be written

$$\mathbf{V}_{\text{out}} = c_1 \mathbf{V}_1' + c_2 \mathbf{V}_2' \tag{7.4-12}$$

where $\mathbf{V}_1'$ and $\mathbf{V}_2'$ are the corresponding output principal polarization states (i.e., $\mathbf{V}_1' = U\mathbf{V}_1$ and $\mathbf{V}_2' = U\mathbf{V}_2$). It can also be written in terms of the frequency dependence, according to Equation (7.4-6),

$$\mathbf{V}_{\text{out}} = c_1 \mathbf{V}_1'(\omega_0) \exp\left(-i \int_{\omega_0}^{\omega} \frac{\tau}{2}\, d\omega\right) + c_2 \mathbf{V}_2'(\omega_0) \exp\left(i \int_{\omega_0}^{\omega} \frac{\tau}{2}\, d\omega\right) \tag{7.4-13}$$

The phase difference between these two polarization components can be written

$$\Delta\phi = -\int_{\omega_0}^{\omega} \tau\, d\omega \tag{7.4-14}$$

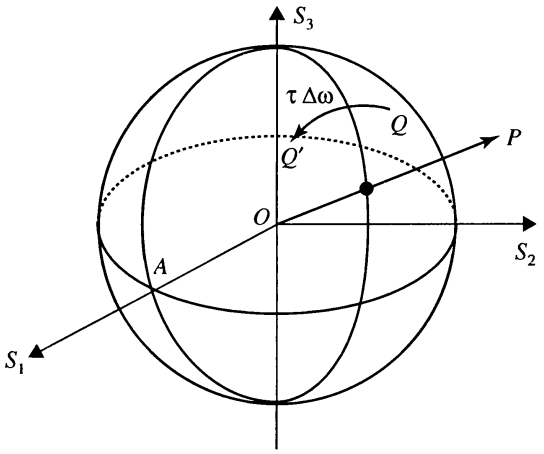


Figure 7.7 Schematic drawing showing the output state of polarization at various frequencies. Q is a point of the Poincaré sphere representing the output state $\mathbf{V}_{out}(\omega_0)$. Q' is a point of the Poincaré sphere representing the output state $\mathbf{V}_{out}(\omega_0 + \Delta\omega)$. OP is a unit vector along the PMD vector. Equivalently, OP is the Stokes vector of the slow PSP. Note that the output state Q' at $\omega_0 + \Delta\omega$ is obtained from the output state Q at ω_0 by rotating an angle of $\tau\,\Delta\omega$ around the PMD vector.

The group delay between these two components is given by, according to Equation (7.4-10),

$$\Delta\tau = -\frac{\partial}{\partial\omega}\Delta\phi = \tau = 2\sqrt{|a'|^2 + |b'|^2} \tag{7.4-15}$$

The group delay between these two polarization components is equivalent to a broadening of the pulse, according to the discussion in Section 7.3.

If the input state of polarization is aligned along one of the principal states of polarization, the broadening due to the birefringence disappears (to first order in frequency). We now examine the output state of polarization over a small frequency spread $\Delta\omega$ around ω_0. Using Equation (7.4-13), we can write the output state as

$$\mathbf{V}_{out}(\omega_0 + \Delta\omega) = c_1\mathbf{V}_1'(\omega_0)\exp(-i\tau\,\Delta\omega/2) + c_2\mathbf{V}_2'(\omega_0)\exp(i\tau\,\Delta\omega/2) \tag{7.4-16}$$

We note that the output polarization state at $\omega_0 + \Delta\omega$ can be obtained from the output polarization state at ω_0, by a rotation around the PMD vector on the Poincaré sphere with an angle of $\tau\,\Delta\omega$. This is illustrated in Figure 7.7.

The above treatment of PSP is based on the Jones matrix formulation. As we know, the polarization states of a beam of light can also be represented in terms of points on the Poincaré sphere. In this case, the output polarization state is obtained by a three-dimensional rotation of the input polarization state in Stokes space. Thus the same subject of PSP can be treated equivalently by using the Stokes formulation. This treatment will be described in Section 7.5.

EXAMPLE: SINGLE UNIFORM BIREFRINGENT FIBER

As an example to illustrate the concept of PSP, we consider the simplest case of a single segment of a birefringent medium with a uniform birefringence. For this case, the Jones matrix is given by

$$M = \begin{bmatrix} e^{-i\Gamma/2} & 0 \\ 0 & e^{i\Gamma/2} \end{bmatrix}$$

where

$$\Gamma = \frac{\omega}{c}(n_s - n_f)d = \frac{\omega}{c}\Delta n\, d$$

So $a = e^{-i\Gamma/2}$, $b = 0$, and

$$a' = -ia\frac{1}{2}\left(\frac{d}{c}\Delta n + \frac{\omega d}{c}\frac{d\Delta n}{d\omega}\right)$$

$$b' = 0$$

$$A = a^*a' + bb'^* = -i\frac{1}{2}\left(\frac{d}{c}\Delta n + \frac{\omega d}{c}\frac{d\Delta n}{d\omega}\right)$$

$$B = a^*b' - a'^*b = 0$$

The eigenvalues are, according to Equation (7.4-10),

$$\delta = +\tau/2 = \frac{1}{2}\left(\frac{d}{c}\Delta n + \frac{\omega d}{c}\frac{d\Delta n}{d\omega}\right) \quad \text{(slow mode)}$$

$$\delta = -\tau/2 = -\frac{1}{2}\left(\frac{d}{c}\Delta n + \frac{\omega d}{c}\frac{d\Delta n}{d\omega}\right) \quad \text{(fast mode)}$$

The principal states are simply the slow mode or the fast mode.

In other words,

$$\mathbf{V}_{in1} = \begin{bmatrix}1\\0\end{bmatrix}, \quad \text{with time of flight } \tau_1 = \tau_0 + \tau/2 \quad \text{(slow mode)}$$

$$\mathbf{V}_{in2} = \begin{bmatrix}0\\1\end{bmatrix}, \quad \text{with time of flight } \tau_2 = \tau_0 - \tau/2 \quad \text{(fast mode)}$$

And the PMD is

$$\Delta\tau = \left(\frac{d}{c}\Delta n + \frac{\omega d}{c}\frac{d\Delta n}{d\omega}\right)$$

Problem 7.9 deals with the PMD of a two-element birefringent fiber.

.5 VECTOR ANALYSIS OF POLARIZATION MODE DISPERSION

In this section, we describe the polarization mode dispersion (PMD) using both the Jones vector and Stokes vector representations. Pauli spin matrices are employed to provide a convenient relationship between these two representations. The Stokes vectors are particularly useful for graphic representation of the variation of the polarization state of the output beam on the Poincaré sphere. We first introduce the definitions and notations of some of the important parameters used in polarization mode dispersion analysis. We then describe the relationship between the Jones vector representation and the Stokes vector representation. In the last section, we describe briefly the concept of principal states of polarization (PSP) by using the Jones matrix formulation. In this section, we introduce the Stokes vector treatment of the same. In addition, we also introduce important concepts, including the PMD vector and the birefringence vector, which are the main tools for analyzing PMD in fiberoptic networks. Near the end, we describe a dynamical equation for the evolution of the PMD vector in arbitrary birefringent fibers, as well as a concatenation rule for PMD vectors. The discussion in this section follows closely that of Reference [2].

Definitions and Notations

Coordinate

z	Direction of propagation along the axis of the fiber
x, y	Transverse coordinates
$\psi(x, y) \exp[i(\omega t - kz]$	Continuous wave with frequency ω and wavenumber k traveling along the fiber

Polarization State Representations and Vectors

$\|s\rangle$	Dirac notation of 2-D complex Jones vector, defined in Chapter 1; using complex numbers, $\|s\rangle$ can be written as a column vector of (s_x, s_y)
$\mathbf{s}$	3-D real Stokes vector $\mathbf{s} = (s_1, s_2, s_3)$, defined in Chapter 1; in terms of the Jones vector elements, the Stokes vector $\mathbf{s}$ can be written $s_1 = s_x s_x^* - s_y s_y^*, \quad s_2 = s_x s_y^* + s_y s_x^*, \quad s_3 = i(s_x s_y^* - s_y s_x^*)$ The definition here is consistent with that of Chapter 1
$\|p\rangle$	Unit Jones vector of slow output PSP
$\|p_-\rangle$	Unit Jones vector of fast output PSP
$\mathbf{p}$	Unit Stokes vector along the direction of slow output PSP
$\mathbf{p}_-$	Unit Stokes vector along the direction of fast output PSP, $\mathbf{p}_- = -\mathbf{p}$
$\mathbf{r}$	Axis of rotation of a rotation matrix $\mathbf{R}$ (3×3) in Stokes space, $\mathbf{r}_- = -\mathbf{r}$
$\boldsymbol{\tau}$	Output PMD vector in Stokes space: its length is $\Delta\tau$, the differential group delay (DGD) between the two PSP, and its direction is that of the Stokes vector of the slow output PSP, that is, $\boldsymbol{\tau} = \Delta\tau\, \mathbf{p}$.
$\boldsymbol{\beta}$	3-D birefringence vector in Stokes space describing local birefringence properties; its length is the angle of rotation on a Poincaré sphere per unit length of fiber, while its direction is the axis of rotation of R

Matrices

I	Identity matrix
U	2×2 Jones matrix Input–output relationship: $\|v\rangle = U\|s\rangle$ U is unitary and can be written $U = \begin{bmatrix} a & b \\ -b^* & a^* \end{bmatrix}$ with $\|a\|^2 + \|b\| = 1$
R	3×3 Rotation matrix in Stokes space isomorphic to U Input–output relationship: $\mathbf{v} = R\mathbf{s}$

$\sigma_1, \sigma_2, \sigma_3$ — 2×2 Pauli spin matrix

$$\sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$$

$\boldsymbol{\sigma}$ — Pauli spin vector in Stokes space
Using Pauli spin matrices, we can write

$$s_1 = \langle s|\sigma_1|s\rangle, \quad s_2 = \langle s|\sigma_2|s\rangle, \quad s_3 = \langle s|\sigma_3|s\rangle,$$

or symbolically

$$\mathbf{s} = \langle s|\boldsymbol{\sigma}|s\rangle$$

$|s\rangle\langle s|$ — Projection operator (matrix) in Jones space

$$|s\rangle\langle s| = \begin{bmatrix} s_x s_x^* & s_x s_y^* \\ s_y s_x^* & s_y s_y^* \end{bmatrix} = \tfrac{1}{2}(\mathrm{I} + \mathbf{s} \cdot \boldsymbol{\sigma})$$

ss — Projection operator (matrix) in Stokes space

$$\mathbf{ss} = \begin{bmatrix} s_1 s_1 & s_1 s_2 & s_1 s_3 \\ s_2 s_1 & s_2 s_2 & s_2 s_3 \\ s_3 s_1 & s_3 s_2 & s_3 s_3 \end{bmatrix}$$

In addition to the above notations, we also introduce the matrix representation of vector cross multiplication. This is purely a matter of mathematical convenience. Consider two vectors in Stokes space and their cross product:

$$\mathbf{A} = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}, \quad \text{and} \quad \mathbf{A} \times \mathbf{B} = \begin{bmatrix} A_2 B_3 - A_3 B_2 \\ A_3 B_1 - A_1 B_3 \\ A_1 B_2 - A_2 B_1 \end{bmatrix}$$

Note that the product can be conveniently written as a matrix–vector multiplication.

$$\mathbf{A} \times \mathbf{B} = \begin{bmatrix} A_2 B_3 - A_3 B_2 \\ A_3 B_1 - A_1 B_3 \\ A_1 B_2 - A_2 B_1 \end{bmatrix} = \begin{bmatrix} 0 & -A_3 & A_2 \\ A_3 & 0 & -A_1 \\ -A_2 & A_1 & 0 \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}$$

So it is convenient to define a matrix representation of the cross multiplication as

$$\mathbf{A}\times = \begin{bmatrix} 0 & -A_3 & A_2 \\ A_3 & 0 & -A_1 \\ -A_2 & A_1 & 0 \end{bmatrix}$$

Jones Vectors and Stokes Vectors

We now describe the relationship between the Jones vector representation and the Stokes vector representation. The Jones matrix is a 2×2 matrix that relates the output states of polarization to the input states of polarization when a beam of polarized light propagates through a birefringent network. The linear relationship can be written

$$|v\rangle = U|s\rangle \tag{7.5-1}$$

where $|s\rangle$ is an input polarization state, $|v\rangle$ is the output polarization state, and U is the 2×2 Jones matrix. We note that the vectors and matrix are complex. Equation (7.5-1) can be viewed as a transformation of the state of polarization by the birefringent network. The Jones matrix operation is equivalent to a "rotation" in the two-dimensional complex Jones space.

If we represent the states of polarization as unit vectors in Stokes space, then the transformation of the state of polarization can be depicted as a rotation in the three-dimensional Stokes space. Mathematically, this is written

$$\mathbf{v} = R\mathbf{s} \tag{7.5-2}$$

where $\mathbf{s}$ is the three-component column vector representing an input state of polarization and $\mathbf{v}$ is the three-component column vector representing the output state of polarization. R is a 3×3 matrix. Note that the Stokes vectors and matrix are all real. We recall that for a beam of polarized light, the Stokes parameter $s_0 = 1$. So we only need the three components (s_1, s_2, s_3) to describe the polarization state. Based on the definition of the Stokes vector components (defined in Chapter 1), they can be written conveniently in terms of the Jones vector and the Pauli spin matrices:

$$s_1 = \langle s|\sigma_1|s\rangle, \quad s_2 = \langle s|\sigma_2|s\rangle, \quad s_3 = \langle s|\sigma_3|s\rangle \tag{7.5-3}$$

or symbolically,

$$\mathbf{s} = \langle s|\boldsymbol{\sigma}|s\rangle \tag{7.5-4}$$

These two equations are the relationships between Jones vectors and Stokes vectors.

Obviously, the Jones matrix U and the rotation matrix R are also related. In fact, there is a unique one-to-one relationship between them. Given a 2×2 Jones matrix, we should be able to find the 3×3 rotation matrix R, or equivalently the axis of rotation and the angle of rotation in Stokes space. Let $\mathbf{r}$ be the axis of rotation of R in Stokes space (see Figure 7.8); then it's obvious that any input Stokes vector collinear with the rotation axis $\mathbf{r}$ will remain invariant upon transmission through the birefringent optical system (the fiber). Let $|r\rangle$ and $|r_\rangle$ be the Jones vectors corresponding to Stokes vectors $\mathbf{r}$ and $-\mathbf{r}$, then the corresponding Jones vectors must be eigenvectors of the 2×2 Jones matrix U. Thus we can write

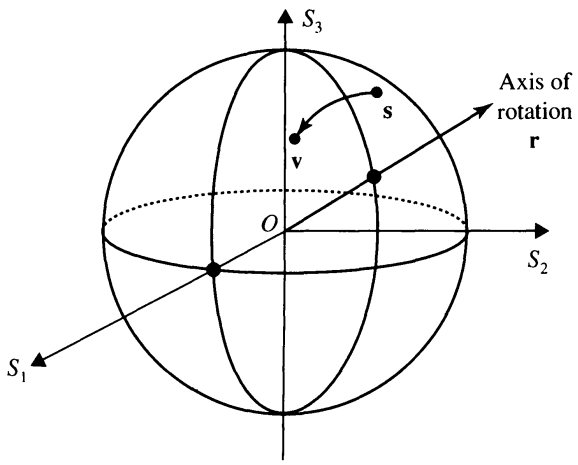


Figure 7.8 The transformation of polarization states by a birefringent network is equivalent to a rotation in Stokes space. The output Stokes vector $\mathbf{v}$ is obtained by a rotation of the input Stokes vector $\mathbf{s}$ around the axis $\mathbf{r}$. The axis of rotation and the angle of rotation are determined by the birefringent network.

$$U|r\rangle = e^{-i\gamma/2}|r\rangle \tag{7.5-5}$$

$$U|r_-\rangle = e^{+i\gamma/2}|r_-\rangle \tag{7.5-6}$$

where $\exp(-i\gamma/2)$ and $\exp(+i\gamma/2)$ are eigenvalues. Since U is both unitary and unimodular, the eigenvalues are unimodular and their product must be unity. According to Equations (7.5-5) and (7.5-6), the Stokes vector $\mathbf{r}$ corresponds to the polarization state of the slow mode of the birefringent network, whereas $-\mathbf{r}$ corresponds to the polarization state of the fast mode of the birefringent network.

Let $|s\rangle$ be an arbitrary input Jones vector. Since $|r\rangle$ and $|r_-\rangle$ constitute a complete orthogonal set of Jones vectors, we can write

$$|s\rangle = c_1|r\rangle + c_2|r_-\rangle \tag{7.5-7}$$

where c_1 and c_2 are constants. Upon transmission through the birefringent optical system, the output Jones vector can be obtained according to Equation (7.5-1). Using Equations (7.5-5) and (7.5-6), we obtain the output state of polarization

$$|v\rangle = c_1|r\rangle e^{-i\gamma/2} + c_2|r_-\rangle e^{+i\gamma/2} \tag{7.5-8}$$

Based on the discussion in Chapter 1, we recognize that the output Stokes vector $\mathbf{v}$ can be obtained by a right-handed rotation of the input Stokes vector $\mathbf{s}$ by an angle γ around the axis $\mathbf{r}$. Thus we are able to obtain the 3×3 rotation matrix R by solving a simple eigenvector problem in Jones space. Explicit expressions of the axis of rotation as well as the angle of rotation can be obtained in terms of the matrix elements of the Jones matrix. After solving the eigenvector problem (7.5-5) and (7.5-6), we obtain the following eigenvalues and eigenvectors:

$$\cos(\gamma/2) = \frac{a + a^*}{2} \tag{7.5-9}$$

$$|r\rangle = N\begin{bmatrix} -b \\ a - e^{-i\gamma/2} \end{bmatrix} \tag{7.5-10}$$

$$|r_-\rangle = N_-\begin{bmatrix} -b \\ a - e^{+i\gamma/2} \end{bmatrix} \tag{7.5-11}$$

where a and b are matrix elements of the Jones matrix, and N and N_- are normalization constants. Note that $|r\rangle$ is a two-component vector in Jones space. These two Jones vectors are mutually orthogonal. The corresponding vector $\mathbf{r} = (r_1, r_2, r_3)$ in Stokes space can be obtained by using Equation (7.5-3). This leads to

$$\begin{aligned} r_1 &= \frac{-2|a|^2 + ae^{+i\gamma/2} + a^*e^{-i\gamma/2}}{2 - ae^{+i\gamma/2} - a^*e^{-i\gamma/2}} \\ r_2 &= \frac{-(ab^* + a^*b) + be^{+i\gamma/2} + b^*e^{-i\gamma/2}}{2 - ae^{+i\gamma/2} - a^*e^{-i\gamma/2}} \\ r_3 &= \frac{-(ia^*b - iab^*) + ibe^{+i\gamma/2} - ib^*e^{-i\gamma/2}}{2 - ae^{+i\gamma/2} - a^*e^{-i\gamma/2}} \end{aligned} \tag{7.5-12}$$

These are explicit expressions of rotation angle γ and rotation axis $\mathbf{r}$ in terms of the matrix elements a and b of the Jones matrix.

EXAMPLE: AZIMUTH ANGLE OF $\psi = 0$

Consider a wave plate of uniaxial crystal with its c axis (or slow axis) oriented at an azimuth angle of $\psi = 0$. According to Chapter 1, the matrix elements are

$$a = e^{-i\Gamma/2}, \quad b = 0$$

Using Equation (7.5-12), we obtain

$$\gamma = \Gamma, \quad r_1 = 1, \quad r_2 = 0, \quad r_3 = 0$$

Note that the angle of rotation is exactly the phase retardation Γ, and the rotation axis is simply the s_1 axis.

EXAMPLE: AZIMUTH ANGLE OF $\psi \neq 0$

Consider now a wave plate of uniaxial crystal with its c axis (or slow axis) oriented at an azimuth angle of $\psi \neq 0$. According to Chapter 1, the matrix elements are

$$a = e^{-i\Gamma/2}\cos^2\Psi + e^{+i\Gamma/2}\sin^2\Psi, \quad b = -i\sin(\Gamma/2)\sin(2\Psi)$$

Using Equation (7.5-12), we obtain

$$\gamma = \Gamma, \quad r_1 = \cos(2\psi), \quad r_2 = \sin(2\psi), \quad r_3 = 0$$

Note that the angle of rotation is exactly the phase retardation Γ, and the rotation axis is on the equatorial plane, at an angle of 2ψ from the s_1 axis. Let R be the point on the Poincaré sphere representing the polarization state of the slow mode; then the output polarization state P' is obtained by rotating P an angle Γ around the axis OR, where O is the origin of the sphere. Figure 7.9 illustrates the polarization transformation on the Poincaré sphere.

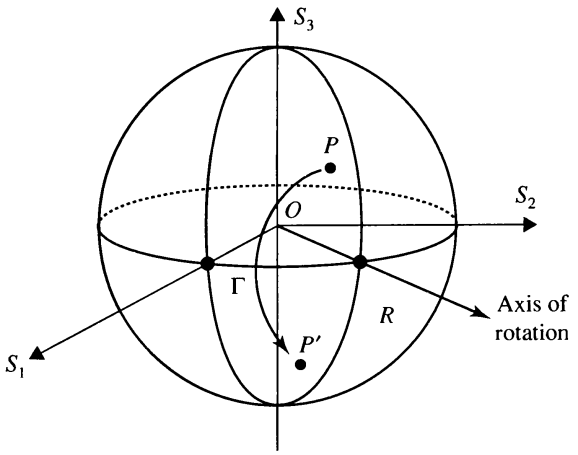


Figure 7.9 Polarization transformation on the Poincaré sphere. P is the input polarization state, P' is the output polarization state, and R is the polarization state of the slow mode of the wave plate.

Furthermore, we are able to obtain the 2×2 Jones matrix from the rotation angle γ and the axis of rotation $\mathbf{r}$. This is done as follows. We can write the Jones matrix in terms of the projection operators and eigenvalues as

$$U = e^{-i\gamma/2}|r\rangle\langle r| + e^{+i\gamma/2}|r_-\rangle\langle r_-| \tag{7.5-13}$$

Using the Pauli representation of the projection operators described in the beginning of this section, the Jones matrix (7.5-13) becomes

$$U = \cos(\gamma/2)I - i\sin(\gamma/2)\mathbf{r}\cdot\boldsymbol{\sigma} \tag{7.5-14}$$

where I is the identity matrix. Thus we can write the Jones matrix by using the explicit form of the Pauli matrices and the components of $\mathbf{r}$,

$$U = \begin{bmatrix} \cos(\gamma/2) - ir_1\sin(\gamma/2) & -(r_3 + ir_1)\sin(\gamma/2) \\ (r_3 - ir_1)\sin(\gamma/2) & \cos(\gamma/2) + ir_1\sin(\gamma/2) \end{bmatrix} \tag{7.5-15}$$

where r_1, r_2, r_3 are the components of $\mathbf{r}$ (the axis of rotation). This is an explicit expression of the Jones matrix in terms of the axis of rotation $\mathbf{r}$ and angle of rotation γ in Stokes space.

Conversely, we can obtain an explicit expression of the 3×3 rotation matrix R in terms of the Jones matrix. This is done as follows. Using Equations (7.5-1), (7.5-2), and (7.5-4), we can write

$$\mathbf{v} = R\mathbf{s} = R\langle s|\boldsymbol{\sigma}|s\rangle = \langle s|R\boldsymbol{\sigma}|s\rangle = \langle v|\boldsymbol{\sigma}|v\rangle = \langle s|U^{\dagger}\boldsymbol{\sigma}U|s\rangle \tag{7.5-16}$$

where the dagger ($\dagger$) stands for Hermitian conjugate (transpose and then complex conjugate). Since this is valid for an arbitrary Jones vector $|s\rangle$, we obtain

$$R\boldsymbol{\sigma} = U^{\dagger}\boldsymbol{\sigma}U \tag{7.5-17}$$

This is an expression of the 3×3 rotation matrix R in terms of the 2×2 Jones matrix U.

Substituting Equation (7.5-14) for U in Equation (7.5-17), using properties of the Pauli matrices, and after several algebraic steps (see Problem 7.11), we obtain

$$R = \mathbf{rr} + \sin(\gamma)\mathbf{r}\times - \cos(\gamma)(\mathbf{r}\times)(\mathbf{r}\times) \tag{7.5-18}$$

or equivalently,

$$R = \begin{bmatrix} r_1r_1 & r_1r_2 & r_1r_3 \\ r_2r_1 & r_2r_2 & r_2r_3 \\ r_3r_1 & r_3r_2 & r_3r_3 \end{bmatrix} + \sin(\gamma)\begin{bmatrix} 0 & -r_3 & r_2 \\ r_3 & 0 & -r_1 \\ -r_2 & r_1 & 0 \end{bmatrix} - \cos(\gamma)\begin{bmatrix} r_1r_1 - 1 & r_1r_2 & r_1r_3 \\ r_2r_1 & r_2r_2 - 1 & r_2r_3 \\ r_3r_1 & r_3r_2 & r_3r_3 - 1 \end{bmatrix} \tag{7.5-19}$$

This is an expression of the 3×3 rotation matrix in terms of the axis of rotation $\mathbf{r}$, and the angle of rotation γ in Stokes space. Table 7.1 lists some examples of Jones matrices and the corresponding 3×3 rotation matrices and the angles of rotation.

PSP (Principal States of Polarization)

We now describe the principal states of polarization in both the Jones space and the Stokes space. Consider the input of an optical wave with a fixed polarization state over a spectral regime of interest. We investigate the variation of the output polarization state as a function of frequency ω. This is done by examining the derivative of the output state with respect to frequency. From Equation (7.5-1), we obtain

$$\frac{\partial}{\partial\omega}|v\rangle = \left(\frac{\partial}{\partial\omega}U\right)|s\rangle \tag{7.5-20}$$

where $|s\rangle$ is a fixed input state of polarization and U is the Jones matrix of the birefringent network. For a general birefringent network, the Jones matrix U is a function of frequency. In general, the output state of polarization will vary with the frequency. We are interested

TABLE 7.1 Jones Matrices and Their Corresponding Rotation Matrices in Stokes Space

Physical Object	Jones Space: Jones Matrix	Stokes Space: Axis of Rotation	Stokes Space: Angle of Rotation	Stokes Space: 3 × 3 Rotation Matrix
Wave plate with phase retardation γ and azimuth angle[a] $\psi = 0$	$U = \begin{bmatrix} e^{-i\gamma/2} & 0 \\ 0 & e^{+i\gamma/2} \end{bmatrix}$	s_1	γ	$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{bmatrix}$
Wave plate with phase retardation γ and azimuth angle[a] $\psi = 45°$	$U = \begin{bmatrix} \cos(\gamma/2) & -i\sin(\gamma/2) \\ -i\sin(\gamma/2) & \cos(\gamma/2) \end{bmatrix}$	s_2	γ	$R = \begin{bmatrix} \cos\gamma & 0 & \sin\gamma \\ 0 & 1 & 0 \\ -\sin\gamma & 0 & \cos\gamma \end{bmatrix}$
Polarization rotator of angle $\gamma/2$	$U = \begin{bmatrix} \cos(\gamma/2) & -\sin(\gamma/2) \\ \sin(\gamma/2) & \cos(\gamma/2) \end{bmatrix}$	s_3	γ	$R = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$

[a] The azimuth angle ψ is defined as the angle of the slow axis measured from the x axis.

in a set of particular input polarization states such that the output states of polarization are insensitive to the variation of frequency (to the first order). Mathematically, this is written

$$\frac{\partial}{\partial\omega}|v\rangle = -i\delta|v\rangle \tag{7.5-21}$$

where δ is a constant. Physically, δ is the group delay due to propagation. According to this equation, the variation in frequency only leads to a phase factor that does not change the state of polarization of an output Jones vector. The states that satisfy the above equation are known as the principal states of polarization (PSP). These states must satisfy the following equation, according to Equations (7.5-1), (7.5-20), and (7.5-21):

$$\left(\frac{\partial}{\partial\omega}U\right)U^{-1}|v\rangle = -i\delta|v\rangle \tag{7.5-22}$$

or equivalently,

$$U'U^{-1}|v\rangle = \begin{bmatrix} a'a^* + b'b^* & ab' - a'b \\ a'^*b^* - a^*b'^* & a'^*a + b'^*b \end{bmatrix}|v\rangle = -i\delta|v\rangle \tag{7.5-23}$$

where a prime indicates differentiation with respect to frequency ω.

As a result of the unitary property of U, the matrix $U'U^{-1}$ has a zero trace. Furthermore, this matrix $U'U^{-1}$ is anti-Hermitian with pure imaginary eigenvalues (see Problem 7.12). So the eigenvalues can be written $-i\tau/2$ and $i\tau/2$, with real τ. The determinant of a square matrix is the product of the eigenvalues. So

$$\det(U'U^{-1}) = \frac{\tau^2}{4} \tag{7.5-24}$$

Since U is unitary, we have

$$\det(U') = \frac{\tau^2}{4} = |a'|^2 + |b'|^2 \tag{7.5-25}$$

The time delay between these two PSP can thus be written, according to Equations (7.5-24) and (7.5-25),

$$\tau = 2\sqrt{\det(U')} = 2\sqrt{|a'|^2 + |b'|^2} \tag{7.5-26}$$

The Jones vectors of the output PSP can be written, according to Equation (7.5-23),

$$\begin{aligned} |p\rangle &= N_- \begin{bmatrix} a'b - b'a \\ a'a^* + b'b^* + i\dfrac{\tau}{2} \end{bmatrix} \quad \text{(slow PSP with eigenvalue } \delta = +\tau/2) \\ |p_-\rangle &= N \begin{bmatrix} a'b - b'a \\ a'a^* + b'b^* - i\dfrac{\tau}{2} \end{bmatrix} \quad \text{(fast PSP with eigenvalue } \delta = -\tau/2) \end{aligned} \tag{7.5-27}$$

where N and N_- are normalization constants. These two output PSP are mutually orthogonal. The Jones vectors of the input PSP can be obtained by multiplying Equation (7.5-27) with U^{-1}.

Using these output PSP as the basis, any output state can be written

$$\begin{aligned} |v(\omega_0)\rangle &= c_1|p\rangle + c_2|p_-\rangle \\ |v(\omega)\rangle &= c_1|p\rangle e^{-i\Delta\omega\tau/2} + c_2|p_-\rangle e^{+i\Delta\omega\tau/2} \end{aligned} \tag{7.5-28}$$

where ω_0 and ω are frequencies, and $\Delta\omega = \omega - \omega_0$. By examining Equation (7.5-28), we note that $\mathbf{v}(\omega)$ is obtained by rotating $\mathbf{v}(\omega_0)$ around the axis $\mathbf{p}$ with an angle $\tau\,\Delta\omega$ in the Stokes space (see Figure 7.7).

It is important to note that this is only valid for a very small frequency range around ω_0. In most fibers, the principal states of polarization depend on the frequency. So this picture is valid only in a small frequency regime around ω_0 where the variation of the PSP with the frequency can be neglected.

Output PMD Vector $\boldsymbol{\tau}$

The output PMD vector is defined as a vector in the Stokes space, which is parallel to the Stokes vector of the slow output PSP mode, with a length of τ. For a single birefringent network, the group delay τ is enough to describe the broadening of the signal. When several birefringent networks are present in series, the total group delay due to PMD is not a simple sum of the individual group delays. The total group delay is a complicated vector summation of the PMD vectors of the individual birefringent elements. This will be discussed near the end of this section. For that purpose, we need to introduce the concept of PMD vector. The PMD vector can be obtained by using the Jones vectors of the output PSP as follows:

$$\boldsymbol{\tau} = \tau\mathbf{p} = \tau\langle p|\boldsymbol{\sigma}|p\rangle \tag{7.5-29}$$

where the group delay τ is given by Equation (7.5-26), and the Jones vector of the slow output PSP $|p\rangle$ is given by Equation (7.5-27). For a uniform birefringent fiber or a wave plate, the PMD vector is parallel to the polarization direction of the slow mode, with a magnitude given by Equation (7.3-1).

We can also obtain the PMD vector by using the projection operators and Pauli matrix representation. The operator (or matrix) can be represented by the projection operators of the eigenvectors as follows:

$$U'U^{-1} = -i\frac{\tau}{2}|p\rangle\langle p| + i\frac{\tau}{2}|p_-\rangle\langle p_-| \tag{7.5-30}$$

where $-i\tau/2$ and $+i\tau/2$ are eigenvalues of $U'U^{-1}$, and $|p\rangle$ and $|p_-\rangle$ are the corresponding eigenvectors. Furthermore, the projection operators can be written, according to definitions in this section,

$$\begin{aligned} |p\rangle\langle p| &= \tfrac{1}{2}(I + \mathbf{p}\cdot\boldsymbol{\sigma}) \\ |p_-\rangle\langle p_-| &= \tfrac{1}{2}(I + \mathbf{p}_-\cdot\boldsymbol{\sigma}) = \tfrac{1}{2}(I - \mathbf{p}\cdot\boldsymbol{\sigma}) \end{aligned} \tag{7.5-31}$$

Thus we obtain from Equations (7.5-30) and (7.5-31)

$$U'U^{-1} = -\frac{i}{2}\boldsymbol{\tau}\cdot\boldsymbol{\sigma} \tag{7.5-32}$$

or equivalently,

$$2iU'U^{-1} = \boldsymbol{\tau}\cdot\boldsymbol{\sigma} = \begin{bmatrix} \tau_1 & \tau_2 - i\tau_3 \\ \tau_2 + i\tau_3 & -\tau_1 \end{bmatrix} \tag{7.5-33}$$

The components of the PMD vector can easily be obtained by examining the matrix elements of the operator $2iU'U^{-1}$. The results are, according to Equation (7.5-23),

$$\begin{aligned} \tau_1 &= 2i(a'a^* + b'b^*) \\ \tau_2 &= 2\,\mathrm{Im}(a'b - ab') \\ \tau_3 &= 2\,\mathrm{Re}(a'b - ab') \end{aligned} \tag{7.5-34}$$

These are explicit expressions of the PMD vector in terms of the matrix elements (a, b) of the Jones matrix U. Once the output PMD vector is obtained, the input PMD vector can be obtained by $\boldsymbol{\tau}_{\text{in}} = R^{-1}\boldsymbol{\tau}$, where R is the rotation matrix representing the polarization transformation due to the birefringent network in the Stokes space. Note that $|\boldsymbol{\tau}_{\text{in}}| = |\boldsymbol{\tau}|$.

EXAMPLE: UNIFORM BIREFRINGENT FIBER

As an example to illustrate the concept of the PMD vector, we consider the simplest case of a single segment of a birefringent medium with a uniform birefringence. For this case, the Jones matrix is given by

$$M = \begin{bmatrix} e^{-i\Gamma/2} & 0 \\ 0 & e^{i\Gamma/2} \end{bmatrix}$$

where

$$\Gamma = \frac{\omega}{c}(n_s - n_f)d = \frac{\omega}{c}\Delta n\, d$$

So $a = e^{-i\Gamma/2}$, $b = 0$, and

$$a' = -ia\frac{1}{2}\left(\frac{d}{c}\Delta n + \frac{\omega d}{c}\frac{d\Delta n}{d\omega}\right)$$

$$b' = 0$$

We obtain, according to Equation (7.5-34),

$$\tau_1 = 2i(a'a^* + b'b^*) = \left(\frac{d}{c}\Delta n + \frac{\omega d}{c}\frac{\partial \Delta n}{\partial \omega}\right)$$

$$\tau_2 = 2\,\mathrm{Im}(a'b - ab') = 0$$

$$\tau_3 = 2\,\mathrm{Re}(a'b - ab') = 0$$

We note that the PMD vector is parallel to the S_1 axis (or the slow axis of the birefringent medium), with a magnitude equal to the group delay between the polarization modes.

Infinitesimal Rotations and Differential Equations

It is known from the mechanics of rigid bodies that the law of rotations becomes particularly simple when the rotations are infinitesimally small. We consider an infinitesimally small segment of birefringent fiber. It is reasonable to believe that the angle of rotation in the Stokes space is proportional to the length of the fiber dz. In other words, we can write

$$\gamma = \beta\, dz = k(n_s - n_f)\, dz = k\,\Delta n\, dz \tag{7.5-35}$$

where $\beta = k\,\Delta n$ is a measure of the birefringence, and n_s and n_f are the refractive indices of the slow mode and fast mode, respectively. The Jones matrix responsible for the input–output relationship can be written, according to Equation (7.5-14),

$$U = I - \frac{i}{2}\beta\, dz\, \mathbf{r}\cdot\boldsymbol{\sigma} \tag{7.5-36}$$

where $\mathbf{r}$ is the axis of rotation in Stokes space. The 3×3 rotation matrix can be written explicitly as, according to Equation (7.5-19),

$$R = I + \beta\, dz \begin{bmatrix} 0 & -r_3 & r_2 \\ r_3 & 0 & -r_1 \\ -r_2 & r_1 & 0 \end{bmatrix} = I + \beta\, dz\, \mathbf{r}\times \tag{7.5-37}$$

Let the output Jones vector after the infinitesimal rotation be written $|v\rangle = |s\rangle + d|s\rangle$, where $d|s\rangle$ is a differential of the Jones vector, representing the change of the polarization state after transmission through the birefringent element. Using these matrices, the input–output relationship can be written

$$|s\rangle + d|s\rangle = U|s\rangle = |s\rangle - \frac{i}{2}\beta\, dz\, \mathbf{r}\cdot\boldsymbol{\sigma}|s\rangle \tag{7.5-38}$$

or equivalently,

$$\frac{d}{dz}|s\rangle = -\frac{i}{2}\boldsymbol{\beta}\cdot\boldsymbol{\sigma}|s\rangle \tag{7.5-39}$$

where

$$\boldsymbol{\beta} = \beta\mathbf{r} = k\,\Delta n\,\mathbf{r} \tag{7.5-40}$$

is known as the birefringence vector. The birefringence vector, a real vector in Stokes space, is defined at each location of a fiber. For each location, a differential element dz of the fiber between z and $z + dz$ is considered as a birefringent element with an infinitesimal rotation in Stokes space. The birefringence vector is parallel to the axis of rotation in the Stokes

space corresponding to the polarization transformation by the differential element, while the magnitude of the vector is the product of the wavenumber with the local birefringence, $k\,\Delta n$. For a uniform birefringent fiber (or a wave plate), the birefringence vector is parallel to the polarization direction (in Stokes space) of the slow mode. With this definition, a birefringent property of a fiber is completely determined if $\boldsymbol{\beta} = \boldsymbol{\beta}(z)$ is given for all z.

Using the rotation matrix R, the input–output relationship in Stokes space can be written, according to Equation (7.5-37),

$$\mathbf{s} + d\mathbf{s} = R\mathbf{s} = \mathbf{s} + \boldsymbol{\beta}\, dz\, \mathbf{r} \times \mathbf{s} \tag{7.5-41}$$

or equivalently,

$$\frac{\partial}{\partial z}\mathbf{s} = \boldsymbol{\beta} \times \mathbf{s} \tag{7.5-42}$$

This equation is particularly useful. If the birefringence vector $\boldsymbol{\beta}(z)$ of a fiber is known, this equation can be integrated to obtain the output polarization state at any location z in the fiber.

We now consider the case of an input beam of light with a fixed state of polarization, over a small spectral regime of interest. As we have discussed earlier, the output state of polarization depends on the frequency. We now investigate the change of the state of polarization of the output beam due to an infinitesimal increase of frequency $d\omega$. The change in the state of output polarization, denoted as $d|s\rangle$, can be described as an infinitesimal rotation in both the Jones space and the Stokes space. In the Jones space, the change of output polarization state $d|s\rangle$ can be written, according to Equations (7.5-22) and (7.5-32),

$$d|s\rangle = \left(\frac{\partial}{\partial\omega}U\right)U^{-1}|s\rangle\, d\omega = -\frac{i}{2}\boldsymbol{\tau}\, d\omega \cdot \boldsymbol{\sigma}|s\rangle \tag{7.5-43}$$

or equivalently,

$$\frac{\partial}{\partial\omega}|s\rangle = -\frac{i}{2}\boldsymbol{\tau} \cdot \boldsymbol{\sigma}|s\rangle \tag{7.5-44}$$

where $\boldsymbol{\tau}$ is the PMD vector Here we are interested in the description of the output polarization state $|v\rangle$ as a continuous function of both z and ω. It suffices to use the notation $|s\rangle \equiv |s(z, \omega)\rangle$ as the Jones vector of the polarization state at any point in the fiber. Following a similar mathematical argument leading from Equation (7.5-39) to Equation (7.5-42), we obtain

$$\frac{\partial}{\partial\omega}\mathbf{s} = \boldsymbol{\tau} \times \mathbf{s} \tag{7.5-45}$$

This differential equation describes the change of the state of polarization of the output beam in Stokes space as a function of frequency given an arbitrary input state of polarization. If $\boldsymbol{\tau}$ is a constant, Equation (7.5-45) can easily be integrated. The solution $\mathbf{s}(\omega)$ is obtained by rotating $\mathbf{s}(\omega_0)$ around $\boldsymbol{\tau}$ by an angle $\tau(\omega - \omega_0)$. This is consistent with the description in Figure 7.7.

The Dynamical PMD Equation

In analyzing the pulse broadening in birefringent fibers, it is desirable to obtain a differential equation for the evolution of the PMD vector along the fibers. As we recall, the magnitude of the PMD vector gives information about the broadening. We now take a partial derivative with respect to ω on both sides of Equation (7.5-42), and a partial derivative with respect to z on both sides of Equation (7.5-45). By eliminating $\partial^2\mathbf{s}/\partial z\, \partial\omega$, we obtain

$$\frac{\partial}{\partial z}(\boldsymbol{\tau} \times \mathbf{s}) = \frac{\partial}{\partial \omega}(\boldsymbol{\beta} \times \mathbf{s}) \tag{7.5-46}$$

Using a vector identity $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{c} \cdot \mathbf{a}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b})$, we obtain, after few mathematical steps,

$$\frac{\partial}{\partial z}\boldsymbol{\tau} = \frac{\partial}{\partial \omega}\boldsymbol{\beta} + \boldsymbol{\beta} \times \boldsymbol{\tau} \tag{7.5-47}$$

This is the differential equation describing the evolution of the PMD vector along the fiber. The same equation has been employed as a basis for the statistical theory of PMD. Given a birefringence vector $\boldsymbol{\beta}(z, \omega)$, Equation (7.6-47) can be employed to obtain the output PMD vector.

PMD Concatenation

The PMD vector defined earlier is an important parameter of a birefringent network. It contains information about the output PSP as well as the group delay between the two PSP. The PMD vector of a sequence of birefringent networks is related to the PMD vectors of each element of the networks. Here we describe a method of obtaining the PMD vector of a sequence of birefringent networks.

Referring to Figure 7.10, we consider a single section of a birefringent fiber. The input–output relationship can be written

$$\mathbf{v} = \mathbf{s}(z) = R\mathbf{s}(0) \tag{7.5-48}$$

where R is the 3×3 rotation matrix. The PMD vector $\boldsymbol{\tau}$ can be obtained in two ways:

1. The Jones matrix approach.
2. The PSP approach.

The Jones Matrix Approach

We start from the Jones equation,

$$|v\rangle = |s(z)\rangle = U|s\rangle \tag{7.5-49}$$

where U is the Jones matrix. Once the Jones matrix is obtained, the PMD vector can be obtained by using Equation (7.5-34).

Alternately, we can find the eigenvectors of U. The Stokes vectors of these two eigenvectors determine the axis of rotation of R, while the phases of the eigenvalues determine the angle of rotation. These eigenvectors and eigenvalues uniquely determine the 3×3 rotation matrix R. Once R is obtained by using Equation (7.5-19), the PMD vector can be obtained as follows. Using Equation (7.5-45), we have

$$\frac{\partial \mathbf{s}(z)}{\partial \omega} = \boldsymbol{\tau} \times \mathbf{s}(z) = \boldsymbol{\tau} \times R\mathbf{s}(0) = \frac{\partial R}{\partial \omega}\mathbf{s}(0) \tag{7.5-50}$$

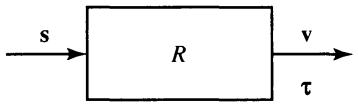


Figure 7.10 Linear input–output relationship of a birefringent fiber system. In the figure, R is the rotation matrix (3×3) that transforms the input Stokes vector $\mathbf{s}$ at $z = 0$ into the output Stokes vector $\mathbf{v} = \mathbf{s}(z)$ at z, and $\boldsymbol{\tau}$ is the PMD vector of the birefringent fiber system.

Since s(0) is arbitrary, we obtain

$$\frac{\partial R}{\partial \omega} = \boldsymbol{\tau} \times R$$

or equivalently,

$$\boldsymbol{\tau}\times = \frac{\partial R}{\partial \omega} R^{-1} \tag{7.5-51}$$

This equation can be employed to obtain all three components of the PMD vector in terms of the rotation matrix.

The PSP Approach

In the PSP approach, we first find the principal states of polarization by solving the eigenvector problem (7.5-23) and obtaining the eigenvectors and the eigenvalues. The magnitude of the PMD vector is the group delay between the two PSP modes, while the direction of the PMD vector is parallel to the polarization direction of the slow PSP mode. Or alternatively, the PMD vector can be obtained by using Equation (7.5-29), where $|p\rangle$ is the Jones vector of the slow PSP mode.

Once the PMD vector of a birefringent network is obtained, we can find the PMD vector of a series of birefringent networks. We start with the discussion of two concatenated birefringent networks.

PMD of Two Concatenated Sections

Referring to Figure 7.11, we consider two sections of birefringent elements in sequence. Let $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ be the PMD vectors of the individual sections.

Let $\boldsymbol{\tau}$ be the overall PMD vector for the whole system shown in Figure 7.9. We need to obtain the PMD vector $\boldsymbol{\tau}$ in terms of $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$. Let R be the rotation matrix of the overall system, $R = R_2R_1$. The PMD vector $\boldsymbol{\tau}$ can be written, according to Equation (7.5-51),

$$\boldsymbol{\tau}\times = \frac{\partial R}{\partial \omega} R^{-1}$$

Using $R = R_2R_1$, where R_1 and R_2 are the 3×3 rotation matrices of the individual sections, we obtain

$$\boldsymbol{\tau}\times = \frac{\partial R}{\partial \omega} R^{-1} = \frac{\partial (R_2R_1)}{\partial \omega}(R_2R_1)^{-1} = \frac{\partial R_2}{\partial \omega} R_1R_1^{-1}R_2^{-1} + R_2 \frac{\partial R_1}{\partial \omega} R_1^{-1}R_2^{-1} \tag{7.5-52}$$

Using $R_1R_1^{-1} = 1$ and $\boldsymbol{\tau}_1\times = (\partial R_1/\partial\omega)R_1^{-1}$, the above equation can be simplified to

$$\boldsymbol{\tau}\times = \frac{\partial R}{\partial \omega} R^{-1} = \frac{\partial R_2}{\partial \omega} R_2^{-1} + R_2\boldsymbol{\tau}_1 \times R_2^{-1} = \boldsymbol{\tau}_2\times + (R_2\boldsymbol{\tau}_1)\times \tag{7.5-53}$$

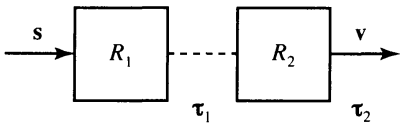


Figure 7.11 Concatenation of two sections of birefringent media.

where the last equality is obtained by using the fact that $R_2\boldsymbol{\tau}_1 \times R_2^{-1}$ is a transformation of the matrix $\boldsymbol{\tau}_1\times$ via a coordinate rotation R_2 and thus can be written $(R_2\boldsymbol{\tau}_1)\times$. So we obtain

$$\boldsymbol{\tau} = \boldsymbol{\tau}_2 + R_2\boldsymbol{\tau}_1 \tag{7.5-54}$$

This is the basic concatenation rule of PMD vectors. According to this equation, the total PMD vector is the sum of the PMD vector of the last element, plus the PMD vector of the previous element after being rotated by the 3 × 3 rotation matrix of the last element.

EXAMPLE: PMD VECTOR FOR *N* CONCATENATED SECTIONS

Consider the concatenation of three sections of birefringent fibers. Let $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$, and $\boldsymbol{\tau}_3$ be the PMD vectors of the individual sections. The overall PMD vector is obtained as follows:

$$\boldsymbol{\tau} = \boldsymbol{\tau}_3 + R_3(\boldsymbol{\tau}_2 + R_2\boldsymbol{\tau}_1) = \boldsymbol{\tau}_3 + R_3\boldsymbol{\tau}_2 + R_3R_2\,\boldsymbol{\tau}_1 \tag{7.5-55}$$

The basic concatenation rule can easily be extended to N sections of birefringent fibers.

$$\boldsymbol{\tau} = \boldsymbol{\tau}_N + R_N\boldsymbol{\tau}_{N-1} + R_NR_{N-1}\boldsymbol{\tau}_{N-2} + R_NR_{N-1}R_{N-2}\boldsymbol{\tau}_{N-3} + \cdots + R_NR_{N-1}\cdots R_2\boldsymbol{\tau}_1 \tag{7.5-56}$$

If we define

$$R(N, n) = R_NR_{N-1}R_{N-2}\cdots R_{n+1}R_n \tag{7.5-57}$$

then the concatenation rule for N sections can be written

$$\boldsymbol{\tau} = \sum_{n=1}^{N} R(N, n+1)\boldsymbol{\tau}_n \tag{7.5-58}$$

where $\boldsymbol{\tau}_n$ is the PMD vector of the nth section.

PMD Vector of an Infinitesimal Section

The PMD vector of an infinitesimal section of birefringent fiber is particularly simple.

According to Equation (7.5-37), the 3 × 3 rotation matrix of an infinitesimal section of birefringent fiber can be written

$$R = I + dz\,\boldsymbol{\beta}\times \tag{7.5-59}$$

where $\boldsymbol{\beta}$ is the birefringence vector of the element and dz is the length of the infinitesimal element. Taking the derivative with respect to frequency, we obtain

$$\frac{\partial}{\partial\omega}R = dz\frac{\partial\boldsymbol{\beta}}{\partial\omega}\times \tag{7.5-60}$$

The inverse rotation matrix can be written

$$R^{-1} = I - dz\,\boldsymbol{\beta}\times \tag{7.5-61}$$

Substituting Equations (7.5-60) and (7.5-61) into Equation (7.5-51), we obtain

$$\boldsymbol{\tau} = \frac{\partial\boldsymbol{\beta}}{\partial\omega}dz \tag{7.5-62}$$

This is the expression of the PMD vector of an infinitesimal element dz. The same result can also be easily obtained by a direct integration of the dynamical equation, and by neglecting

TABLE 7.2 Birefringence Vectors and PMD Vectors of Some Birefringent Elements

Physical Object	Birefringence Vector	PMD Vector
Wave plate with birefringence Δn, thickness d, and azimuth angle $\psi = 0$ (angle between slow axis and x axis)	$\boldsymbol{\beta} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} k\,\Delta n$	$\boldsymbol{\tau} = \tau \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$
Wave plate with birefringence Δn, thickness d, and azimuth angle ψ (angle between slow axis and x axis)	$\boldsymbol{\beta} = \begin{bmatrix} \cos(2\psi) \\ \sin(2\psi) \\ 0 \end{bmatrix} k\,\Delta n$	$\boldsymbol{\tau} = \tau \begin{bmatrix} \cos(2\psi) \\ \sin(2\psi) \\ 0 \end{bmatrix}$
Polarization rotator of angle $\gamma/2$ (circular birefringence Δn and thickness d)	$\boldsymbol{\beta} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} k\,\Delta n$	$\boldsymbol{\tau} = \tau \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

Note: Here $\tau = d\partial(k\,\Delta n)/\partial\omega$ is the group delay between the two polarization modes.

the second-order term $\boldsymbol{\beta} \times \boldsymbol{\tau}\, dz$. For a uniform birefringent fiber, the PMD vector and the birefringence vector are both parallel to the polarization direction of the slow mode. Table 7.2 is a list of PMD vectors and birefringence vectors of some simple birefringent elements.

Using Equation (7.5-62) for $\boldsymbol{\tau}_n$ in Equation (7.5-58), we obtain the following integral form of the concatenation rule:

$$\boldsymbol{\tau}(L) = \int_0^L R(L, z) \frac{\partial \boldsymbol{\beta}(z)}{\partial \omega} dz \tag{7.5-63}$$

where L is the length of the birefringent fiber, and $R(L, z)$ is the 3×3 rotation matrix corresponding to the segment between (z, L).

PMD Vector and the Broadening of Pulses

The PMD vector discussed earlier is a measure of the pulse broadening due to PMD in birefringent networks. As discussed in Section 7.3, the pulse broadening of a beam of light due to PMD in a birefringent network can be written in terms of the input PMD vector as, according to Equation (7.3-11),

$$(\Delta\tau_b)^2 = \tfrac{1}{4}[|\boldsymbol{\tau}_{\text{in}}|^2 - (\boldsymbol{\tau}_{\text{in}} \cdot \boldsymbol{s})^2] = \tfrac{1}{4}[|\boldsymbol{\tau}|^2 - (\boldsymbol{\tau}_{\text{in}} \cdot \mathbf{s})^2] \tag{7.5-64}$$

where $\boldsymbol{\tau}_{\text{in}}$ is the input PMD vector, $\boldsymbol{\tau} = R\boldsymbol{\tau}_{\text{in}}$ is the output PMD vector, s and is the Stokes vector of the input wave (note that $|\boldsymbol{\tau}_{\text{in}}| = |\boldsymbol{\tau}|$). We note that when the input Stokes vector s is aligned along the input PMD vector $\boldsymbol{\tau}_{\text{in}}$ (or the input PSP), the PMD broadening (first order) is zero. For a beam of unpolarized wave $\boldsymbol{\tau}_{\text{in}} \cdot \mathbf{s} = 0$, the PMD broadening is, according to Equation (7.5-64) or (7.3-11),

$$(\Delta\tau_b)^2 = \tfrac{1}{4}|\boldsymbol{\tau}|^2 \tag{7.5-65}$$

where $\boldsymbol{\tau}$ is the PMD vector of the birefringent network. If the birefringent network consists of a sequence of N birefringent elements (fibers), the PMD vector of the whole network can be obtained from the PMD vectors ($\boldsymbol{\tau}_n$, $n = 1, 2, 3 \ldots, N$) of the individual elements by using the concatenation rule (7.5-56) described earlier. To illustrate the concatenation rule of PMD vectors, we consider the following example.

EXAMPLE: TWO UNIFORM BIREFRINGENT ELEMENTS IN SEQUENCE

Referring to Figure 7.12, we consider the transmission of a beam of light through two uniform birefringent elements in sequence. Let the orientation of the slow axes be ψ_1, ψ_2 measured relative to the x axis. According to the discussion above and Table 7.2, the PMD vectors in Stokes space for each of the elements are

$$\boldsymbol{\tau}_1 = \tau_1 \begin{bmatrix} \cos(2\psi_1) \\ \sin(2\psi_1) \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\tau}_2 = \tau_2 \begin{bmatrix} \cos(2\psi_2) \\ \sin(2\psi_2) \\ 0 \end{bmatrix}$$

with $\tau_1 = d_1 \partial(k\,\Delta n_1)/\partial\omega$ and $\tau_2 = d_2 \partial(k\,\Delta n_2)/\partial\omega$, where Δn_1 and Δn_2 are birefringence, and d_1 and d_2 are lengths of the birefringent elements. These two PMD vectors lie in the equatorial plane of the Poincaré sphere. They are oriented at $2\psi_1$ and $2\psi_2$ from the s_1 axis (see Figure 7.12).

The overall PMD vector of the whole system is obtained by using the concatenation rule (7.5-54). We first rotate PMD vector $\boldsymbol{\tau}_1$ by an angle Γ_2 around the PMD vector $\boldsymbol{\tau}_2$ of the second element. The overall PMD vector is then obtained by a vector summation of $\boldsymbol{\tau}_2$ and $R_2\boldsymbol{\tau}_1$. This is illustrated in Figure 7.12. Following a simple algebra, we obtain the length of the overall PMD vector:

$$\tau^2 = \tau_1^2 + \tau_2^2 + 2\tau_1\tau_2 \cos[2(\psi_2 - \psi_1)] \tag{7.5-66}$$

The length of the overall PMD vector is a measure of the pulse broadening for a beam of unpolarized light, according to Equation (7.5-65). Depending on the relative orientation between the slow axes of the two birefringent elements, the length of the overall PMD vector can vary from $(\tau_1 + \tau_2)$ to $|\tau_1 - \tau_2|$, depending on the angle $(\psi_2 - \psi_1)$. If the slow axes of the second element is completely random relative to the orientation of the rotated PMD vector

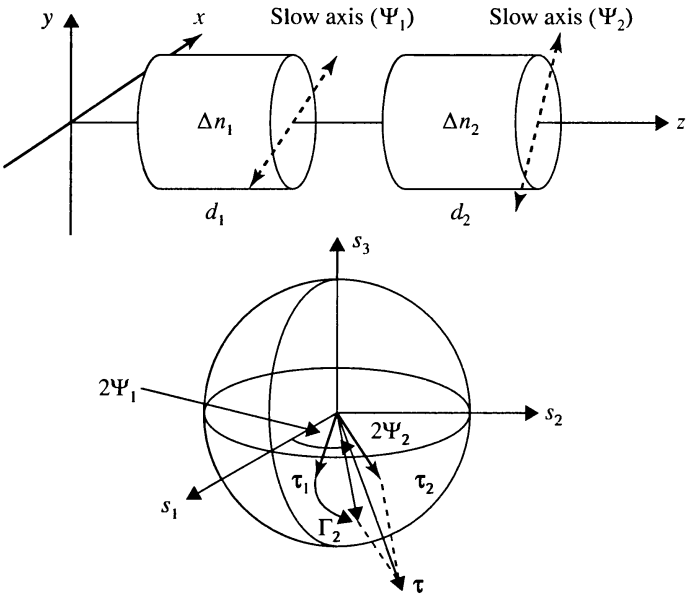


Figure 7.12 Schematic drawing of transmission of a beam of light through a series of two birefringent elements, with their slow axes oriented at azimuth angles ψ_1 and ψ_2, respectively. The lower figure is the Poincaré sphere representation of the PMD vectors, and the vector addition of $\boldsymbol{\tau}_2$ and $R_2\boldsymbol{\tau}_1$. R_2 is a rotation of angle Γ_2 around the vector $\boldsymbol{\tau}_2$, where $\Gamma_2 = d_2 k\,\Delta n_2$.

$R_2\boldsymbol{\tau}_1$, then the third term on the right side of Equation (7.5-66) will average out to zero. In this case, the pulse broadening is $(\tau_1^2 + \tau_2^2)/4$, according to Equation (7.6-65). In other words, the last element with a PMD vector $\boldsymbol{\tau}_2$ contributes an additional broadening of $\tau_2^2/4$. Based on this argument, the pulse broadening of a sequence of N birefringent elements with completely random orientation of the slow axes can be written

$$(\Delta\tau_b)^2 = \frac{1}{4}\sum_{n=1}^{N} |\boldsymbol{\tau}_n|^2 \tag{7.5-67}$$

where $\boldsymbol{\tau}_n$ is the PMD vector of the nth element. It is important to note that this is a purely hypothetical case. In real fiberoptic networks, the orientation of the slow axes of each of the birefringent elements may not be completely random.

7.6 HIGH-ORDER PMD AND COMPENSATORS

The principal states of polarization (PSP) described earlier are useful in the so-called first-order PMD compensator (PMDC). Figure 7.13 shows a schematic drawing of such a compensator. An input pulse of a beam of unpolarized light is split into two smaller pulses. The two smaller pulses are polarized along the fast PSP mode and the slow PSP mode, respectively. These two modes are mutually orthogonal in their polarization states that are generally elliptical. At the output of the birefringent network, these two pulses are separated by a polarization group delay. They are polarized along the output PSP directions. Using a quarter-wave plate at the appropriate orientation, these two polarized modes can be converted into linearly polarized states. A birefringent wave plate with the proper retardance can then be employed to compensate for the time delay between the two pulses. This leads to a single pulse at output and thus removes the pulse spreading (or splitting) due to PMD.

Polarization mode dispersion (PMD) is currently one of the primary obstacles to ultrahigh-speed optical communications systems. As discussed earlier in this chapter, the effects of PMD were traditionally described in terms of a pair of principal states of polarization (PSP) and the differential group delay (DGD) between them. At low communication speeds, the PSP and the DGD at the center of the signal spectrum (this is usually the carrier frequency) are sufficient to fully describe the effects of PMD over the spectrum of the transmitted signal. As the communication speed increases, the spectral bandwidth of the signals increases accordingly. The PSP and DGD at the center frequency are no longer a good representation

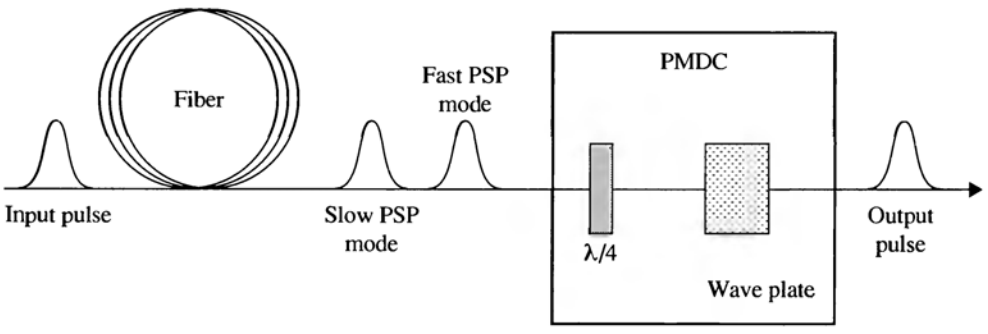


Figure 7.13 Schematic drawing of a first-order PMDC.

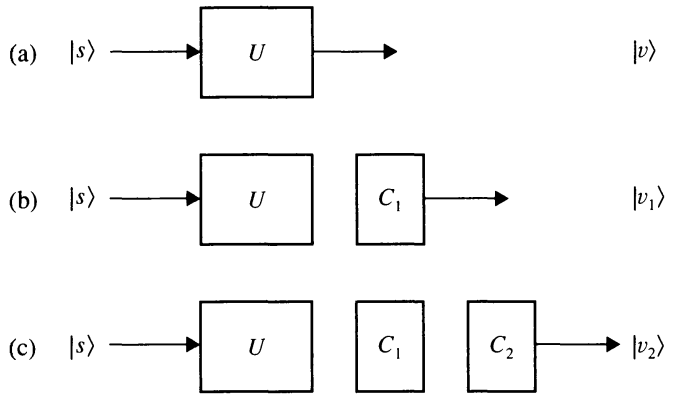


Figure 7.14 Schematic drawing of PMD compensators and the input–output Jones vectors. U is the Jones matrix of the fiber; C_1 and C_2 are the Jones matrices of PMD compensators.

of the PSP and DGD within the whole signal spectral bandwidth. Under the circumstances, we need to consider the second-order variation of the PSP and DGDs over the bandwidth [3–5]. Furthermore, we need to find compensators that can undo the PMD over the bandwidth for these high-speed signals.

Commonly termed as high-order PMD, the variation of first-order PSP and DGD over the bandwidth of high-speed signals necessitates the development of analytical tools for its description. Several approaches appear in the literature for the description of high-order PMD. The most common approach is based on the output PMD vector, $\boldsymbol{\tau}$, which in the absence of polarization-dependent loss (PDL) is a real, three-dimensional vector whose magnitude is equal to the DGD and whose orientation is parallel to the Poincaré sphere representation of the slow output PSP. Using this approach, the high-order PMD can be described in terms of the derivatives of the PMD vector with respect to frequency [2, 4].

Here we provide a general description of high-order PMD and discuss optical means to compensate for the high-order PMD. Using the Jones matrix method, we generalize the phenomenological approach of Poole and Wagner [1], by considering the high-order derivatives of the Jones matrix equation [6–8]. Referring to Figure 7.14, we consider a fiber whose birefringent property is described by a Jones matrix U. The input–output relationship is written

$$|v\rangle = U|s\rangle \tag{7.6-1}$$

where $|s\rangle$ is an arbitrary input Jones vector and $|v\rangle$ is the corresponding output Jones vector.

The first-order PSP satisfy the following vector equation:

$$\frac{\partial}{\partial\omega}|v\rangle = \left(\frac{\partial}{\partial\omega}U\right)|s\rangle = -i\delta_1|v\rangle \tag{7.6-2}$$

where $|s\rangle$ is a fixed input state of polarization and δ_1 is an eigenvalue. The DGD is the difference between the two eigenvalues $(\delta_{1+} - \delta_{1-})$. For the purpose of discussion, we use the following notations for the PSP and their eigenvalues:

$|p_{1+}\rangle$ Unit Jones vector of slow first-order output PSP with eigenvalue δ_{1+}

$|p_{1-}\rangle$ Unit Jones vector of fast first-order output PSP with eigenvalue δ_{1-}

Let C_1 represent the Jones matrix of a first-order PMD compensator. Then, by definition, C_1 must satisfy the following equation:

$$\frac{\partial}{\partial\omega}|v_1\rangle = \left(\frac{\partial}{\partial\omega}C_1U\right)|s\rangle = 0 \tag{7.6-3}$$

where $|s\rangle$ is a fixed input state of polarization and C_1U is the Jones matrix of the fiber plus the compensator (see Figure 7.14b). With the presence of the compensator C_1, the DGD of the overall system becomes zero. Carrying out the differentiation in Equation (7.6-3), we obtain

$$(C_1'U + C_1U')|s\rangle = 0 \tag{7.6-4}$$

where a prime indicates differentiation with respect to ω. To solve for the Jones matrix C_1, we let $|s\rangle$ be the input PSP. According to Equations (7.6-1) and (7.6-2), we have $U|s\rangle = |p_1\rangle$ and $U'|s\rangle = -i\delta_1|p_1\rangle$, where $|p_1\rangle$ is a first-order output PSP. The above equation becomes

$$[C_1' + C_1(-i\delta_1)]|p_1\rangle = 0 \tag{7.6-5}$$

The Jones matrix of the first-order PMD compensator C_1 can thus be written, using the first-order output PSP as the basis,

$$C_1 = \begin{bmatrix} \exp(-i\delta_{1+}\omega) & 0 \\ 0 & \exp(-i\delta_{1-}\omega) \end{bmatrix}_{\text{PSP}} \tag{7.6-6}$$

In xy coordinates, the Jones matrix C_1 can be obtained by using the following formula:

$$\langle i|C_1|j\rangle = \sum_{\alpha\beta} \langle i|\alpha\rangle\langle\alpha|C_1|\beta\rangle\langle\beta|j\rangle \quad \text{with } i, j = x, y \quad \text{and} \quad \alpha, \beta = |p_{1+}\rangle, |p_{1-}\rangle \tag{7.6-7}$$

The Jones matrix C_1 in xy coordinates can also be written

$$C_1 = \begin{bmatrix} p_{1+x} & p_{1-x} \\ p_{1+y} & p_{1-y} \end{bmatrix} \begin{bmatrix} \exp(-i\delta_{1+}\omega) & 0 \\ 0 & \exp(-i\delta_{1-}\omega) \end{bmatrix} \begin{bmatrix} p_{1+x} & p_{1+y} \\ p_{1-x} & p_{1-y} \end{bmatrix} \tag{7.6-8}$$

where p_{1+x}, p_{1+y}, p_{1-x}, p_{1+y} are xy components of the Jones vectors of the first-order output PSP. We note that the first matrix in the above equation consists of two column vectors of the first-order output PSP, and the third matrix consists of two row vectors of the first-order output PSP. The above equation is an explicit expression of the Jones matrix of the first-order PMD compensator.

We now consider the second-order effect. Referring to Figure 7.14b, the input–output Jones vector relationship is now written

$$|v_1\rangle = C_1U|s\rangle \tag{7.6-9}$$

With the presence of the first-order PMD compensator, all input states are first-order principal states and the DGD is zero to the first order. Since the DGD is zero for the first order, we consider the second-order effect. This is particularly important in high-speed transmission where the signal bandwidth is quite broad. The second-order PMD is described as the second-order differential variation of the output polarization states and the associated DGD. The second-order PSP satisfy the following vector equation:

$$\frac{\partial^2}{\partial\omega^2}|v_1\rangle = \left(\frac{\partial^2}{\partial\omega^2}C_1U\right)|s\rangle = -i\delta_2|v_1\rangle \tag{7.6-10}$$

where $|s\rangle$ is a fixed input state of polarization and δ_2 is an eigenvalue. Let the second-order output PSP and their eigenvalues be written

$|p_{2+}\rangle$ Unit Jones vector of slow second-order output PSP with eigenvalue δ_{2+}

$|p_{2-}\rangle$ Unit Jones vector of fast second-order output PSP with eigenvalue δ_{2-}

The DGD between the two eigenmodes is

$$\text{DGD} = (\delta_{2+} - \delta_{2-})(\omega - \omega_0) \tag{7.6-11}$$

where ω_0 is the center of the signal spectrum, and δ_{2+} and δ_{2-} are the eigenvalues.

Let C_2 represent the Jones matrix of the second-order PMD compensator (see Figure 7.14c). Then, by definition, C_2 must satisfy the following equation:

$$\frac{\partial^2}{\partial\omega^2}|v_2\rangle = \left(\frac{\partial^2}{\partial\omega^2}C_2M_1\right)|s\rangle = 0 \tag{7.6-12}$$

where $|s\rangle$ is a fixed input state of polarization and M_1 is the Jones matrix of the fiber plus the compensator (see Figure 7.14b),

$$M_1 = C_1U \tag{7.6-13}$$

With the presence of the compensator C_2, the second-order DGD of the overall system (Figure 7.14c) becomes zero.

Carrying out the differentiation in Equation (7.6-12), we obtain

$$(C_2''M_1 + 2C_2'M_1' + C_2M_1'')|s\rangle = 0 \tag{7.6-14}$$

To solve for the Jones matrix C_2, we let $|s\rangle$ be the input second-order PSP. According to Equations (7.6-3), (7.6-9), and (7.6-10), we have $M_1|s\rangle = |p_2\rangle$, $M_1'|s\rangle = 0$, and $M_1''|s\rangle = -i\delta_2|p_2\rangle$, where $|p_2\rangle$ is a second-order output PSP. Equation (7.6-14) becomes

$$(C_2'' + C_2(-i\delta_2))|p_2\rangle = 0 \tag{7.6-15}$$

The Jones matrix of the second-order PMD compensator C_2 can thus be written, using the second-order PSP as the basis,

$$C_2 = \begin{bmatrix} \exp(-i\frac{1}{2}\delta_{2+}\omega^2) & 0 \\ 0 & \exp(-i\frac{1}{2}\delta_{2-}\omega^2) \end{bmatrix}_{\text{PSP}} \tag{7.6-16}$$

This is the matrix in the coordinate system where the second-order output PSP are the axes.

Using Equation (7.6-7), we can write the Jones matrix C_2 in xy coordinates as

$$C_2 = \begin{bmatrix} p_{2+x} & p_{2-x} \\ p_{2+y} & p_{2-y} \end{bmatrix} \begin{bmatrix} \exp(-i\frac{1}{2}\delta_{2+}\omega^2) & 0 \\ 0 & \exp(-i\frac{1}{2}\delta_{2-}\omega^2) \end{bmatrix} \begin{bmatrix} p_{2+x} & p_{2+y} \\ p_{2-x} & p_{2-y} \end{bmatrix} \tag{7.6-17}$$

where p_{2+x}, p_{2+y}, p_{2-x}, p_{2+y} are the xy components of the Jones vectors of the second-order output PSP.

We note that the first matrix in the above equation consists of two column vectors of the second-order PSP, and the third matrix consists of two row vectors of the second-order PSP. The above equation is an explicit expression of the Jones matrix of the second-order PMD compensator. With the presence of both the first-order and second-order PMD compensator, the overall DGD is zero to the second order of variation in frequency. The second matrix in Equation (7.6-17) represents a wave plate with a DGD given by Equation (7.6-11). Such a wave plate would require a birefringent material with birefringence dispersion.

Once the second-order PMD is eliminated with the presence of the second-order PMD compensator, we can now move on to consider the third-order PMD effect. Similar analysis can be carried out, and the Jones matrix of the third-order PMD compensator can be obtained. The process can continue indefinitely until the effect of PMD is eliminated completely within the bandwidth of interest.

PROBLEMS

7.1 Show that in a quadratic index fiber in which g^2 is independent of ω, the spreading of a pulse of spectral width $\Delta\omega$ can be described by

$$\Delta\tau \equiv \frac{L}{c}\left|\frac{n^2 g^2}{ck^3}(l+m+1)^2 - \frac{dn}{d\omega}\right|\Delta\omega$$

where l and m are the transverse mode indices. In the expression, the term proportional to $(l+m+1)^2$ is due to modal dispersion, whereas the term proportional to $dn/d\omega$ is due to material dispersion.

7.2 Show that the FWHM bandwidth of the square of the Fourier spectrum (7.2-3) is given by

$$\Delta\Omega_{\text{FWHM}} = \frac{4\ln 2}{\tau_0} = 2\pi\,\Delta\nu_{\text{FWHM}}$$

where $\tau_0 = \sqrt{2\ln 2/\alpha}$ is the full width at half-maximum. Show that the same bandwidth can be written

$$\Delta\lambda = \frac{2\ln 2}{\pi c}\frac{\lambda^2}{\tau_0}$$

7.3 In optical communications, signal pulses are often in the shape of a trapezoid, as shown in Figure P7.3.

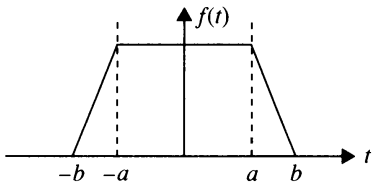


Figure P7.3

(a) Show that the pulse can be written analytically as

$$f(t) = \begin{cases} 1, & 0 < |t| < a \\ \dfrac{b-|t|}{b-a}, & a < |t| < b \\ 0, & \text{elsewhere} \end{cases}$$

(b) Show that the Fourier transform can be written

$$F(\Omega) = a\frac{\sin\Omega a}{\pi\Omega a} + b\frac{\sin\Omega b - \sin\Omega a}{\pi\Omega(b-a)} - \frac{\cos\Omega b - \cos\Omega a + \Omega b\sin\Omega b - \Omega a\sin\Omega a}{\pi\Omega^2(b-a)} = \frac{\cos\Omega a - \cos\Omega b}{\pi\Omega^2(b-a)}$$

(c) Show that, in the limit when $b = a$, the Fourier transform becomes

$$F(\Omega) = a\frac{\sin\Omega a}{\pi\Omega a}$$

(d) Assume the propagation of a trapezoid pulse with $a = 40$ ps and $b = 60$ ps in a single-mode fiber with $D = 17$ ps/nm-km. Estimate the pulse width at the end of a 100 km fiber. Try to find the pulse shape.

7.4 Consider a pulse given by

$$f(t) = \begin{cases} \cos\left(\dfrac{\pi t}{2b}\right), & -b < t < b \\ 0, & \text{elsewhere} \end{cases}$$

(a) Find the full width at half-maximum of the pulse intensity. [Answer: b.]

(b) Show that the Fourier transform is given by

$$F(\Omega) = \frac{1}{2\pi}\left(\frac{\sin(\Omega-\alpha)b}{(\Omega-\alpha)} + \frac{\sin(\Omega+\alpha)b}{(\Omega+\alpha)}\right)$$

where $\alpha = \pi/2b$.

(c) Assume propagation of a cosine pulse with $b = 100$ ps in a single-mode fiber with $D = 17$ ps/nm-km. Estimate the pulse width at the end of a 100 km fiber. Try to find the pulse shape.

7.5 Consider a general pulse $f(t)$ with a Fourier transform $F(\Omega)$:

$$f(t) = \int F(\Omega)e^{i\Omega t}\,d\Omega$$

Assume both the pulse function and its Fourier transform have a mean at zero. Let $(\Delta t)^2$ be the variance of $f(t)$ and $(\Delta\Omega)^2$ be the variance of $F(\Omega)$. They are defined as

$$(\Delta t)^2 = \frac{\int f^*(t)t^2 f(t)\,dt}{\int f^*(t)f(t)\,dt} \equiv \frac{\langle f|t^2|f\rangle}{\langle f|f\rangle},$$

$$(\Delta\Omega)^2 = \frac{\int F^*(\Omega)\Omega^2 F(\Omega)\,d\Omega}{\int F^*(\Omega)F(\Omega)\,d\Omega} \equiv \frac{\langle F|\Omega^2|F\rangle}{\langle F|F\rangle}$$

(a) Show that

$$\frac{\partial}{\partial t}f(t) = \int i\Omega F(\Omega)e^{i\Omega t}\,d\Omega$$

and

$$(\Delta\Omega)^2 = \frac{1}{\int f^*(t)f(t)\,dt}\int f^*(t)\left(-\frac{\partial^2}{\partial t^2}\right)f(t)\,dt$$

(b) Define the following pair of Hermitian operators:

$$A = \lambda t + i\Omega/\lambda, \quad A^\dagger = \lambda t - i\Omega/\lambda$$

where Ω is replaced with its operator $-i\,\partial/\partial t$. Show that

$$\langle f|A^\dagger A|f\rangle = \langle f|\lambda^2 t^2 - \frac{\partial^2}{\lambda^2\,\partial t^2} - 1|f\rangle$$

The above integral is positive definite for all real λ.

(c) Using (b), show that

$$\Delta t\,\Delta\Omega \geq \tfrac{1}{2}$$

This is the uncertainty principle.

(d) Show that the equality holds for Gaussian pulses.

(e) The standard deviations Δt and $\Delta\Omega$ are different from the FWHM values discussed in this chapter and Problem 7.2. Show that

$$\Delta t_{\text{FWHM}}\Delta\Omega_{\text{FWHM}} = 4\ln 2 = 2.77$$

7.6 Let n_s and n_f be the indices of refraction for the slow mode and the fast mode, respectively. The PMD is defined as

$$\Delta\tau = \frac{d}{v_{gs}} - \frac{d}{v_{gf}}$$

where v_{gs} and v_{gf} are the group velocities of the modes. Show that

$$\Delta\tau = \left(\frac{d}{c}\Delta n + \frac{\omega d}{c}\frac{\partial\Delta n}{\partial\omega}\right)$$

7.7 PMD in a short segment of fiber is proportional to the length. In a long segment of fiber with random birefringence, the PMD is proportional to the square root of the length. Consider a long fiber consisting of N segments of the following parameters:

$$d_i = L/N$$

$$\tau_i = \tau_1$$

(a) Show that

$$(\Delta\tau_b)^2 = \frac{1}{4}N\tau_1^2 = \frac{1}{4}N\left(\frac{\Delta n}{c} + \frac{\omega}{c}\frac{\partial\Delta n}{\partial\omega}\right)^2 d_1^2$$

(b) Show that the broadening can be written

$$(\Delta\tau_b) = \frac{1}{2}\sqrt{N}\left(\frac{\Delta n}{c} + \frac{\omega}{c}\frac{\partial\Delta n}{\partial\omega}\right)d_1$$

$$= \frac{1}{2}\left(\frac{\Delta n}{c} + \frac{\omega}{c}\frac{\partial\Delta n}{\partial\omega}\right)\sqrt{Ld_1}$$

7.8 Consider a birefringent network consisting of two segments of uniform birefringent fibers. Let θ be the angle between the polarization directions of the slow modes of the two fibers. Define τ_1 and τ_2 as

$$\tau_1 = \left(\frac{d_1}{c}\Delta n_1 + \frac{\omega d_1}{c}\frac{\partial\Delta n_1}{\partial\omega}\right), \quad \tau_2 = \left(\frac{d_2}{c}\Delta n_2 + \frac{\omega d_2}{c}\frac{\partial\Delta n_2}{\partial\omega}\right)$$

$$t_1 = \left(\frac{d_1}{c}\bar{n}_1 + \frac{\omega d_1}{c}\frac{\partial\bar{n}_1}{\partial\omega}\right), \qquad t_2 = \left(\frac{d_2}{c}\bar{n}_2 + \frac{\omega d_2}{c}\frac{\partial\bar{n}_2}{\partial\omega}\right)$$

where d_1 and d_2 are lengths of the fibers, Δn_1 and Δn_2 are the birefringence values of the fibers, and $\bar{n}_1$ and $\bar{n}_2$ are the average indices. A beam of unpolarized light is launched into the fiber system with an input pulse of intensity $I_0(t)$. For simplicity, we assume $I_0(t)$ has a centroid at $t = 0$.

(a) Show that, at the end of the first segment of fibers, the intensity is

$$I_1(t) = \tfrac{1}{2}I_0(t - t_1 - \tau_1/2) + \tfrac{1}{2}I_0(t - t_1 + \tau_1/2)$$
$$\equiv I_{1s}(t) + I_{1f}(t)$$

where t_1 is the average group delay of the two polarization modes, I_{1s} is the intensity of the slow mode, and I_{1f} is the intensity of the fast mode.

(b) Show that, at the end of the second element, the intensity is

$$I_2(t) = I_{2s}(t) + I_{2f}(t)$$

where

$$I_{2s}(t) = \gamma I_{1s}(t - t_2 - \tau_2/2) + (1-\gamma)I_{1f}(t - t_2 - \tau_2/2)$$

$$I_{2f}(t) = \gamma I_{1f}(t - t_2 + \tau_2/2) + (1-\gamma)I_{1s}(t - t_2 + \tau_2/2)$$

where $\gamma = \cos^2\theta$.

(c) Show that

$$\tau_{g1} \equiv \int tI_1(t)\,dt = t_1 \quad \text{and} \quad \tau_{g2} \equiv \int tI_2(t)\,dt = t_1 + t_2$$

(d) Show that

$$b_1^2 \equiv \langle(t - \tau_{g1})^2\rangle \equiv \int (t - \tau_{g1})^2 I_1(t)\,dt = b_0^2 + \tfrac{1}{4}\tau_1^2$$

$$b_2^2 \equiv \langle(t - \tau_{g2})^2\rangle \equiv \int (t - \tau_{g2})^2 I_2(t)\,dt$$

$$= b_0^2 + \tfrac{1}{4}(\tau_1 - \tau_2)^2 + \gamma\tau_1\tau_2$$

where

$$b_0^2 \equiv \int t^2 I_0(t)\,dt$$

Note that $b_2^2 = b_0^2 + \frac{1}{4}\tau_1^2 + \frac{1}{4}\tau_2^2$ is valid only when $\gamma = \frac{1}{2}$.

7.9 Consider the same birefringent network of Problem 7.8. Let the azimuth angle of the slow axis of the first element be zero ($\psi_1 = 0$) and let θ be the azimuth angle of the slow axis of the second element ($\psi_2 = \theta$).

(a) The Jones matrix of the system can be written

$$M = \begin{bmatrix} e^{-i\Gamma_2/2}\cos^2\theta + e^{i\Gamma_2/2}\sin^2\theta & -i\sin(\Gamma_2/2)\sin(2\theta) \\ -i\sin(\Gamma_2/2)\sin(2\theta) & e^{-i\Gamma_2/2}\sin^2\theta + e^{i\Gamma_2/2}\cos^2\theta \end{bmatrix} \begin{bmatrix} e^{-i\Gamma_1/2} & 0 \\ 0 & e^{i\Gamma_1/2} \end{bmatrix}$$

where Γ_1 and Γ_2 are

$$\Gamma_1 = \Delta n_1 d_1 \omega/c \quad \text{and} \quad \Gamma_2 = \Delta n_2 d_2 \omega/c$$

Multiply the matrices and find the matrix elements. Write down expressions for a and b.

(b) Show that τ_1 and τ_2 defined in Problem 7.8 can be written

$$\tau_1 = \frac{\partial \Gamma_1}{\partial \omega} \quad \text{and} \quad \tau_2 = \frac{\partial \Gamma_2}{\partial \omega}$$

(c) Show that

$$|a'|^2 + |b'|^2 = \tfrac{1}{4}(\tau_2 - \tau_1)^2 + \tau_2\tau_1\cos^2\theta$$

The group delay between the two principal states of polarization is thus

$$\Delta\tau = 2\sqrt{|a'|^2 + |b'|^2} = \sqrt{(\tau_2 - \tau_1)^2 + 4\tau_2\tau_1\cos^2\theta}$$

Note that the PMD is minimum when $\theta = 90°$ when the slow axes of the two elements are mutually orthogonal, and maximum when the slow axes are parallel ($\theta = 0$).

(d) Find the Jones vectors of the principal states of polarization.

(e) For an input beam of unpolarized light, the broadening is given by, according to Equations (7.3-9) and (7.3-11),

$$(\Delta\tau_b)^2 = \tfrac{1}{4}(\Delta\tau)^2$$

Show that this broadening is consistent with the result of Problem 7.8(d).

7.10

(a) Using the matrix elements obtained in Problem 7.9(a), show that

$$A = a^*a' + bb'^* = -\frac{i}{2}[\tau_1 + \tau_2(\cos^2\theta - \sin^2\theta)]$$

$$B = a^*b' - a'^*b = -\frac{i}{2}\tau_2 \sin 2\theta \exp(i\Gamma_1)$$

(b) Show that the group delay obtained from

$$\tau = \pm\sqrt{AD - BC} = \pm\sqrt{|A|^2 + |B|}$$

agrees with the result of Problem 7.9(c).

(c) The principal states of polarization are given by Equation (7.4-9):

$$\mathbf{V}_{\text{in}} = \begin{bmatrix} -B \\ A + i\delta \end{bmatrix}$$

where $\delta = \tau/2, -\tau/2$. Investigate the frequency dependence of the PSP by plotting the polarization ellipse as a function of frequency. You may plot the ellipticity and the inclination angle of the ellipse as functions of the frequency.

7.11

(a) Show that the eigenvectors of the Jones matrix in Equations (7.5-10) and (7.5-11) are mutually orgthogonal.

(b) Derive Equation (7.5-12).

(c) Using Equation (7.5-14) for U, show that

$$U^\dagger\boldsymbol{\sigma}U = \cos^2(\gamma/2)\boldsymbol{\sigma} + \sin^2(\gamma/2)[(\mathbf{r}\cdot\boldsymbol{\sigma})\boldsymbol{\sigma}(\mathbf{r}\cdot\boldsymbol{\sigma})] + i\sin(\gamma/2)\cos(\gamma/2)[(\mathbf{r}\cdot\boldsymbol{\sigma})\boldsymbol{\sigma} - \boldsymbol{\sigma}(\mathbf{r}\cdot\boldsymbol{\sigma})]$$

(d) Using $\sigma_1\sigma_2 = -\sigma_2\sigma_1 = i\sigma_3$, $\sigma_2\sigma_3 = -\sigma_3\sigma_2 = i\sigma_1$, $\sigma_3\sigma_1 = -\sigma_1\sigma_3 = i\sigma_2$, and $\sigma_1^2 = \sigma_2^2 = \sigma_3^3 = 1$, show that

$$(\mathbf{r}\cdot\boldsymbol{\sigma})\boldsymbol{\sigma} - \boldsymbol{\sigma}(\mathbf{r}\cdot\boldsymbol{\sigma}) = -2i(\mathbf{r}\times\boldsymbol{\sigma})$$

$$(\mathbf{r}\cdot\boldsymbol{\sigma})\boldsymbol{\sigma}(\mathbf{r}\cdot\boldsymbol{\sigma}) = 2\mathbf{r}(\mathbf{r}\cdot\boldsymbol{\sigma}) - \boldsymbol{\sigma}$$

(e) Show that $U^\dagger\boldsymbol{\sigma}U = \cos(\gamma)\boldsymbol{\sigma} + 2\sin^2(\gamma/2)[\mathbf{r}(\mathbf{r}\cdot\boldsymbol{\sigma})] + \sin(\gamma)(\mathbf{r}\times\boldsymbol{\sigma})$.

(f) Using $\mathbf{r}\times(\mathbf{r}\times\boldsymbol{\sigma}) = \mathbf{r}(\mathbf{r}\cdot\boldsymbol{\sigma}) - \boldsymbol{\sigma} = (\mathbf{rr} - 1)\boldsymbol{\sigma}$, derive Equations (7.5-18) and (7.5-19).

7.12 By definition an anti-Hermitian matrix A satisfies $A^\dagger = -A$.

(a) Using $|a|^2 + |b|^2 = 1$, show that the trace of the matrix in Equation (7.5-23) is zero.

(b) Show that the matrix is anti-Hermitian.

(c) Show that the eigenvalues of an anti-Hermitian matrix are pure imaginary. [Hint: An anti-Hermitian matrix can be written $A = iH$, where H is a Hermitian matrix.]

REFERENCES

1. Poole, C. D., and R. E. Wagner, Phenomenological approach to polarization dispersion in long single-mode fibers. *Electron. Lett.* **22**(19):1029 (1986).
2. Gordon, J. P., and H. Kogelnik, *Proc. Natl. Acad. Sci. USA* **97**(9):4541 (2000).
3. Ciprut, P., B. Gisin, N. Gisin, R. Passy, J. P. Von der Weid, F. Prieto, and C. W. Zimmer, Second-order polarization mode dispersion: impact on analog and digital transmissions. *J. Lightwave Techno.* **16**(5):757 (1998).
4. Poole, C. D., and C. R. Giles, Polarization-dependent pulse compression and broadening due to polarization dispersion in dispersion shifted fibers. *Opt. Lett.* **13**(2):155 (1988).
5. Francia, C., F. Bruyere, D. Penninckx, and M. Chbat, PMD second-order effects on pulse propagation in single-mode optical fibers. *IEEE Photonics Technol. Lett.* **10**(12):1739 (1998).
6. Eyal, A., W. K. Marshall, M. Tur, and A. Yariv, Representation of second-order polarization mode dispersion. *Electron Lett.* **35**(19):1658 (1999).
7. Eyal, A., Y. Li, W. K. Marshall, M. Tur, and A. Yariv, Statistical determination of the length dependence of high-order PMD. *Opt. Lett.* **25**:875 (2000).
8. Li, Y., A. Eyal, P.-O. Hedekvist, and A. Yariv, Measurement of high-order polarization mode dispersion. *IEEE Photonics Techno. Lett.* **12**:861 (2000).

CHAPTER 8

NONLINEAR OPTICS

8.0 INTRODUCTION

In Chapter 1 we considered the propagation of electromagnetic radiation in linear media in which the induced polarization is proportional to the electric field. In this chapter we consider some of the consequences of the nonlinear dielectric properties of media in which, in addition to the linear response, an electric field produces a polarization that is a nonlinear function of the field. The nonlinear response can give rise to an exchange of energy between a number of electromagnetic fields of different frequencies. Some of the most important applications of this phenomenon include: (1) second-harmonic generation (SHG) in which part of the energy of an optical wave of frequency ω propagating through a crystal is converted to that of a wave at 2ω; (2) parametric oscillation in which a strong pump wave at ω_3 causes the simultaneous generation in a nonlinear crystal of radiation at lower frequencies ω_1 and ω_2, where $\omega_3 = \omega_1 + \omega_2$; and (3) stimulated Raman scattering (SRS), stimulated Brillouin scattering (SBS), and so on. Starting from the physical origin of the dielectric nonlinearity, we first describe a general methodology of nonlinear interaction of optical beams. We then provide an electromagnetic formulation of the coupling between optical beams in nonlinear media. Second-harmonic generation, parametric amplification, and third-order phenomena including the Kerr effect and Raman scattering will be addressed in detail in this chapter.

8.1 ON THE PHYSICAL ORIGIN OF NONLINEAR POLARIZATION

In any real atomic system, the polarization induced in the medium by the presence of an electric field is not exactly proportional to the electric field, but can be expressed in a Taylor series expansion, which in a lossless medium takes the form

$$P_i = \varepsilon_0 \chi_{ij} E_j + 2 d_{ijk} E_j E_k + 4 \chi_{ijkl} E_j E_k E_l + \cdots \tag{8.1-1}$$

where P_i is the ith component of the instantaneous polarization and E_i is the ith component of the instantaneous electric field of the optical beams ($i, j, k, l = x, y, z$). Summations over repeated indices are assumed. χ_{ij} is the linear susceptibility, while d_{ijk} and χ_{ijkl} are the second-order and third-order nonlinear optical susceptibilities, respectively. Since the order of E_j, E_k, E_l are immaterial in the above definition, the nonlinear optical susceptibilities observe the following symmetry:

$$d_{ijk} = d_{ikj}$$
$$\chi_{ijkl} = \chi_{i[jkl]} \tag{8.1-2}$$

where $[jkl]$ is any permutation of (jkl). It is known that the linear dielectric susceptibility tensor χ_{ij} is symmetric (i.e., $\chi_{ij} = \chi_{ji}$) in lossless media (see Problem 1.37). In the extreme case of a dc field (or low frequencies), the dispersion of the nonlinear coefficients d_{ijk} and χ_{ijkl} can be neglected. In this regime, all the relevant frequencies are far away from the absorption lines (or resonance frequencies), the system is lossless, and the material response is instan-taneous. In this case, the high-order terms can be viewed as a correction to the dielectric tensor. As a result, the nonlinear coefficients d_{ijk} and χ_{ijkl} are invariant under any reshuffling of their subscripts; for example, $\chi_{1231} = \chi_{2113}$ (see Problem 8.15). For the second-order nonlinear optical coefficient d_{ijk}, this symmetry reduces the 27 tensor elements to 10 independent elements. In lossy and consequently dispersive systems, d_{ijk} will in general depend on the frequencies of the waves involved in the nonlinear processes (e.g., sum or difference mixing, $\omega_3 = \omega_1 + \omega_2$ or $\omega_3 = \omega_1 - \omega_2$). In this case, the symmetry properties of the nonlinear coefficients d_{ijk} and χ_{ijkl} are limited to Equation (8.1-2). Throughout this book, we will use (1, 2, 3) and (x, y, z) interchangeably for the subscript indices of the nonlinear optical coefficients. For example, we write $\chi_{1111} = \chi_{xxxx}$, $\chi_{1122} = \chi_{xxyy}$.

The nonlinear optical response characterized by the parameters d_{ijk} and χ_{ijkl} give rise to numerous interesting phenomena and useful applications. The second-order nonlinearity $P_i = 2d_{ijk}E_jE_k$ is responsible for second-harmonic generation (SHG, or frequency doubling), for sum and difference frequency generation, and for parametric amplification and oscillation. The third-order term $P_i = 4\chi_{ijkl}E_jE_kE_l$ figures in such diverse phenomena as third-harmonic generation (THG), Raman and Brillouin scattering, self-focusing, the Kerr effect, the optical soliton, four-wave mixing, and phase conjugation. In this chapter, we will not concern ourselves with the physical origin of the nonlinear coefficients d_{ijk} and χ_{ijkl}. We will take them as material parameters and merely explore the electromagnetic phenomena and applications made possible by the nonlinearities.

8.2 SECOND-ORDER NONLINEAR PHENOMENA—GENERAL METHODOLOGY

Consider the nonlinear coupling of two optical fields. The first field, described in terms of its electric-field components, is given by

$$E_j^{\omega_1}(t) = \mathrm{Re}(E_{0j}^{\omega_1}e^{i\omega_1 t}) = \tfrac{1}{2}(E_{0j}^{\omega_1}e^{i\omega_1 t} + \text{c.c.}) \quad (j = x, y, z) \tag{8.2-1}$$

while the second field at ω_2 is

$$E_k^{\omega_2}(t) = \mathrm{Re}(E_{0k}^{\omega_2}e^{i\omega_2 t}) = \tfrac{1}{2}(E_{0k}^{\omega_2}e^{i\omega_2 t} + \text{c.c.}) \quad (k = x, y, z) \tag{8.2-2}$$

where c.c. stands for complex conjugate, and $E_{0j}^{\omega_1}$ and $E_{0k}^{\omega_1}$ are constant amplitudes of the two optical fields, respectively. If the medium is nonlinear, the presence of these field components can give rise to polarizations at frequencies $(n\omega_1 + m\omega_2)$, where n and m are any integers. To illustrate the general methodology, we consider the ith Cartesian component of the nonlinear polarization at $\omega_3 = \omega_1 + \omega_2$ as

$$P_i^{\omega_3=\omega_1+\omega_2}(t) = \mathrm{Re}(P_{0i}^{\omega_3}e^{i\omega_3 t}) = \tfrac{1}{2}(P_{0i}^{\omega_3}e^{i\omega_3 t} + \text{c.c.}) \quad (i = x, y, z) \tag{8.2-3}$$

where $P_{0i}^{\omega_3}$ is a constant amplitude of the polarization. Limiting our attention to the second-order terms in Equation (8.1-1)—that is,

$$P_i = 2d_{ijk}E_jE_k \tag{8.2-4}$$

we obtain

$$P_i(t) = 2d_{ijk}\tfrac{1}{2}(E_{0j}^{\omega_1}e^{i\omega_1 t} + E_{0j}^{\omega_2}e^{i\omega_2 t} + \text{c.c.}) \times \tfrac{1}{2}(E_{0k}^{\omega_1}e^{i\omega_1 t} + E_{0k}^{\omega_2}e^{i\omega_2 t} + \text{c.c.}) \tag{8.2-5}$$

Consider only the sum frequency term

$$P_i^{\omega_1+\omega_2}(t) = (\tfrac{1}{2}d_{ijk}E_{0j}^{\omega_1}E_{0k}^{\omega_2}e^{i(\omega_1+\omega_2)t} + \tfrac{1}{2}d_{ikj}E_{0k}^{\omega_2}E_{0j}^{\omega_1}e^{i(\omega_2+\omega_1)t} + \text{c.c.}) \tag{8.2-6}$$

where we recall the convention of summation over repeated indices. In a lossless (instantaneous response) system, $d_{ijk} = d_{ikj}$, so that

$$P_i^{\omega_1+\omega_2}(t) = \tfrac{1}{2}P_{0i}^{\omega_1+\omega_2}e^{i(\omega_1+\omega_2)t} + \text{c.c.} = d_{ijk}E_{0j}^{\omega_1}E_{0k}^{\omega_2}e^{i(\omega_1+\omega_2)t} + \text{c.c.} \tag{8.2-7a}$$

or in terms of the field amplitudes,

$$P_{0i}^{\omega_1+\omega_2} = 2d_{ijk}E_{0j}^{\omega_1}E_{0k}^{\omega_2} \tag{8.2-7b}$$

It is important to note that d_{ijk} will in general depend on ω_1 and ω_2 and whether the sum or their difference is considered. Our primary interest will be in the case of transparent lossless media where the coefficients d_{ijk} do not depend on the frequencies involved or on whether their sum or difference is generated.

Only noncentrosymmetric crystals can possess a nonvanishing d_{ijk} tensor. This follows from the requirement that in a centrosymmetric crystal a reversal of the signs of $E_j^{\omega_1}$ and $E_k^{\omega_2}$ must also cause a reversal in the sign of $P_i^{\omega_1+\omega_2}$ and will not affect the amplitude. Using Equation (8.2-7b), we get

$$-2d_{ijk}E_{0j}^{\omega_1}E_{0k}^{\omega_2} = 2d_{ijk}(-E_{0j}^{\omega_1})(-E_{0k}^{\omega_2}) \tag{8.2-8}$$

so that $d_{ijk} = 0$. Lack of inversion symmetry is also the prerequisite for a linear electro-optic effect (Pockels effect) and piezoelectricity, so that all electro-optic and piezoelectric crystals can be expected to display second-order ($P \propto E^2$) nonlinear optical properties. The same argument can also be used to show that all crystals as well as liquids and gases can display third-order optical nonlinearities.

The nonlinear coefficients d_{ijk} are measured most often in second-harmonic generation experiments, where $\omega_1 = \omega_2 = \omega$. In this case, we have, according to Equation (8.2-5),

$$P_i(t) = (\tfrac{1}{2}P_{0i}^{2\omega}e^{i2\omega t} + \text{c.c.}) = 2d_{ijk}(\tfrac{1}{2}E_{0j}^{\omega}e^{i\omega t} + \text{c.c.}) \times (\tfrac{1}{2}E_{0k}^{\omega}e^{i\omega t} + \text{c.c.}) \tag{8.2-9}$$

or equivalently,

$$P_{0i}^{2\omega} = d_{ijk}E_{0j}^{\omega}E_{0k}^{\omega} \tag{8.2-10}$$

where we observe the convention of summation over repeated indices. Note the factor of 2 difference between Equations (8.2-10) and (8.2-7b). It is important to note that we cannot obtain Equation (8.2-10) by setting $\omega_1 = \omega_2$ in Equation (8.2-7b). This is a result of the field normalization in Equation (8.2-5) when ω_2 approaches ω_1.

Since no physical significance can be attached to the interchange of j and k in Equation (8.2-10), we can replace the subscripts kj and jk by the contracted indices as follows:

$$xx = 1, \quad yy = 2, \quad zz = 3$$

$$yz = zy = 4, \quad xz = zx = 5, \quad xy = yx = 6$$

The resulting d_{iK} components ($i = 1, 2, 3$ and $K = 1, 2, 3, 4, 5, 6$) form a 3×6 matrix that operates on the E^2 column tensor to yield the amplitude of the second-harmonic nonlinear polarization $\mathbf{P}^{2\omega}$. The second-harmonic polarization amplitude can thus be written, according to Equation (8.2-10),

$$\begin{bmatrix} P_{0x}^{2\omega} \\ P_{0y}^{2\omega} \\ P_{0z}^{2\omega} \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \end{bmatrix} \begin{bmatrix} E_{0x}^{\omega} E_{0x}^{\omega} \\ E_{0y}^{\omega} E_{0y}^{\omega} \\ E_{0z}^{\omega} E_{0z}^{\omega} \\ 2E_{0y}^{\omega} E_{0z}^{\omega} \\ 2E_{0z}^{\omega} E_{0x}^{\omega} \\ 2E_{0x}^{\omega} E_{0y}^{\omega} \end{bmatrix} \tag{8.2-11}$$

The contracted d_{iK} tensor obeys the same symmetry restrictions as the piezoelectric tensor and the electro-optic tensor (see Chapter 9), and in crystals of a given point-group symmetry it has the same form. The tensor forms are given in Table 8.1.

TABLE 8.1 The Form of the Second-Order Nonlinear Optical Tensor in Contracted Notation for All Crystal Classes

Centrosymmetric: ($\bar{1}$, 2/*m*.*mmm*.4/*m*.4/*mmm*, $\bar{3}$, $\bar{3}m$, 6/*m*, 6/*mmm*, *m*3, *m*3*m*)

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Triclinic: (1)

$$\begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \end{bmatrix}$$

Monoclinic: 2 (2 $\parallel x_2$)

$$\begin{bmatrix} 0 & 0 & 0 & d_{14} & 0 & d_{16} \\ d_{21} & d_{22} & d_{23} & 0 & d_{25} & 0 \\ 0 & 0 & 0 & d_{34} & 0 & d_{36} \end{bmatrix}$$

2 (2 $\parallel x_3$)

$$\begin{bmatrix} 0 & 0 & 0 & d_{14} & d_{15} & 0 \\ 0 & 0 & 0 & d_{24} & d_{25} & 0 \\ d_{31} & d_{32} & d_{33} & 0 & 0 & d_{36} \end{bmatrix}$$

m (*m* $\perp x_2$)

$$\begin{bmatrix} d_{11} & d_{12} & d_{13} & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{24} & 0 & d_{26} \\ d_{31} & d_{32} & d_{33} & 0 & d_{35} & 0 \end{bmatrix}$$

m (*m* $\perp x_3$)

$$\begin{bmatrix} d_{11} & d_{12} & d_{13} & 0 & 0 & d_{16} \\ d_{21} & d_{22} & d_{23} & 0 & 0 & d_{26} \\ 0 & 0 & 0 & d_{34} & d_{35} & 0 \end{bmatrix}$$

Orthorhombic: 222

$$\begin{bmatrix} 0 & 0 & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{25} & 0 \\ 0 & 0 & 0 & 0 & 0 & d_{36} \end{bmatrix}$$

2*mm*

$$\begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{24} & 0 & 0 \\ d_{31} & d_{32} & d_{33} & 0 & 0 & 0 \end{bmatrix}$$

Tetragonal: 4

$$\begin{bmatrix} 0 & 0 & 0 & d_{14} & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & -d_{14} & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{bmatrix}$$

$\bar{4}$

$$\begin{bmatrix} 0 & 0 & 0 & d_{14} & d_{15} & 0 \\ 0 & 0 & 0 & -d_{15} & d_{14} & 0 \\ d_{31} & -d_{31} & 0 & 0 & 0 & d_{36} \end{bmatrix}$$

422

$$\begin{bmatrix} 0 & 0 & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & -d_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

4*mm*

$$\begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & 0 & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{bmatrix}$$

$\bar{4}2m$ (2 $\parallel x_1$)

$$\begin{bmatrix} 0 & 0 & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & d_{36} \end{bmatrix}$$

Trigonal: 3

$$\begin{bmatrix} d_{11} & -d_{11} & 0 & d_{14} & d_{15} & -d_{22} \\ -d_{22} & d_{22} & 0 & d_{15} & -d_{14} & -d_{11} \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{bmatrix}$$

32

$$\begin{bmatrix} d_{11} & -d_{11} & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & -d_{14} & -d_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

TABLE 8.1 (cont'd)

	$3m\ (m \perp x_1)$: $\begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} & -d_{22} \\ -d_{22} & d_{22} & 0 & d_{15} & 0 & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{bmatrix}$	$3m\ (m \perp x_2)$: $\begin{bmatrix} d_{11} & -d_{11} & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & 0 & -d_{11} \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{bmatrix}$
Hexagonal:	6: $\begin{bmatrix} 0 & 0 & 0 & d_{14} & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & -d_{14} & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{bmatrix}$	$6mm$: $\begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & 0 & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{bmatrix}$
	622: $\begin{bmatrix} 0 & 0 & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & -d_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	
	$\bar{6}$: $\begin{bmatrix} d_{11} & -d_{11} & 0 & 0 & 0 & -d_{22} \\ -d_{22} & d_{22} & 0 & 0 & 0 & -d_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\bar{6}m2\ (m \perp x_1)$: $\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -d_{22} \\ -d_{22} & d_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
	$\bar{6}m2\ (m \perp x_2)$: $\begin{bmatrix} d_{11} & -d_{11} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -d_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	
Cubic:	$\bar{4}3m$, 23: $\begin{bmatrix} 0 & 0 & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & d_{14} \end{bmatrix}$	432: $\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

EXAMPLE: SHG IN KDP (KH_2PO_4)

As a result of the crystal symmetry $\bar{4}2m$, the nonvanishing nonlinear coefficients are, according to Table 8.1, d_{14}, $d_{25} = d_{14}$, and d_{36}. So the second-order nonlinear polarization amplitude can be written, according to Equation (8.2-11),

$$P_{0x} = 2d_{xzy}E_{0z}E_{0y} = 2d_{14}E_{0z}E_{0y}$$
$$P_{0y} = 2d_{yzx}E_{0z}E_{0x} = 2d_{25}E_{0z}E_{0y} = 2d_{14}E_{0z}E_{0y}$$
$$P_{0z} = 2d_{zxy}E_{0x}E_{0y} = 2d_{36}E_{0x}E_{0y}$$

where E_{0x}, E_{0y}, E_{0z} are the amplitudes of the electric field of the fundamental optical wave at frequency ω. Table 8.2 lists the nonlinear optical coefficients for a number of crystals.

8.3 ELECTROMAGNETIC FORMULATION AND OPTICAL SECOND-HARMONIC GENERATION

We consider the interaction of three waves with $\omega_3 = \omega_1 + \omega_2$ via the second-order optical nonlinearity. Let the field be written

$$E_i(t) = E_i^{\omega_1}(t) + E_i^{\omega_2}(t) + E_i^{\omega_3}(t) \quad (i = x', y', z') \tag{8.3-1}$$

with

$$E_i^{\omega_1}(t) = \tfrac{1}{2}(E_{0i}^{\omega_1}e^{i(\omega_1 t - k_1 z)} + \text{c.c.}) = \tfrac{1}{2}(a_{1i}E_1 e^{i(\omega_1 t - k_1 z)} + \text{c.c.}) \quad (i = x', y', z')$$
$$E_i^{\omega_2}(t) = \tfrac{1}{2}(E_{0i}^{\omega_2}e^{i(\omega_2 t - k_2 z)} + \text{c.c.}) = \tfrac{1}{2}(a_{2i}E_2 e^{i(\omega_2 t - k_2 z)} + \text{c.c.}) \quad (i = x', y', z')$$
$$E_i^{\omega_3}(t) = \tfrac{1}{2}(E_{0i}^{\omega_3}e^{i(\omega_3 t - k_3 z)} + \text{c.c.}) = \tfrac{1}{2}(a_{3i}E_3 e^{i(\omega_3 t - k_3 z)} + \text{c.c.}) \quad (i = x', y', z')$$

where we assume that all three fields are normal modes of propagation along the z direction in the nonlinear medium, with unique wavenumbers k_1, k_2, and k_3 and amplitudes E_1, E_2, and

TABLE 8.2 Second-Order Nonlinear Optical Coefficients for a Number of Crystals

Crystal	$d_{ijk}^{(2\omega)}$ in Units of $\frac{1}{9} \times 10^{-22}$ MKS[a]
Ag_3AsS_3	$d_{22} = 22.5$
(Proustite)	$d_{36} = 13.5$
$AgGaSe_2$	$d_{36} = 27 \pm 3$
$AgSbS_3$	$d_{36} = 9.5$
$AlPO_4$	$d_{11} = 0.38 \pm 0.03$
β-BaB_2O_4 (BBO)	$d_{11} = 5.8 \times d_{36}$ (KDP)
	$d_{31} = 0.05 \times d_{11}$
	$d_{22} < 0.05 \times d_{11}$
$Ba_2NaNb_5O_{15}$	$d_{33} = 10.4 \pm 0.7$
	$d_{32} = 7.4 \pm 0.7$
$BaTiO_3$	$d_{33} = 6.4 \pm 0.5$
	$d_{31} = 18 \pm 2$
	$d_{15} = 17 \pm 2$
CdS	$d_{33} = 28.6 \pm 2$
	$d_{31} = 30 \pm 10$
	$d_{36} = 33$
CdSe	$d_{31} = 22.5 \pm 3$
$CdGeAs_2$	$d_{36} = 363 \pm 70$
GaP	$d_{14} = 80 \pm 14$
GaAs	$d_{14} = 72$
KH_2PO_4	$d_{36} = 0.45 \pm 0.03$
(KDP)	$d_{14} = 0.35$
KD_2PO_4	$d_{36} = 0.42 \pm 0.02$
	$d_{14} = 0.42 \pm 0.02$
KH_2AsO_4	$d_{36} = 0.48 \pm 0.03$
	$d_{14} = 0.51 \pm 0.03$
LiB_3O_5 (LBO)	$d_{31} = 0.84$
	$d_{32} = -0.78$
$LiIO_3$	$d_{15} = 4.4$
$LiNbO_3$	$d_{15} = 4.4$
	$d_{22} = 2.3 \pm 1.0$
$NH_4H_2PO_4$	$d_{36} = 0.45$
(ADP)	$d_{14} = 0.50 \pm 0.02$
Quartz (SiO_2)	$d_{11} = 0.37 \pm 0.02$
Se	$d_{11} = 130 \pm 30$
Te	$d_{11} = 517$
ZnO	$d_{33} = 6.5 \pm 0.2$
	$d_{31} = 1.95 \pm 0.2$
	$d_{15} = 2.1 \pm 0.2$
ZnS	$d_{36} = 13$

[a] Some authors define the nonlinear coefficient d by $P = \varepsilon_0 dE^2$ (where d is in units of pm/V) rather than by the relation $P = dE^2$ used in this book. To convert: 1 pm/V = $0.7969 \times (\frac{1}{9} \times 10^{-22}$ MKS) and $(\frac{1}{9} \times 10^{-22}$ MKS) = 1.255 pm/V.

E_3. In the above equations, $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ are unit vectors representing the polarization direction of the normal modes of propagation of these three fields in the nonlinear medium. In formulating the interaction we adopt the z axis as the common direction of propagation, and (x', y', z') as the Cartesian components in the principal coordinates of the nonlinear medium. These three fields are propagating in the same direction (z axis) but may assume different polarization states. Generally, the direction of propagation (z) may not be parallel to one of the principal axes. The collinear propagation in the nonlinear medium is to ensure maximum

physical overlap. In the absence of a nonlinear dielectric response, these three fields are solutions of the wave equation and are propagating independently in the medium.

We now start with the wave equation in a form that includes the polarization explicitly:

$$\nabla^2\mathbf{E} = \mu_0 \frac{\partial^2}{\partial t^2}(\varepsilon_0\mathbf{E} + \mathbf{P}) = \mu_0\varepsilon \frac{\partial^2\mathbf{E}}{\partial t^2} + \mu_0 \frac{\partial^2}{\partial t^2}\mathbf{P}_{\mathrm{NL}} \tag{8.3-2}$$

where $\mathbf{P}_{\mathrm{NL}}$ stands for the nonlinear polarization and can be written

$$(P_{\mathrm{NL}})_i = 2d_{ijk}E_jE_k \quad (i, j, k = x', y', z') \tag{8.3-3}$$

It is important to note that $\mathbf{E}$ in the above equation is the sum of the three fields given by Equation (8.3-1). As a result of the nonlinear polarization, the three fields are coupled. To obtain the coupled equations for the field amplitudes, we start with the nonlinear polarization at $\omega_1 = \omega_3 - \omega_2$. According to Equation (8.2-7a), we obtain

$$[P_{\mathrm{NL}}^{\omega_3-\omega_2}(z, t)]_i = d_{ijk}a_{3j}a_{2k}E_3E_2^* e^{i[(\omega_3-\omega_2)t-(k_3-k_2)z]} + \text{c.c.} \tag{8.3-4}$$

Similarly, we obtain the other two nonlinear polarizations

$$[P_{\mathrm{NL}}^{\omega_3-\omega_1}(z, t)]_i = d_{ijk}a_{3j}a_{1k}E_3E_1^* e^{i[(\omega_3-\omega_1)t-(k_3-k_1)z]} + \text{c.c.} \tag{8.3-5}$$

$$[P_{\mathrm{NL}}^{\omega_1+\omega_2}(z, t)]_i = d_{ijk}a_{1j}a_{2k}E_1E_2 e^{i[(\omega_1+\omega_2)t-(k_1+k_2)z]} + \text{c.c.} \tag{8.3-6}$$

where we observe the convention of summation over repeated indices ($i, j, k = x', y', z'$). These nonlinear polarizations can be viewed as distributed dipole sources, which can radiate and generate waves at the oscillating frequencies. We now substitute these nonlinear polarizations into the wave equation. By carrying out the indicated differentiation and assuming the following slowly varying amplitude approximation,

$$\frac{d^2}{dz^2}E_s \ll k_s \frac{d}{dz}E_s \quad (s = 1, 2, 3) \tag{8.3-7}$$

we obtain, after a few steps of algebra,

$$\begin{aligned}
\frac{d}{dz}E_1 &= -i\omega_1\sqrt{\frac{\mu_0}{\varepsilon_1}}\, dE_3E_2^* e^{-i(k_3-k_2-k_1)z} \\
\frac{d}{dz}E_2^* &= +i\omega_2\sqrt{\frac{\mu_0}{\varepsilon_2}}\, dE_1E_3^* e^{+i(k_3-k_2-k_1)z} \\
\frac{d}{dz}E_3 &= -i\omega_3\sqrt{\frac{\mu_0}{\varepsilon_3}}\, dE_1E_2 e^{+i(k_3-k_2-k_1)z}
\end{aligned} \tag{8.3-8}$$

where d is the effective second-order nonlinear coefficient

$$d = \sum_{ijk} d_{ijk}a_{1i}a_{2j}a_{3k} \tag{8.3-9}$$

The coupled equations (8.3-8) constitute the main result of this section. In arriving at the coupled equations, we have employed the cyclic symmetry of the nonlinear coefficients, that is, $d_{ijk} = d_{jik} = d_{ikj}$. We will apply them in the following sections to some specific cases. We note the coupled equations are in agreement with the conservation of energy. It can be shown that

$$\frac{d}{dz}\left(\sqrt{\varepsilon_1}\,|E_1|^2 + \sqrt{\varepsilon_2}\,|E_2|^2 + \sqrt{\varepsilon_3}\,|E_3|^2\right) = 0 \tag{8.3-10}$$

provided $\omega_3 = \omega_1 + \omega_2$.

We define the following new field variables A_1, A_2, and A_3:

$$A_m = \sqrt{\frac{n_m}{\omega_m}}\, E_m, \quad m = 1, 2, 3 \tag{8.3-11}$$

where n_m is the index of refraction associated with wave E_m and ω_m is the corresponding frequency. The beam intensity can be written

$$I_m = \frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu_0}}\, n_m |E_m|^2 = \frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu_0}}\, \omega_m |A_m|^2 \tag{8.3-12}$$

Since a photon's energy is $\hbar\omega$, it follows that $|A_m|^2$ is proportional to the photon flux of the beam at frequency ω_m, the proportional constant being independent of frequency. The coupled equations (8.3-8) can now be written

$$\begin{aligned} \frac{d}{dz} A_1 &= -i\kappa A_3 A_2^* e^{-i\Delta kz} \\ \frac{d}{dz} A_2^* &= +i\kappa A_1 A_3^* e^{+i\Delta kz} \\ \frac{d}{dz} A_3 &= -i\kappa A_1 A_2 e^{+i\Delta kz} \end{aligned} \tag{8.3-13}$$

where the momentum mismatch Δk (or wavenumber mismatch) and the coupling constant κ are given by

$$\Delta k = k_3 - (k_1 + k_2) \tag{8.3-14}$$

$$\kappa = d\sqrt{\frac{\mu_0\omega_1\omega_2\omega_3}{\varepsilon_0 n_1 n_2 n_3}} = \left(\sum_{ijk} d_{ijk} a_{1i} a_{2j} a_{3k}\right)\sqrt{\frac{\mu_0\omega_1\omega_2\omega_3}{\varepsilon_0 n_1 n_2 n_3}} \tag{8.3-15}$$

where the summations are over all the components of the polarization unit vectors. Using Equation (8.3-13), the conservation of energy becomes

$$\frac{d}{dz}(\omega_1|A_1|^2 + \omega_2|A_2|^2 + \omega_3|A_3|^2) = 0 \tag{8.3-16}$$

Second-Harmonic Generation

For the case of second-harmonic generation (SHG), we consider the interaction of two waves with frequencies ω_1 and ω_3 (with $\omega_3 = 2\omega_1$) via the second-order optical nonlinearity. Let the fields be written

$$\begin{aligned} E_i^{\omega_1}(t) &= \tfrac{1}{2}(E_{0i}^{\omega_1} e^{i(\omega_1 t - k_1 z)} + \text{c.c.}) = \tfrac{1}{2}(a_{1i} E_1 e^{i(\omega_1 t - k_1 z)} + \text{c.c.}) \quad (i = x', y', z') \\ E_i^{\omega_3}(t) &= \tfrac{1}{2}(E_{0i}^{\omega_3} e^{i(\omega_3 t - k_3 z)} + \text{c.c.}) = \tfrac{1}{2}(a_{3i} E_3 e^{i(\omega_3 t - k_3 z)} + \text{c.c.}) \quad (i = x', y', z') \end{aligned} \tag{8.3-17}$$

where a_1 and a_3 are unit vectors representing the polarization states of the waves.

The nonlinear polarizations for SHG can be written

$$[P_{\text{NL}}^{\omega_3 - \omega_1}(z, t)]_i = d_{ijk} a_{3j} a_{1k} E_3 E_1^* e^{i[(\omega_3 - \omega_1)t - (k_3 - k_1)z]} + \text{c.c.} \tag{8.3-18}$$

$$[P_{\text{NL}}^{2\omega_1}(z, t)]_i = \tfrac{1}{2} d_{ijk} a_{1j} a_{1k} E_1 E_1 e^{i(2\omega_1 t - 2k_1 z)} + \text{c.c.} \tag{8.3-19}$$

where summations over repeated indices are assumed. The coupled equations for SHG can now be written

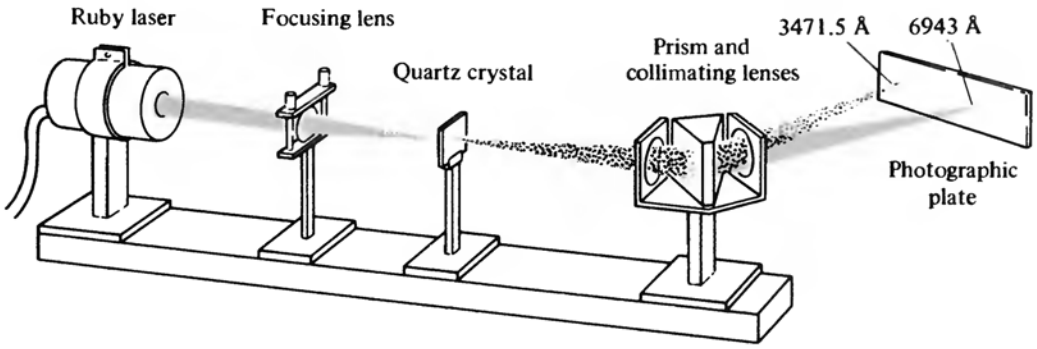


Figure 8.1 Arrangement used in the first experimental demonstration of second-harmonic generation [1]. A ruby laser beam at $\lambda_0 = 0.694$ μm is focused on a quartz crystal, causing generation of a (weak) beam at $\lambda_0/2 = 0.347$ μm. The two beams are then separated by a prism and detected on a photographic plate.

$$\frac{d}{dz}E_1 = -i\omega_1\sqrt{\frac{\mu_0}{\varepsilon_1}}\,dE_3E_1^*e^{-i(k_3-2k_1)z}$$
$$\frac{d}{dz}E_3 = -i\omega_3\sqrt{\frac{\mu_0}{\varepsilon_3}}\,\frac{1}{2}dE_1E_1e^{+i(k_3-2k_1)z} \tag{8.3-20}$$

where d is the effective second-order nonlinear coefficient

$$d = \sum_{ijk} d_{ijk}a_{3i}a_{1j}a_{1k} \tag{8.3-21}$$

The coupled equations (8.3-20) can now be applied for investigating the energy coupling between the fundamental and the second-harmonic wave.

It can be shown that

$$\frac{d}{dz}(\sqrt{\varepsilon_1}\,|E_1|^2 + \sqrt{\varepsilon_3}\,|E_3|^2) = 0 \tag{8.3-22}$$

provided $\omega_3 = 2\omega_1$.

The second-harmonic generation experiment that ushered in the field of nonlinear optics was performed by Franken, Hill, Peters, and Weinreich [1] in 1961. In their experiment (see Figure 8.1), a ruby laser beam at 694.3 nm was focused on the front surface of a crystalline quartz plate. The emergent radiation was examined with a spectrometer and was found to contain radiation at twice the input frequency (i.e., at λ = 347.15 nm). The conversion efficiency in this first experiment was $\sim 10^{-8}$. The utilization of more efficient materials, higher-intensity lasers, and phase-matching techniques has resulted, in the last few years, in conversion efficiencies approaching 100%. These factors will be discussed later in the section.

Phase Matching and Coherence Length

To illustrate the concept of phase matching, we consider a case of nondepletion approximation in which the amount of power lost from the input (ω_1) beam (by conversion to $2\omega_1$) is negligible, so that we may treat E_1 as a constant in the coupled equations (8.3-20) and we need consider only the second of the equations (8.3-20).

$$\frac{d}{dz}E_3 = -i\omega\sqrt{\frac{\mu_0}{\varepsilon}}\,dE_1^2e^{+i\Delta kz} \tag{8.3-23}$$

where

$$\omega = \omega_1 = \omega_3/2, \quad \varepsilon = \varepsilon_3$$

and

$$\Delta k = k_3 - 2k_1 \tag{8.3-24}$$

We now integrate Equation (8.3-23), assuming $E_3(0) = 0$ (i.e., no second-harmonic input at $z = 0$) and for a crystal of length L. We obtain

$$E_3(L) = -i\omega\sqrt{\frac{\mu_0}{\varepsilon}}\, dE_1^2 \frac{e^{+i(\Delta k)L} - 1}{i(\Delta k)L} L \tag{8.3-25}$$

and the beam intensity $I^{(2\omega)}$ of the second-harmonic output, or the conversion efficiency η_{SHG}, can be written

$$\eta_{SHG} \equiv \frac{I^{(2\omega)}}{I^{(\omega)}} = \frac{2\omega^2 d^2 L^2}{n^3}\left(\frac{\mu_0}{\varepsilon_0}\right)^{3/2} \frac{\sin^2[(\Delta k)L/2]}{[(\Delta k)L/2]^2} I^{(\omega)} \tag{8.3-26}$$

where $I^{(\omega)}$ is the beam intensity of the fundamental wave, $n = n^{(\omega)} \approx n^{(2\omega)}$, and Δk is the wavenumber mismatch given by Equation (8.3-24). According to this equation, a prerequisite for efficient second-harmonic generation is that $\Delta k = 0$, or

$$k^{(2\omega)} = 2k^{(\omega)} \tag{8.3-27}$$

If $\Delta k \neq 0$, the second-harmonic wave generated at some plane (e.g., z_1), having propagated to some other plane (e.g., z_2), is not in phase with the second-harmonic wave generated at z_2. This results in the interference described by the sinusoidal factor on the right side of Equation (8.3-26). Two adjacent peaks of this spatial interference pattern are separated by the "coherence length,"

$$l_c = \frac{2\pi}{\Delta k} = \frac{2\pi}{k^{(2\omega)} - 2k^{(\omega)}} \tag{8.3-28}$$

The coherence length l_c is thus a measure of the maximum crystal length that is useful in producing the second-harmonic power. In terms of the indices of refraction, the wavenumber mismatch can be written

$$\Delta k = k^{(2\omega)} - 2k^{(\omega)} = \frac{2\omega}{c} n^{(2\omega)} - 2\frac{\omega}{c} n^{(\omega)} = \frac{2\omega}{c}(n^{(2\omega)} - n^{(\omega)}) \tag{8.3-29}$$

and the coherence length can be written

$$l_c = \frac{2\pi}{\Delta k} = \frac{2\pi}{k^{(2\omega)} - 2k^{(\omega)}} = \frac{2c\pi}{2\omega(n^{(2\omega)} - n^{(\omega)})} = \frac{\lambda}{2(n^{(2\omega)} - n^{(\omega)})} \tag{8.3-30}$$

where λ is the wavelength of the fundamental wave. Under ordinary circumstances, it may be no longer than 10^{-2} cm. This is because the index of refraction normally increases with ω. To illustrate this, we consider the following example.

EXAMPLE: COHERENCE LENGTH OF FREQUENCY DOUBLING IN KDP

Let $\lambda = 1$ μm, and assume ordinary waves in KDP. The refractive indices (from Table 8.3) are

$$n(2\omega) = 1.514928$$

$$n(\omega) = 1.496044$$

Using Equation (8.3-30), we obtain

$$l_c = \frac{\lambda}{2(n^{(2\omega)} - n^{(\omega)})} = \frac{10^{-6}}{2(1.514928 - 1.496044)}\ \text{m} = 26.5\ \mu\text{m}$$

TABLE 8.3 Index of Refraction Dispersion Data of KH_2PO_4

Wavelength (μm)	n_o (Ordinary Ray)	n_e (Extraordinary Ray)
0.2000	1.622630	1.563913
0.3000	1.545570	1.498153
0.4000	1.524481	1.480244
0.5000	1.514928	1.472486
0.6000	1.509274	1.468267
0.7000	1.505235	1.465601
0.8000	1.501924	1.463708
0.9000	1.498930	1.462234
1.0000	1.496044	1.460993
1.1000	1.493147	1.459884
1.2000	1.490169	1.458845
1.3000	1.487064	1.457838
1.4000	1.483803	1.456838
1.5000	1.480363	1.455829
1.6000	1.476729	1.454797
1.7000	1.472890	1.453735
1.8000	1.468834	1.452636
1.9000	1.464555	1.451495
2.0000	1.460044	1.450308

After Reference [2].

The technique that is used widely (see References [3, 4]) to satisfy the *phase-matching* requirement, $\Delta k = 0$, takes advantage of the natural birefringence of anisotropic crystals, which was discussed in Chapter 1. Using the relation $k^{(\omega)} = \omega\sqrt{\mu\varepsilon_0 n^{\omega}}$, the phase-matching condition (8.3-27) becomes

$$n^{2\omega} = n^{\omega} \tag{8.3-31}$$

so the indices of refraction at the fundamental and second-harmonic frequencies must be equal. In normally dispersive materials, the index of the ordinary wave or the extraordinary wave along a given direction increases with ω, as can be seen from Table 8.3. This makes it impossible to satisfy (8.3-31) when both the ω and 2ω beams are of the same polarization type—that is, when both waves are extraordinary or ordinary. We can, however, under certain circumstances, satisfy (8.3-31) by making the two waves be of different types (different polarization states). To illustrate the point, consider the dependence of the index of refraction of the extraordinary wave in a uniaxial crystal on the angle θ between the propagation direction and the crystal optic (z') axis. It is given by

$$\frac{1}{n_e^2(\theta)} = \frac{\cos^2\theta}{n_o^2} + \frac{\sin^2\theta}{n_e^2} \tag{8.3-32}$$

According to Equation (8.3-32), we note that the index of refraction associated with the propagation of the extraordinary wave can vary from n_o at $\theta = 0$ to n_e at $\theta = 90°$. In a negative birefringent crystal (e.g., KDP), if $n_e^{2\omega} < n_o^{\omega}$, there exists an angle θ_m at which $n_e^{2\omega}(\theta_m) = n_o^{\omega}$; so if the fundamental beam (at ω) is launched along θ_m as an ordinary ray, the second-harmonic beam will be generated along the *same direction* as an extraordinary ray. The situation is illustrated by Figure 8.2. The angle θ_m is determined by the intersection between the sphere (shown as a circle in the figure) corresponding to the index surface of the ordinary beam at ω, and the index surface of the extraordinary ray $n_e^{2\omega}(\theta_m)$. The angle θ_m, which defines a

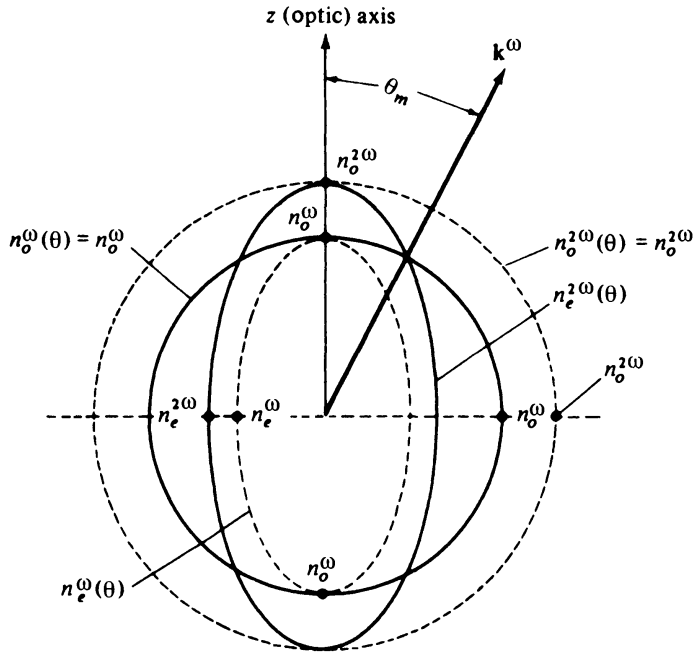


Figure 8.2 Normal (index) surfaces for the ordinary and extraordinary rays in a negative ($n_e < n_o$) uniaxial crystal. If $n_e^{2\omega} - n_o^{\omega}$, the condition $n_e^{2\omega}(\theta) = n_o^{\omega}$ is satisfied at $\theta = \theta_m$. The eccentricities shown are vastly exaggerated.

cone, for negative uniaxial crystals—that is, crystals in which $n_e^{2\omega} < n_o^{\omega}$ —is that satisfying $n_e^{2\omega}(\theta_m) = n_o^{\omega}$ or, using Equation (8.3-32),

$$\frac{\cos^2\theta_m}{(n_o^{2\omega})^2} + \frac{\sin^2\theta_m}{(n_e^{2\omega})^2} = \frac{1}{(n_o^{\omega})^2} \tag{8.3-33}$$

and, solving for θ_m,

$$\sin^2\theta_m = \frac{(n_o^{\omega})^{-2} - (n_o^{2\omega})^{-2}}{(n_e^{2\omega})^{-2} - (n_o^{2\omega})^{-2}} \tag{8.3-34}$$

The method of phase matching depicted in Figure 8.2 is known as type I phase matching. In type I phase matching, the polarization states of the fundamental and the second harmonic are mutually orthogonal. In a uniaxial crystal, if the fundamental is an ordinary wave, then the second harmonic is an extraordinary, or vice versa. Another mode of phase matching is possible in which the fundamental wave is a mixture of ordinary wave and extraordinary wave, and the second-harmonic wave can be either the ordinary or the extraordinary wave. The latter is known as type II phase matching. Generally, type I phase matching in uniaxial crystals can be obtained provided the birefringence ($|\Delta n| = |n_e - n_o|$) is larger than the index change ($n^{(2\omega)} - n^{(\omega)}$) due to chromatic dispersion of the crystal. If the birefringence is larger than twice the index change due to chromatic dispersion, then type II phase matching is possible. Table 8.4 is a summary of phase matching in uniaxial crystals.

TABLE 8.4 Summary of Phase Matching in Uniaxial Crystals

Birefringence	Type	Fundamental ω	Second Harmonic 2 ω	Condition
Negative ($n_e < n_o$)	I	O	E	$n_e^{(2\omega)} - n_e^{(\omega)} < n_o^{(\omega)} - n_e^{(\omega)}$
	II	O + E	E	$2(n_e^{(2\omega)} - n_e^{(\omega)}) < (n_o^{(\omega)} - n_e^{(\omega)})$
Positive ($n_o < n_e$)	I	E	O	$n_o^{(2\omega)} - n_o^{(\omega)} < n_e^{(\omega)} - n_o^{(\omega)}$
	II	O + E	O	$2(n_o^{(2\omega)} - n_o^{(\omega)}) < (n_e^{(\omega)} - n_o^{(\omega)})$

Note: E = extraordinary wave; O = ordinary wave.

EXAMPLE: SECOND-HARMONIC GENERATION

Consider the problem of second-harmonic generation using the output of a pulsed ruby laser ($\lambda_0 = 0.6943$ μm) in a KH_2PO_4 crystal (KDP). The appropriate d coefficient can be obtained by using Equation (8.3-21). The unit vectors representing the polarization states of the waves are given by

$$\mathbf{a}_3 = \begin{bmatrix} \cos\theta_m/\sqrt{2} \\ -\cos\theta_m/\sqrt{2} \\ \sin\theta_m \end{bmatrix} \quad \text{and} \quad \mathbf{a}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$$

where $\mathbf{a}_3$ is the polarization unit vector of the second-harmonic wave, while $\mathbf{a}_1$ is that of the fundamental wave. Using Equation (8.3-21) and carrying out the summation, we obtain

$$d = d_{312} \sin\theta_m = d_{36} \sin\theta_m$$

where d_{36} is listed in Table 8.2. The angle θ_m between the optic axis (c axis) and the direction of propagation for which $\Delta k = 0$ is given by Equation (8.3-34). The appropriate indices are taken from Table 8.2 and are

$$n_e(\lambda = 0.694\ \mu\text{m}) = 1.466, \quad n_e(\lambda = 0.347\ \mu\text{m}) = 1.490$$

$$n_o(\lambda = 0.694\ \mu\text{m}) = 1.506, \quad n_o(\lambda = 0.347\ \mu\text{m}) = 1.534$$

Substituting the foregoing data into Equation (8.3-34) gives

$$\theta_m = 52°$$

To obtain phase matching along this direction, the fundamental beam in the crystal must be polarized as appropriate to an ordinary ray in accordance with the discussion following Equation (8.3-32).

Experimental Verification of Phase Matching

According to Equation (8.3-26), if the phase-matching condition $\Delta k = 0$ is violated, the output power is reduced by a factor

$$F = \frac{\sin^2[(\Delta k)L/2]}{[(\Delta k)L/2]^2} \tag{8.3-35}$$

from its (maximum) phase-matched value. The phase mismatch $(\Delta k)L/2$ is given, according to Equation (8.3-29), by

$$\frac{(\Delta k)L}{2} = \frac{\omega L}{c}[n_e^{2\omega}(\theta) - n_o^{\omega}] \tag{8.3-36}$$

and is thus a function of θ. If we use Equation (8.3-32) to expand $n_e^{2\omega}(\theta)$ as a Taylor series near $\theta \cong \theta_m$, retain the first two terms only, and assume perfect phase matching at $\theta = \theta_m$ so $n_e^{2\omega}(\theta_m) = n_o^{\omega}$, we obtain

$$\Delta k(\theta)L = -\frac{2\omega L}{c}\sin(2\theta_m)\frac{(n_e^{2\omega})^{-2} - (n_o^{2\omega})^{-2}}{2(n_o^{\omega})^{-3}}(\theta - \theta_m) \equiv 2\beta(\theta - \theta_m) \tag{8.3-37}$$

where β, as defined by Equation (8.3-37), is a constant depending on $n_e^{2\omega}$, $n_o^{2\omega}$, n_o^{ω}, ω, and L. If we plot the output power at 2ω as a function of θ, we would expect, according to Equations (8.3-26) and (8.3-37), to find it varying as

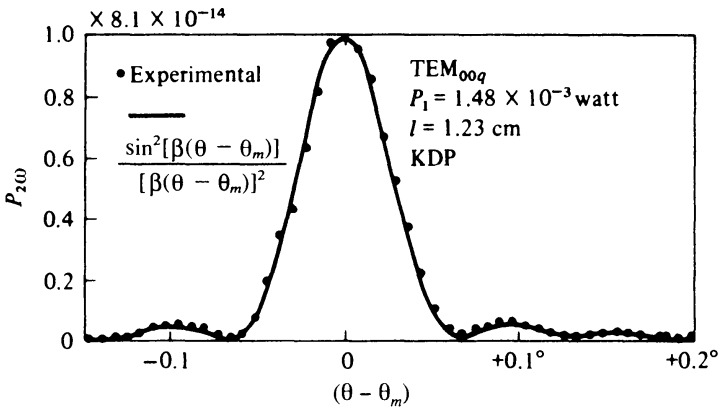


Figure 8.3 Variation of the second-harmonic power $P_{2\omega}$ with the angular departure $(\theta - \theta_m)$ from the phase-matching angle. (After Reference [5].)

$$P_{2\omega}(\theta) \propto \frac{\sin^2[\beta(\theta - \theta_m)]}{[\beta(\theta - \theta_m)]^2} \tag{8.3-38}$$

Figure 8.3 shows an experimental plot of $P_{2\omega}(\theta)$ as well as a theoretical plot of Equation (8.3-38). Another phase-matching technique involves the introduction of an artificial spatial periodicity $\Lambda = 2\pi/\Delta k$ into the beams' path. This method is discussed later in this chapter.

Second-Harmonic Generation with a Depleted Input

Expression (8.3-26) for the conversion efficiency in second-harmonic generation was derived assuming negligible depletion of the fundamental beam at ω. It is therefore valid only for cases where the conversion efficiency is small, that is, $\eta_{SHG} \ll 1$. Here we consider the general case of high-efficiency second-harmonic generation. In this case, we need to consider the depletion of the fundamental beam. Using the photon flux amplitude A_m defined in Equation (8.3-11), the coupled equations for second-harmonic generation can be written

$$\begin{aligned} \frac{dA_1}{dz} &= -i\kappa A_3 A_1^* e^{-i(\Delta k)z} \\ \frac{dA_3}{dz} &= -i\frac{\kappa}{2} A_1^2 e^{+i(\Delta k)z} \end{aligned} \tag{8.3-39}$$

where A_1 is the flux amplitude of the fundamental beam and A_3 is that of the second harmonic beam, and the coupling constant and the phase mismatch are given by

$$\kappa \equiv \sqrt{\left(\frac{\mu}{\varepsilon_0}\right)\frac{\omega_1\omega_1\omega_3}{n_1 n_1 n_3}}\, d \approx \left(\frac{\omega}{n}\right)^{3/2} \sqrt{2\left(\frac{\mu_0}{\varepsilon_0}\right)}\, d \tag{8.3-40}$$

$$\Delta k \equiv k_3 - (k_1 + k_2) = k^{(2\omega)} - 2k^{(\omega)} \tag{8.3-41}$$

In Equation (8.3-40), we used $\omega_3 = 2\omega_1 \equiv 2\omega$ and assumed $n = n(\omega) \approx n(2\omega)$. For the case of SHG with phase matching ($\Delta k = 0$), the coupled equations become

$$\begin{aligned} \frac{dA_1}{dz} &= -i\kappa A_3 A_1^* \\ \frac{dA_3}{dz} &= -i\frac{\kappa}{2} A_1^2 \end{aligned} \tag{8.3-42}$$

It follows from Equation (8.3-42) that if we choose, without loss of generality, $A_1(0)$ as a real number, then $A_1(z)$ is real and Equation (8.3-42) can be rewritten in the form

$$\frac{dA_1}{dz} = -\kappa A_3' A_1$$
$$\frac{dA_3'}{dz} = \frac{1}{2}\kappa A_1^2 \tag{8.3-43}$$

where $A_3 \equiv -iA_3'$. It follows from Equation (8.3-43) that

$$\frac{d}{dz}(A_1^2 + 2A_3'^2) = 0$$

(i.e., for every two photons "removed" from beam 1 (the fundamental beam at ω), one photon is added to beam 3 (the second harmonic at 2ω)).

Assuming no input at ω_3 (i.e., $A_3(0) = 0$), we have $A_1^2 + 2A_3'^2 = A_1^2(0)$, and the second of (8.3-43) becomes

$$\frac{dA_3'}{dz} = \frac{1}{2}\kappa\left(A_1^2(0) - 2A_3'^2\right) \tag{8.3-44}$$

A direct integration of the above equation leads to a solution

$$A_3'(z) = \frac{1}{\sqrt{2}} A_1(0) \tanh\left(\frac{1}{\sqrt{2}}\kappa A_1(0) z\right) \tag{8.3-45}$$

for the photon flux amplitude of the second-harmonic beam.

We note that as $\kappa A_1(0)z \to \infty$, $A_3'(z) \to A_1(0)/\sqrt{2}$ so that all the input photons at ω are converted into half as many photons at 2ω and the power conversion efficiency approaches unity. In the general case,

$$\eta_{SHG} \equiv \frac{P^{(2\omega)}}{P^{(\omega)}} = \frac{\hbar\omega_3 |A_3(z)|^2}{\hbar\omega_1 |A_1(z)|^2} = \frac{2|A_3(z)|^2}{|A_1(z)|^2} = \tanh^2[\tfrac{1}{2}\kappa A_1(0)z] \tag{8.3-46}$$

A plot of the theoretically predicted relation (8.3-46) as well as of experimental data obtained in converting from $\lambda = 1.06$ to $\lambda = 0.53$ μm is shown in Figure 8.4.

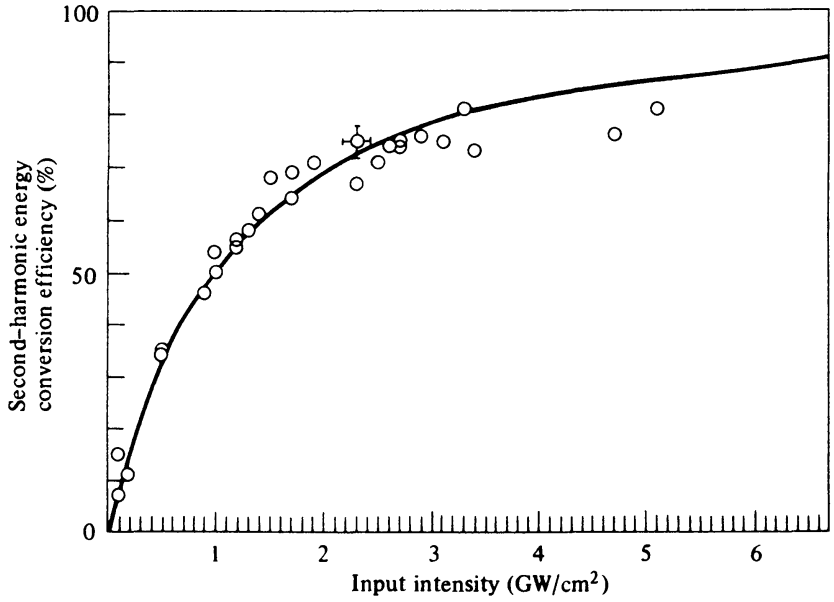


Figure 8.4 Frequency doubling energy conversion efficiency. Solid curve: theoretical prediction of Equation (8.3-46) (recall that $A_1(0) \propto \sqrt{I_{in}}$). The circles correspond to experimental points. (After supplementary Reference [6].)

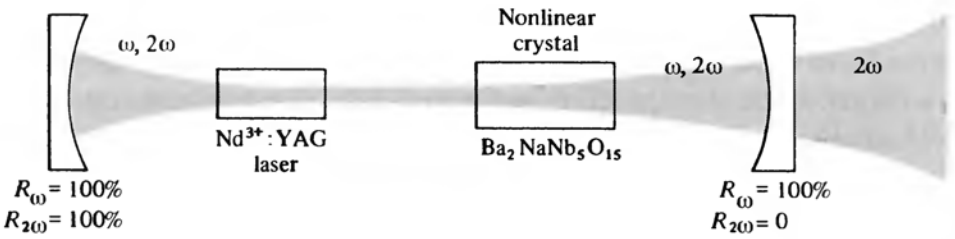


Figure 8.5 Typical setup for second-harmonic conversion inside a laser resonator. (After Reference [8].)

As a nonlinear optical process, an efficient second-harmonic generation usually requires a fundamental beam with a very high intensity. In the example shown in Figure 8.4, an input intensity of GW/cm^2 is needed to achieve a significant conversion efficiency. For such a high intensity, the nonlinear optical crystal must have a high damage threshold. To achieve a high conversion efficiency with a relatively low input intensity would require the development of new materials with a large nonlinear optical coefficient d. Another approach to achieve high conversion efficiency is to take advantage of the high intensity in an optical resonator and to place the nonlinear crystal inside the laser resonator where the energy flux can be made very large. This approach has been used successfully [7]. Figure 8.5 shows a schematic drawing of the experiment. In this example, the fundamental beam is a Nd:YAG laser beam oscillating at $\lambda = 1.064$ μm. The resonator mirrors are designed so that the Nd:YAG laser beam is confined inside the resonator. A crystal of $Ba_2NaNb_5O_{15}$ is placed inside the laser resonator as the nonlinear optical medium for SHG.

8.4 OTHER SECOND-ORDER NONLINEAR PROCESSES

In addition to the second-harmonic generation (SHG), there are other second-order nonlinear processes, including optical rectification, optical parametric amplification (OPA), optical parametric oscillation (OPO), and frequency up-conversion. In this section, we will discuss some of the most important processes. Optical parametric amplification can be employed to produce squeezed states of optical beams. This subject will be discussed in Chapter 18. Optical parametric oscillation is widely used in tunable solid-state lasers (e.g., Ti: sapphire lasers).

Optical Parametric Amplification

Optical parametric amplification (OPA) in its simplest form involves the transfer of power from a "pump" wave at ω_3 to two waves at lower frequencies ω_1 and ω_2, with $\omega_3 = \omega_1 + \omega_2$. It is fundamentally similar to the case of second-harmonic generation treated in Section 8.3. The only difference is in the direction of power flow. In SHG, power is fed from the low-frequency optical field at ω to the field at 2ω. In OPA, the power flow is from the high-frequency field (ω_3) to the low-frequency fields at ω_1 and ω_2. In the special case where $\omega_1 = \omega_2$, we have the exact reverse of SHG. These processes are the result of the nonlinear optical mixing of three beams. The equations governing the interaction among the three waves are identical to those described in Section 8.2. The only differences are the boundary conditions. Let the field of the three waves be written

$$\mathbf{E} = \sum_{m=1,2,3} \mathbf{a}_m E_m \exp[i(\omega_m t - k_m z)] \tag{8.4-1}$$

where $\mathbf{a}_m$ is the unit vector representing the polarization state of the mth beam in the nonlinear crystal, and E_m is the amplitude of the mth beam with frequency ω_m. Each of the waves is a normal mode of propagation with a well-defined state of polarization in the medium. The field amplitudes $E_1(z)$, $E_2(z)$, and $E_3(z)$ are functions of z, to reflect the coupling. We assume that these three beams are copropagating along the same direction (z axis), with $\omega_3 = \omega_1 + \omega_2$. As a result of the second-order optical nonlinearity, these three waves are coupled. The coupled equations can be written, according to Equation (8.3-8),

$$\begin{aligned}
\frac{dE_1}{dz} &= -\frac{i\omega_1}{2}\sqrt{\frac{\mu}{\varepsilon_1}}\, dE_3E_2^*e^{-i(k_3-k_2-k_1)z} \\
\frac{dE_2^*}{dz} &= +\frac{i\omega_2}{2}\sqrt{\frac{\mu}{\varepsilon_2}}\, dE_1E_3^*e^{-i(k_1-k_3+k_2)z} \\
\frac{dE_3}{dz} &= -\frac{i\omega_3}{2}\sqrt{\frac{\mu}{\varepsilon_3}}\, dE_1E_2e^{-i(k_1+k_2-k_3)z}
\end{aligned} \tag{8.4-2}$$

where the effective nonlinear coefficient d is given by

$$d = \sum_{ijk} d_{ijk}a_{1i}a_{2j}a_{3k} \tag{8.4-3}$$

In arriving at Equations (8.4-2), we have employed the permutation symmetry of the nonlinear coefficients (i.e., $d_{ijk} = d_{jik} = d_{ikj}$). We now transform to a new set of field variables A_l defined by

$$A_m \equiv \sqrt{\frac{n_m}{\omega_m}}\, E_m, \quad m = 1, 2, 3 \tag{8.4-4}$$

where $n_m^2 = \varepsilon_m/\varepsilon_0$; that is, n_m is the index of refraction of wave m, so that the beam intensity (power flow per unit area) at ω_m is given by

$$I_m = \frac{P_m}{A} = \frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu}}\, n_m |E_m|^2 = \frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu}}\, \omega_m |A_m|^2 \tag{8.4-5}$$

The power flow P_m/A per unit area is related to the flux N_m (photons per square meter per second) by

$$\frac{P_m}{A} = N_m\hbar\omega_m = \frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu}}\, |A_m|^2\omega_m \tag{8.4-6}$$

so that $|A_m|^2$ is proportional to the photon flux at ω_m. The coupled equations (8.4-2) for the photon flux amplitudes A_m become

$$\begin{aligned}
\frac{dA_1}{dz} &= -\frac{\alpha_1}{2}A_1 - \frac{i}{2}\kappa A_2^*A_3e^{-i(\Delta k)z} \\
\frac{dA_2^*}{dz} &= -\frac{\alpha_2}{2}A_2^* + \frac{i}{2}\kappa A_1A_3^*e^{i(\Delta k)z} \\
\frac{dA_3}{dz} &= -\frac{\alpha_3}{2}A_3 - \frac{i}{2}\kappa A_1A_2e^{i(\Delta k)z}
\end{aligned} \tag{8.4-7}$$

where

$$\kappa \equiv d\sqrt{\left(\frac{\mu}{\varepsilon_0}\right)\frac{\omega_1\omega_2\omega_3}{n_1n_2n_3}} \tag{8.4-8}$$

$$\Delta k \equiv k_3 - (k_1 + k_2) \tag{8.4-9}$$

and α_m is the linear attenuation coefficient for the mth beam. The advantage of using A_m instead of E_m is now apparent since, unlike (8.4-2), relations (8.4-7) involve a single coupling parameter κ.

We will now use (8.4-7) to obtain the field amplitudes $A_1(z)$, $A_2(z)$, and $A_3(z)$ for the case in which three waves with amplitudes $A_1(0)$, $A_2(0)$, and $A_3(0)$ at frequencies ω_1, ω_2, and ω_3, respectively, are incident on a nonlinear crystal at $z = 0$. We take $\omega_3 = \omega_1 + \omega_2$, $\alpha_1 = \alpha_2 = \alpha_3 = 0$ (no losses), and $\Delta k = k_3 - k_1 - k_2 = 0$. In addition, we assume that $\omega_1|A_1(z)|^2$ and $\omega_2|A_2(z)|^2$ remain small compared to $\omega_3|A_3(z)|^2$ throughout the interaction region. This last condition, in view of (8.4-7), is equivalent to assuming that the power drained off the "pump" (at ω_3) by the "signal" (ω_1) and idler (ω_2) is negligible compared to the input power at ω_3. This enables us to view $A_3(z)$ as a constant. With the assumptions stated above, Equations (8.4-7) become

$$\frac{dA_1}{dz} = -\frac{ig}{2}A_2^* \quad \text{and} \quad \frac{dA_2^*}{dz} = \frac{ig^*}{2}A_1 \tag{8.4-10}$$

where g is a constant given by, according to Equations (8.4-4) and (8.4-8),

$$g \equiv \kappa A_3(0) = \sqrt{\left(\frac{\mu}{\varepsilon_0}\right)\frac{\omega_1\omega_2}{n_1n_2}}\,E_3(0)d \tag{8.4-11}$$

We note that the coupling constant g is proportional to the (pump) field amplitude $E_3(0)$. Generally, g is a complex number, depending on the phase of the pump field amplitude $E_3(0)$.

The solutions of the coupled equations (8.4-10), subject to the boundary conditions $A_1(z = 0) \equiv A_1(0)$, $A_2(z = 0) \equiv A_2(0)$, are

$$\begin{aligned} A_1(z) &= A_1(0)\cosh\gamma z - i\frac{g}{|g|}A_2^*(0)\sinh\gamma z \\ A_2^*(z) &= A_2^*(0)\cosh\gamma z + i\frac{g}{|g|}A_1(0)\sinh\gamma z \end{aligned} \tag{8.4-12}$$

where the gain coefficient γ is given by

$$\gamma = \left|\frac{g}{2}\right| = \left|\frac{\kappa A_3(0)}{2}\right| = \frac{1}{2}\left|\sqrt{\left(\frac{\mu}{\varepsilon_0}\right)\frac{\omega_1\omega_2}{n_1n_2}}\,E_3(0)d\right| \tag{8.4-13}$$

where we recall that d is the effective nonlinear optical coefficient.

Equations (8.4-12) describe the growth of the signal and idler waves under phase-matching conditions. In the case of parametric amplification, the input will consist of the pump (ω_3) wave and one of the other two fields, say, ω_1. In this case $A_2(0) = 0$, and using the relation $N_i \propto A_iA_i^*$ for the photon flux, we obtain from (8.4-12)

$$\begin{aligned} N_1(z) \propto A_1^*(z)A_1(z) &= |A_1(0)|^2\cosh^2\gamma z \xrightarrow[\gamma z \gg 1]{} \frac{|A_1(0)|^2}{4}e^{2\gamma z} \\ N_2(z) \propto A_2^*(z)A_2(z) &= |A_1(0)|^2\sinh^2\gamma z \xrightarrow[\gamma z \gg 1]{} \frac{|A_1(0)|^2}{4}e^{2\gamma z} \end{aligned} \tag{8.4-14a}$$

Thus, for $\gamma z \gg 1$, the photon fluxes at ω_1 and ω_2 grow exponentially. If we limit our attention to the signal wave at ω_1, we note that the signal power (or signal photon flux) undergoes an amplification by a factor

$$\frac{A_1^*(z)A_1(z)}{A_1^*(0)A_1(0)} \underset{\gamma z \gg 1}{=} \frac{1}{4}e^{2\gamma z} \tag{8.4-14b}$$

EXAMPLE: PARAMETRIC AMPLIFICATION

The magnitude of the gain coefficient γ available in a traveling-wave parametric interaction is estimated for the following case involving the use of a $LiNbO_3$ crystal:

$$d_{15} = d_{113} = 5 \times 10^{-23} \text{ MKS (see Table 8.2)}$$

$$\nu_1 \cong \nu_2 = 3 \times 10^{14} \text{ Hz}$$

P_3/Area = Pump power $= 5 \times 10^6$ W/cm^2

$$n_1 \cong n_3 = 2.2$$

Converting P_3 to $|E_3|^2$ with the use of Equation (8.4-5) and then substituting into Equation (8.4-11) or (8.4-13) yields

$$\gamma = 0.35 \text{ cm}^{-1}$$

This shows that traveling-wave parametric amplification is not expected to lead to large values of gain except for extremely large pump-power densities. The main attraction of the parametric amplification just described is probably in giving rise to optical parametric oscillation (OPO), which will be described later in this section.

In the preceding section the analysis of parametric amplification assumed that the phase-matching condition

$$k_3 = k_1 + k_2 \tag{8.4-15}$$

is satisfied. It is important to determine the consequences of violating this condition.

We start with Equations (8.4-10) by adding the exponential terms accounting for the phase mismatch, and taking the loss coefficients $\alpha_1 = \alpha_2 = 0$:

$$\begin{aligned} \frac{dA_1}{dz} &= -i\frac{g}{2}A_2^* e^{-i(\Delta k)z} \\ \frac{dA_2^*}{dz} &= +i\frac{g^*}{2}A_1 e^{i(\Delta k)z} \end{aligned} \tag{8.4-16}$$

The above coupled equations can be solved as follows. We multiply the first equation of (8.4-16) by $\exp[+i(\Delta k)z]$ and then differentiate with respect to z. This leads to

$$\frac{d}{dz}\left(e^{+i(\Delta k)z}\frac{dA_1}{dz}\right) = -i\frac{g}{2}\frac{d}{dz}A_2^* \tag{8.4-17}$$

We can now eliminate A_2^* by using the second equation of (8.4-16). After a few steps of algebra, we obtain

$$\frac{d^2A_1}{dz^2} + i(\Delta k)\frac{dA_1}{dz} = \left|\frac{g}{2}\right|^2 A_1 \tag{8.4-18}$$

This is an ordinary second-order differential equation. The general solution can be written

$$A_1(z) = e^{-i(\Delta k/2)z}[C_1 \cosh(sz) + C_2 \sinh(sz)] \tag{8.4-19}$$

where C_1 and C_2 are constants, and

$$s = \sqrt{|g/2|^2 - (\Delta k/2)^2} \tag{8.4-20}$$

The general solutions of (8.4-16) subject to the boundary condition of $A_1(0)$ and $A_2^*(0)$ can be written

$$\begin{aligned} A_1(z)e^{i(\Delta k/2)z} &= A_1(0)\left[\cosh(sz) + \frac{i(\Delta k)}{2s}\sinh(sz)\right] - i\frac{g}{2s}A_2^*(0)\sinh(sz) \\ A_2^*(z)e^{-i(\Delta k/2)z} &= A_2^*(0)\left[\cosh(sz) - \frac{i(\Delta k)}{2s}\sinh(sz)\right] + i\frac{g}{2s}A_1(0)\sinh(sz) \end{aligned} \tag{8.4-21}$$

The last result reduces, as it should, to (8.4-12) if we put $\Delta k = 0$.

By examining Equations (8.4-21), we note that exponential gain is possible only when the gain coefficient s is a real number. This is possible provided

$$|g| = 2\gamma \geq \Delta k \tag{8.4-22}$$

according to Equation (8.4-20). When $|g| < \Delta k$, no sustained growth of the signal (A_1) and idler (A_2) waves is possible, since in this case the sinh and cosh functions in (8.4-21) become

$$\begin{aligned} &i\sin\{\tfrac{1}{2}[(\Delta k)^2 - |g|^2]^{1/2}z\} \\ &i\cos\{\tfrac{1}{2}[(\Delta k)^2 - |g|^2]^{1/2}z\} \end{aligned} \tag{8.4-23}$$

respectively, and the energies of the waves at ω_1 and ω_2 oscillate as functions of the distance z.

The problem of phase matching in parametric amplification is fundamentally the same as that in second-harmonic generation. Instead of satisfying the condition $k^{2\omega} = 2k^{\omega}$, we have, according to Equation (8.4-15), to satisfy the condition

$$k_3 = k_1 + k_2$$

This is done, as in second-harmonic generation, by using the dependence of the phase velocity of the extraordinary wave in uniaxial crystals on the direction of propagation. In a negative uniaxial crystal ($n_e < n_o$), we can, as an example, choose the signal and idler waves as ordinary while the pump at ω_3 is applied as an extraordinary wave. This is type I phase matching. Using Equation (8.3-32) and the relation $k^{\omega} = (\omega/c)n^{\omega}$, the phase-matching condition (8.4-15) is satisfied when all three waves propagate at an angle θ_m to the z (optic) axis, where

$$n_e^{\omega_3}(\theta_m) = \left[\left(\frac{\cos\theta_m}{n_o^{\omega_3}}\right)^2 + \left(\frac{\sin\theta_m}{n_e^{\omega_3}}\right)^2\right]^{-1/2} = \frac{\omega_1}{\omega_3}n_o^{\omega_1} + \frac{\omega_2}{\omega_3}n_o^{\omega_2} \tag{8.4-24}$$

Optical Parametric Oscillation (OPO)

In the previous discussion, we have demonstrated that a pump wave at ω_3 can cause a simultaneous amplification in a nonlinear medium of signal and idler waves at frequencies ω_1 and ω_2, respectively, where $\omega_3 = \omega_1 + \omega_2$. If the nonlinear crystal is placed within an optical resonator (as shown in Figure 8.6) that provides resonance for the signal or idler waves (or

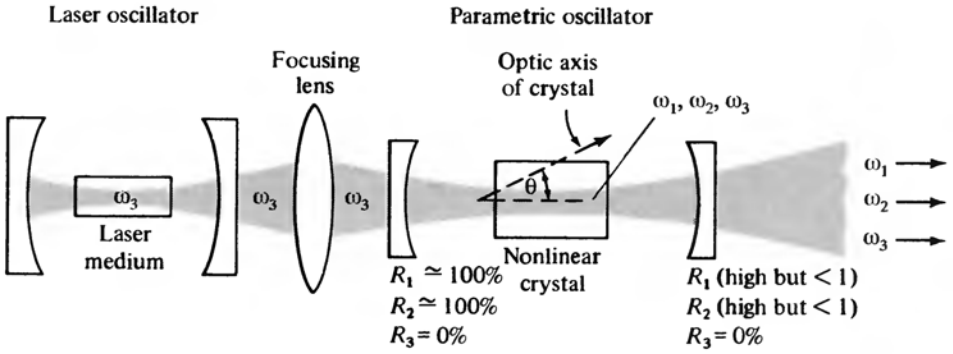


Figure 8.6 Schematic diagram of an optical parametric oscillator in which the laser output at ω_3 is used as the pump, giving rise to oscillations at ω_1 and ω_2 (where $\omega_3 = \omega_1 + \omega_2$) in an optical cavity that contains the nonlinear crystal and resonates at ω_1 and ω_2.

both), the parametric gain will, at some threshold pumping intensity, cause a simultaneous oscillation at the signal and idler frequencies. The threshold pumping corresponds to the point at which the parametric gain just balances the losses of the signal and idler waves. This is the physical basis of the optical parametric oscillator. Its practical importance derives from its ability to convert the power output of the pump laser to power at the signal and idler frequencies that, as will be shown later, can be tuned continuously over large spectral ranges.

To analyze this situation, we return to (8.4-7). We take $\Delta k = 0$ and neglect the depletion of the pump waves, so $A_3(z) = A_3(0)$. The result is

$$\begin{aligned} \frac{dA_1}{dz} &= -\frac{1}{2}\alpha_1 A_1 - i\frac{g}{2}A_2^* \\ \frac{dA_2^*}{dz} &= -\frac{1}{2}\alpha_2 A_2^* + i\frac{g^*}{2}A_1 \end{aligned} \tag{8.4-25}$$

where, as in Equation (8.4-11), the complex constant g is given by

$$g \equiv \sqrt{\left(\frac{\mu}{\varepsilon_0}\right)\frac{\omega_1\omega_2}{n_1 n_2}}\, E_3(0)d \tag{8.4-26}$$

Equations (8.4-25) describe traveling-wave parametric interactions. We will use them to describe the interaction inside a resonator. This procedure seems plausible if we think of propagation inside an optical resonator as a folded optical path. The magnitudes of the spatial distributed loss constants α_1 and α_2 must then be chosen so that they account for the actual losses in the resonator. The latter will include losses caused by the less than perfect reflection at the mirrors, as well as distributed loss in the nonlinear crystal and that due to diffraction. The effective loss constant α_m is chosen so that $\exp(-\alpha_m l)$ is the total attenuation in intensity per resonator pass at ω_m, where l is the crystal length.

If the parametric gain is sufficiently high to overcome the losses, steady-state oscillation results. When this is the case,

$$\frac{dA_1}{dz} = \frac{dA_2^*}{dz} = 0 \tag{8.4-27}$$

and thus the power gained via the parametric interaction just balances the losses. Putting $d/dz = 0$ in (8.4-25) gives

$$-\frac{1}{2}\alpha_1 A_1 - i\frac{g}{2}A_2^* = 0$$
$$i\frac{g^*}{2}A_1 - \frac{1}{2}\alpha_2 A_2^* = 0 \tag{8.4-28}$$

The condition for nontrivial solutions for A_1 and A_2^* is that the determinant of (8.4-28) vanish; that is,

$$\det\begin{vmatrix} -\frac{\alpha_1}{2} & -i\frac{g}{2} \\ i\frac{g^*}{2} & -\frac{\alpha_2}{2} \end{vmatrix} = 0$$

and, therefore,

$$|g|^2 = \alpha_1\alpha_2 \tag{8.4-29}$$

This is the *threshold condition* for *parametric oscillation.*

We have shown above that the pair of signals (ω_1) and idler frequencies that are caused to oscillate by parametric pumping at ω_3 satisfy the condition $k_3 = k_1 + k_2$. Using $k_i = \omega_i n_i/c$, we can write it as

$$\omega_3 n_3 = \omega_1 n_1 + \omega_2 n_2 \tag{8.4-30}$$

In a crystal the indices of refraction generally depend, as shown in Section 8.3, on the frequency, crystal orientation (if the wave is extraordinary), electric field (in electro-optic crystals), and the temperature. If, as an example, we change the crystal orientation in the oscillator, the oscillation frequencies ω_1 and ω_2 will change so as to compensate for the change in indices, and thus condition (8.4-30) will be satisfied at the new frequencies. Such an angular dependence can be employed as a tuning mechanism for tunable OPO. We will discuss the tuning of OPO in the last section of this chapter.

Frequency Up-Conversion

Parametric interactions in a crystal can be used to convert a signal from a "low" frequency ω_1 to a "high" frequency ω_3 by mixing it with a strong laser beam at ω_2, where

$$\omega_1 + \omega_2 = \omega_3 \tag{8.4-31}$$

This is particularly useful in situations where the detection system is more efficient in the spectral regime around ω_3. Using the quantum mechanical photon picture, we can consider the basic process taking place in frequency up-conversion as one in which a signal (ω_1) photon and a pump (ω_2) photon are annihilated while, simultaneously, one photon at ω_3 is generated. Since a photon energy is $\hbar\omega$; conservation of energy dictates that $\omega_3 = \omega_1 + \omega_2$ and, in a manner similar to Equation (8.4-15), the conservation of momentum leads to the relationship

$$\mathbf{k}_3 = \mathbf{k}_1 + \mathbf{k}_2 \tag{8.4-32}$$

between the wavevectors at the three frequencies. This point of view also suggests that the number of output photons at ω_3 cannot exceed the input number of photons at ω_1.

The analysis of frequency up-conversion starts with Equations (8.4-7). Assuming negligible depletion of the pump wave A_2, no losses ($\alpha = 0$) at ω_1 and ω_3, and taking $\Delta k = 0$, we can write the first and third of these equations as

$$\frac{dA_1}{dz} = -i\frac{g}{2}A_3$$
$$\frac{dA_3}{dz} = -i\frac{g}{2}A_1 \tag{8.4-33}$$

where, choosing without loss of generality the pump phase as zero so that $A_2(0) = A_2^*(0)$,

$$g \equiv \sqrt{\left(\frac{\mu}{\varepsilon_0}\right)\frac{\omega_1\omega_3}{n_1n_3}}\,E_2 d \tag{8.4-34}$$

where E_2 is the amplitude of the electric field of the pump laser. Taking the input waves with initial (complex) amplitudes $A_1(0)$ and $A_3(0)$, the general solutions of (8.4-33) are

$$A_1(z) = A_1(0)\cos\left(\frac{g}{2}z\right) - iA_3(0)\sin\left(\frac{g}{2}z\right)$$
$$A_3(z) = A_3(0)\cos\left(\frac{g}{2}z\right) - iA_1(0)\sin\left(\frac{g}{2}z\right) \tag{8.4-35}$$

In the case of a single (low) frequency input at ω_1, we have $A_3(0) = 0$. In this case,

$$|A_1(z)|^2 = |A_1(0)|^2\cos^2\left(\frac{g}{2}z\right)$$
$$|A_3(z)|^2 = |A_3(0)|^2\sin^2\left(\frac{g}{2}z\right) \tag{8.4-36}$$

therefore,

$$|A_1(z)|^2 + |A_3(z)|^2 = |A_1(0)|^2 \tag{8.4-37}$$

In the discussion following Equation (8.4-5), we pointed out that $|A_l(z)|^2$ is proportional to the photon flux (photons per square meter per second) at ω_l. Using this fact, we may interpret (8.4-35) as stating that the photon flux at ω_1 plus that at ω_3 at any plane z is a constant equal to the input ($z = 0$) flux at ω_1. If we rewrite (8.4-36) in terms of powers, we obtain

$$P_1(z) = P_1(0)\cos^2\left(\frac{g}{2}z\right)$$
$$P_3(z) = \frac{\omega_3}{\omega_1}P_1(0)\sin^2\left(\frac{g}{2}z\right) \tag{8.4-38}$$

In a crystal of length l, the conversion efficiency is thus

$$\frac{P_3(l)}{P_1(0)} = \frac{\omega_3}{\omega_1}\sin^2\left(\frac{g}{2}l\right) \tag{8.4-39}$$

and can have a maximum value of ω_3/ω_1, corresponding to the case in which all the input (ω_1) photons are converted to ω_3 photons.

In most practical situations, the conversion efficiency is small (see the following example) so using $\sin x \sim x$ for $x \ll 1$, we get

$$\frac{P_3(l)}{P_1(0)} \approx \frac{\omega_3}{\omega_1}\left(\frac{g^2l^2}{4}\right)$$

which, by the use of Equations (8.4-34) and (8.4-5), can be written

$$\frac{P_3(l)}{P_1(0)} \approx \frac{\omega_3^2 l^2 d^2}{2n_1 n_2 n_3}\left(\frac{\mu}{\varepsilon_0}\right)^{3/2}\left(\frac{P_2}{A}\right) \tag{8.4-40}$$

where A is the cross-sectional area of the interaction region.

EXAMPLE: FREQUENCY UP-CONVERSION

The main practical interest in parametric frequency up-conversion stems from the fact that it offers a means of detecting infrared radiation (a region where detectors are either inefficient, very slow, or require cooling to cryogenic temperatures) by converting the frequency into the visible or near-visible part of the spectrum. The radiation can then be detected by means of efficient and fast detectors such as photomultipliers or photodiodes; see References [9–12].

As an example of this application, consider the problem of up-converting a 10.6 μm signal, originating in a CO_2 laser, to 0.96 μm by mixing it with the 1.06 μm output of a Nd^{3+}:YAG laser. The nonlinear crystal chosen for this application has to have low losses at 1.06 μm and 10.6 μm, as well as at 0.96 μm. In addition, its birefringence has to be such as to make phase matching possible. The crystal proustite (Ag_3AsS_3) meets these requirements [12].

Using the data

$$\frac{P_{1.06\mu m}}{A} = 10^4 \text{ W/cm}^2 = 10^8 \text{ W/cm}^2$$

$l = 10^{-2}$ m

$n_1 \sim n_2 \sim n_3 = 2.6$ (an average number based on the data of Reference [12])

$d_{\text{eff}} = 1.1 \times 10^{-22}$ (MKS) (taken conservatively as a little less than half the value for d_{22})

we obtain, from Equation (8.4-40),

$$\frac{P_{\lambda=0.96\mu m}(l = 1 \text{ cm})}{P_{\lambda=10.6\mu m}(l = 0)} = 7.1 \times 10^{-4}$$

indicating a useful amount of conversion efficiency.

8.5 QUASI PHASE MATCHING

An alternative technique for achieving phase matching in crystals is referred to as *quasi phase matching* [13], a reference to a crystal fashioned in such a way that the direction of one of its principal axes, say, z, is reversed periodically. This, in properly chosen crystal orientation and polarization directions of the participating optical fields, results in a periodic modulation of the nonlinear coefficient tensor element d_{ijk} responsible for the interaction. The coupled wave equations (8.4-2) remain unchanged, except that d is replaced by $d(z)$, which, being periodic, can be expanded in a Fourier series

$$d(z) = d_{\text{bulk}}\left(\sum_{m=-\infty}^{\infty} c_m \exp\left(im\frac{2\pi}{\Lambda}z\right)\right) \tag{8.5-1}$$

where c_m are the Fourier expansion coefficients and Λ is the period of $d(z)$. The effect on the first of Equations (8.4-2), as an example, is to transform it to

$$\frac{dE_1}{dz} = -\frac{i\omega_1}{2}\sqrt{\frac{\mu}{\varepsilon_1}}\, d_{\text{bulk}} E_3 E_2^* \sum_{m=-\infty}^{\infty} c_m \exp\left[i\left(m\frac{2\pi}{\Lambda} - k_3 + k_2 + k_1\right)z\right] \tag{8.5-2}$$

Phase matching can be obtained if, for some integer m, the condition

$$m\frac{2\pi}{\Lambda} = k_3 - k_2 - k_1 \tag{8.5-3}$$

is satisfied. Ignoring non-phase-matched terms in Equation (8.5-2) (their contribution averages out to zero over distances that are large compared to the coherence length), we rewrite (8.5-2) as

$$\frac{dE_1}{dz} = -i\frac{\omega_1}{2}\sqrt{\frac{\mu}{\varepsilon_1}}\, d_{\text{bulk}} c_m \exp\left[i\left(m\frac{2\pi}{\Lambda} - k_3 + k_2 + k_1\right)z\right] \tag{8.5-4}$$

$$c_m = \frac{1}{\Lambda}\int_0^{\Lambda} \frac{d(z)}{d_{\text{bulk}}} \exp\left(-im\frac{2\pi}{\Lambda}z\right) dz \tag{8.5-5}$$

where m is an integer that satisfies Equation (8.5-3).

The simplest case of a spatially periodic $d(z)$ is one in which $d(z)$ switches from d_{bulk} to $-d_{\text{bulk}}$ every $\Lambda/2$. In this case,

$$c_m = \frac{1 - \cos m\pi}{m\pi} \quad \text{for } m \neq 0 \tag{8.5-6}$$

so that, choosing $m = 1$, the effective nonlinear constant is

$$d_{\text{eff}} = c_m d_{\text{bulk}} = \frac{2}{\pi} d_{\text{bulk}} \tag{8.5-7}$$

For $m = 1$ quasi phase matching, the spatial period is given by $\Lambda = 2\pi/\Delta k = 2L_c$, according to Equation (8.5-3). It is clear from Equation (8.5-4) that, in principle, quasi-phase-matched configurations can give rise to the same conversion efficiency as in the ideal, $\Delta k = 0$, phase-matched case, except that we require a longer interaction path to achieve it. The length penalty factor is $d_{\text{bulk}}/d_{\text{eff}} = c_m^{-1}$.

To appreciate quasi phase matching on an intuitive basis, we note that it involves reversing the sign of the nonlinear interaction at

$$z = L_c, 2L_c, \ldots \quad \text{with} \quad L_c = \frac{\pi}{\Delta k} \tag{8.5-8}$$

These are the locations where the amplitude of the newly generated second-harmonic (SH) wave is opposite in sign relative to the already generated second-harmonic wave in the non-phase-matched case. Reversing the sign of the newly generated SH wave leads to constructive buildup of $E^{2\omega}(z)$. This keeps the power flowing from ω to 2ω along the length of the crystal and leads to cumulative buildup of $E^{2\omega}(z)$. The buildup of $E^{2\omega}(z)$ via quasi phase matching can best be visualized using a phasor plot of the interaction. Taking the specific case of second-harmonic generation as an example, we can divide the interaction path L into sufficiently short segments, each of length δz, such that $(\Delta k)\,\delta z \ll \pi$ and obtain from Equation (8.3-23)

$$\Delta E^{(2\omega)}(z) = -i\omega\sqrt{\frac{\mu}{\varepsilon}}\, d(E^{(\omega)})^2 e^{i(\Delta k)z}\delta z \tag{8.5-9}$$

where $\Delta E^{(2\omega)}$ is the complex increment to the phasor $E^{(2\omega)}(z)$ due to the segment of length δz centered at z. By adding the increments $\Delta E^{(2\omega)}$ vectorially, or rather phasorially, we obtain the phasor diagram shown in Figure 8.7. In the non-phase-matched case (Figure 8.7a), the generated second-harmonic field keeps growing, reaching a maximum, usually insignificantly small, at

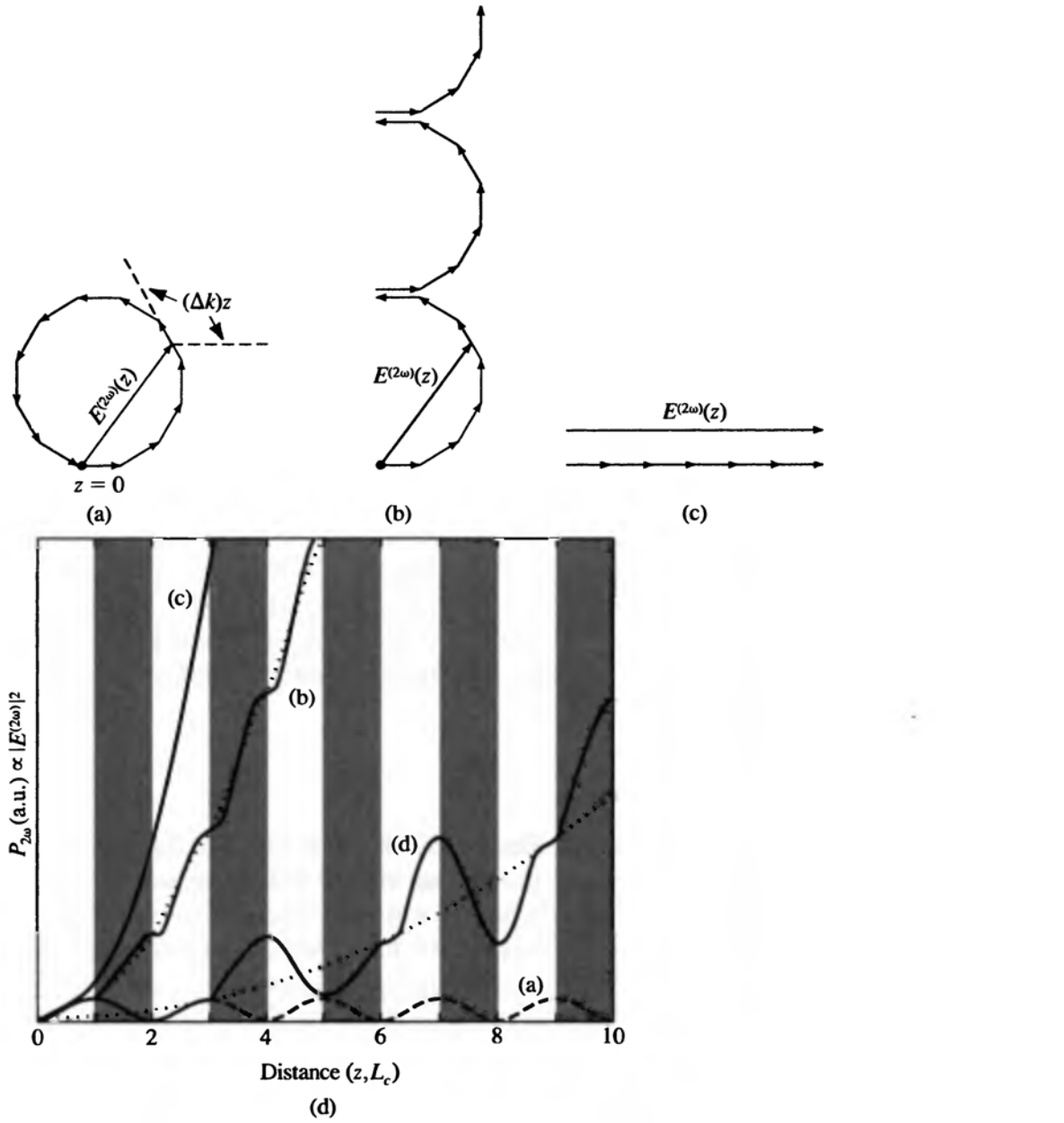


Figure 8.7 Evolution of the second-harmonic phasor $E^{(2\omega)}(z)$ in (a) a non-phase-matched case, (b) quasi-phase-matched case, and (c) bulk birefringent phase matching case ($\Delta k = 0$). (d) The second-harmonic power field $E^{(2\omega)}$ in a crystal for cases (a), (b), and (c), as well as curve (d) for quasi-phase-matched operation using the third Fourier coefficient $m = 3$, $L_c \equiv \pi/\Delta k$. Part (d) is reproduced from Reference [14].

$z = \pi/\Delta k$. At longer distances, $E^{(2\omega)}$ begins to shrink, returning to zero at $z = 2\pi/\Delta k$. In the quasi-phase-matched case (Figure 8.7b), the sign of the interaction reverses every $\Delta z = \pi/\Delta k$. This is done by reversing the sign of $d(z)$. The resultant $E^{(2\omega)}(z)$ thus keeps growing monotonically, albeit at a (spatial) rate smaller than in the case of ideal bulk phase matched (Figure 8.7c).

Quasi Phase Matching in Crystal Dielectric Waveguides

Quasi phase matching has been practiced almost exclusively in configurations involving dielectric waveguiding. The main reason is that in a single-mode waveguide, the small

cross-sectional area ($\sim 10^{-8}$ cm^2) leads to very large intensities even at modest input power levels. This enables efficient conversion efficiencies with reasonable, say, <1 cm, crystal length. In addition, the lack of diffraction makes it possible to maintain the high intensity throughout the crystal length. Most of the work, to date, involves optical waveguides fabricated in $LiNbO_3$ [15]. The spatial modulation of the nonlinear coefficient is achieved by reversing locally the direction of the c axis of the crystal by means of periodically applied electric fields [16, 17] or diffusion of Ti through openings in a mask with a period $\Lambda = 2L_c$ [17]. This periodic reversal is caused by a periodic reversal of the direction of the crystal's permanent electric polarization by a periodically (spatially) reversing electric field or impurity diffusion.

8.6 THIRD-ORDER NONLINEAR OPTICAL PROCESSES

The third-order nonlinear polarization

$$P_i^{(3)} = 4\chi_{ijkl}E_jE_kE_l \tag{8.6-1}$$

is responsible for third-harmonic generation (THG), stimulated Raman scattering (SRS), stimulated Brillouin scattering (SBS), four-wave mixing (FWM), optical Kerr effect, and so on. In this section, we will treat the coupling of waves due to the third-order nonlinearity. Particularly, we will treat some of the important third-order processes that are relevant to optical communications.

Optical Kerr Effect

It was first discovered by J. Kerr in 1875 that a transparent liquid becomes doubly refracting (birefringent) when placed in a strong electric field. Because the effect is quadratic with respect to the applied electric field E, it is sometimes known as the quadratic electro-optical effect. The amount of birefringence due to the Kerr effect can be written

$$\Delta n = n_{\parallel} - n_{\perp} = \lambda K E^2 \tag{8.6-2}$$

where K is the Kerr constant and λ is the vacuum wavelength. In Equation (8.6-2), $n_{\parallel}$ is the refractive index for the optical wave polarized along the direction of the applied electric field, while $n_{\perp}$ is for the optical wave polarized perpendicular to the applied electric field.

According to Equation (8.6-1), the presence of an optical beam in a nonlinear medium can cause a change of the index of refraction. Let the field of an optical beam be written

$$E_j(t) = \mathrm{Re}(E_{0j}^{\omega}e^{i\omega t}) = \tfrac{1}{2}(E_{0j}^{\omega}e^{i\omega t} + \text{c.c.}) \quad (j = x, y, z) \tag{8.6-3}$$

where E_{0j}^{ω} is a constant. The third-order nonlinear polarization can be written

$$P_i^{(3)}(t) = 4\chi_{ijkl}\tfrac{1}{2}(E_{0j}^{\omega}e^{i\omega t} + \text{c.c.})\tfrac{1}{2}(E_{0k}^{\omega}e^{i\omega t} + \text{c.c.})\tfrac{1}{2}(E_{0l}^{\omega}e^{i\omega t} + \text{c.c.}) \tag{8.6-4}$$

where summations over repeated indices (j, k, l) are assumed.

We focus our attention on the amplitude of the nonlinear term that oscillates at the same frequency as the input optical beam ω. Ignoring all other terms, we obtain the amplitude of the nonlinear polarization that oscillates at frequency ω:

$$P_{\mathrm{NL}i}^{(\omega)} = \chi_{ijkl}\{E_{0j}^{\omega}E_{0k}^{\omega}E_{0l}^{\omega *} + E_{0j}^{\omega *}E_{0k}^{\omega}E_{0l}^{\omega} + E_{0j}^{\omega}E_{0k}^{\omega *}E_{0l}^{\omega}\} \tag{8.6-5}$$

where summations over repeated indices (j, k, l) are assumed. This nonlinear polarization oscillating at frequency ω is equivalent to a self-induced change in the dielectric susceptibility.

Consider a simple case of an optical beam polarized along the x axis in an isotropic medium with an amplitude E_{0x}^{ω}. The amplitude of nonlinear polarization at ω is written

$$P_{\mathrm{NL}x}^{(\omega)} = 3\chi_{xxxx}E_{0x}^{\omega}(E_{0x}^{\omega}E_{0x}^{\omega *}) \tag{8.6-6}$$

This is equivalent to a change of the linear susceptibility $\Delta\chi$. We recall that $n^2 = 1 + \chi$. Thus the third-order nonlinear polarization leads to a change of the index of refraction of

$$\Delta n = n_2 I = \frac{\Delta\chi}{2n} = \frac{3\chi_{xxxx}}{2n\varepsilon_0}|E_{0x}|^2 = \frac{3c\mu_0\chi_{xxxx}}{n^2\varepsilon_0} I \tag{8.6-7}$$

where I is the optical beam intensity and n_2 is known as the Kerr coefficient. For silica fibers doped with germanium oxide, the Kerr coefficient is around (depending on the density of impurities)

$$n_2 \approx 3.0 \times 10^{-20}\ \mathrm{m^2/W} \tag{8.6-8}$$

Like most of the third-order nonlinear processes, the Kerr effect is relatively weak in most media. In optical fibers, the nonlinear Kerr effect can be significant due to the high optical intensity in the core and the long interaction lengths afforded by these waveguides. The third-order nonlinearity in optical fibers can lead to four-wave mixing, self-phase modulation, cross-phase modulation, and so on.

According to Equation (8.6-7), the Kerr coefficient for a beam of linearly polarized light is given by

$$n_2 = \frac{3c\mu_0\chi_{xxxx}}{n^2\varepsilon_0} = \frac{3}{cn^2\varepsilon_0^2}\chi_{xxxx} \tag{8.6-9}$$

For a general polarization state, the Kerr coefficient is slightly different from (8.6-9). Let the electric field of the optical wave with an arbitrary polarization state be written

$$\mathbf{E}(t) = \tfrac{1}{2}(\hat{\mathbf{x}}E_{0x}^{\omega}e^{i\omega t} + \hat{\mathbf{y}}E_{0y}^{\omega}e^{i\omega t} + \text{c.c.}) \tag{8.6-10}$$

where E_{0x} and E_{0y} are the components. A similar derivation leads to the following amplitudes of the nonlinear polarization at frequency ω (see also Section 14.2):

$$\begin{aligned} P_{\mathrm{NL}x}^{(\omega)} &= \chi_{xxxx}\{(3|E_{0x}^{\omega}|^2 + 2|E_{0y}^{\omega}|^2)E_{0x}^{\omega} + E_{0y}^{\omega 2}E_{0x}^{\omega *}\} \\ P_{\mathrm{NL}y}^{(\omega)} &= \chi_{xxxx}\{(3|E_{0y}^{\omega}|^2 + 2|E_{0x}^{\omega}|^2)E_{0y}^{\omega} + E_{0x}^{\omega 2}E_{0y}^{\omega *}\} \end{aligned} \tag{8.6-11}$$

where we have used the following symmetry relationships (see Section 14.2 and Problem 14.8):

$$\chi_{1111} = \chi_{1122} + \chi_{1212} + \chi_{1221} \quad \text{and} \quad \chi_{1111} = \chi_{2222} \tag{8.6-12}$$

We note that cross-phase modulation exists when both polarization components are present. In other words, the amplitude of the x component can induce a change of the refractive index for y-polarized light, and vice versa. Kerr coefficients for circularly polarized light and unpolarized light can still be defined. Using Equation (8.6-10), we can show that the Kerr coefficients are given by (see Problem 8.14)

$$n_{2,\mathrm{unpolarized}} = \frac{5}{6}n_2 = \frac{5}{2cn^2\varepsilon_0^2}\chi_{xxxx} \tag{8.6-13}$$

and

$$n_{2,\mathrm{circular}} = \frac{2}{3}n_2 = \frac{2}{cn^2\varepsilon_0^2}\chi_{xxxx} \tag{8.6-14}$$

TABLE 8.5 Kerr Coefficients of Some Selected Substances

Material	Wavelength (μm)	Index of Refraction	χ_{1111} (esu) (10^{-13})	χ_{1111} (MKS) (10^{-33})	n_2 (MKS) (10^{-18} m^2/W)
CCl_4	0.694	1.454	0.05	0.62	0.04
CS_2	1.06	1.594	4.65	57.50	2.89
	0.694	1.612	3.57	44.10	2.16
Lucite	1.06	1.49	0.11	1.34	0.08
Ruby	1.06	1.76	0.07	0.85	0.04
SF_6 glass	1.06	1.77	0.42	5.22	0.21
SiO_2	0.694	1.455	0.01	0.15	0.01
YAG	1.06	1.83	0.15	1.90	0.07
	0.694	1.829	0.06	0.74	0.03

Data adapted from Reference [18].

Table 8.5 lists the Kerr coefficients of selected nonlinear materials. Many of the early publications on third-order susceptibility are in units of esu. We have converted them into MKS units for convenience (see Problem 8.13). According to (8.6-11), a beam of elliptically polarized light may undergo a change of the state of polarization as the beam propagates in a Kerr medium.

Stimulated Raman Scattering

The Raman effect, discovered in 1928 by Sir Chandrasekhera Venkata Raman, is an inelastic scattering of light by elementary excitations of matter. When a photon interacts with an atomic system (e.g., molecule), it can be scattered in one of the three ways: (1) it can be elastically scattered and thus retain its incident photon energy (Rayleigh scattering), (2) it can be inelastically scattered by quasi-particle excitations (e.g., molecular vibration) of the atomic system, thereby either giving energy to the medium (Stokes scattering), or (3) it can be scattered by removing energy from it (anti-Stokes scattering). Thus the Raman effect involves a coupling between incident photons and the quasi-particle excitations such as phonons (lattice vibrations). In terms of energy-level diagrams, the Raman effect is often described as in Figure 8.8.

The Raman frequency shifts ($\omega_L - \omega_S$), in the range of 10^{12}–10^{13} Hz, are determined by different vibration frequencies of the molecules. Raman spectroscopy is an excellent probe to

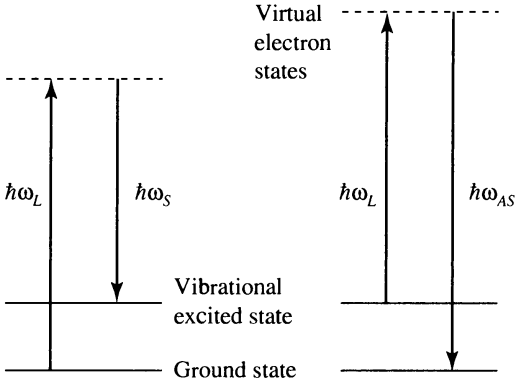


Figure 8.8 Schematic drawing of the energy diagram of Raman scattering. In Stokes scattering, the atomic system is excited from the ground state to an excited state (vibrational excited state). In anti-Stokes scattering, the atomic system is donating energy to the optical beam by decaying from an excited state to the ground state.

investigate the low-energy elementary excitations in materials as well as to characterize their structural, electronic, vibrational, and magnetic properties. As depicted in Figure 8.8, for each molecular vibration we observe two Raman lines. The Stokes and anti-Stokes lines are connected with transitions from the ground state to the first excited vibrational state and vice versa. Since the excited states are populated only slightly at room temperature, the anti-Stokes intensities are relatively small compared with the Stokes intensities. The lifetime of the molecular vibration is often deduced from the width of the Raman line.

Like the Kerr effect, Raman scattering also is relatively weak in most media. In optical fibers, however, the nonlinear effect due to Raman scattering can be significant due to the high optical intensity of a laser beam in the core and the long interaction lengths (in the range of 100 km) afforded by these waveguides. Both stimulated Raman scattering (SRS) and stimulated Brillouin scattering (SBS) occur when the light launched into the fiber exceeds a threshold power level for each process. Under the conditions of stimulated scattering, optical power is more efficiently converted from the input pump wave to the scattered Stokes wave.

The scattered wave is frequency shifted from the pump and in the case of SBS propagates in the opposite direction. This means that the amount of optical power leaving the far end of the fiber no longer increases linearly with the input power. The maximum launch power becomes clamped and the excess is simply reflected back out of the fiber. For long distance or highly branched fiber links, it is important that as much power as possible can be launched into the fiber to compensate for attenuation and power splitting. Limits on the maximum output power due to SBS must therefore be avoided.

In SRS, the Stokes wave can be shifted from the pump wave by typically 10–100 nm and continues to propagate forward along the fiber with the pump wave. If the pump is actually one channel of a multiwavelength WDM communications system, then its Stokes wave may overlap with other channels at longer wavelengths—leading to crosstalk and Raman amplification. Raman amplification causes shorter wavelength channels to experience power depletion and act as a pump for the amplification of longer wavelength channels. This can skew the power distribution among the WDM channels—reducing the signal-to-noise ratio of the lowest frequency channels and introducing crosstalk on the high-frequency channels. Both of these effects can lower the information-carrying capacity of the optical system. On the other hand, with a proper spacing of the channels, SRS can be employed for the amplification of signal waves. This subject is taken up in Chapter 17.

We now consider the energy coupling in SRS. For simplicity, we assume that both the laser and the Stokes scattering are polarized along the x axis. The electric field can be written

$$E_x(t) = \mathrm{Re}(E_L e^{i(\omega_L t - k_L z)} + E_S e^{i(\omega_S t - k_S z)}) = \tfrac{1}{2}(E_L e^{i(\omega_L t - k_L z)} + E_S e^{i(\omega_S t - k_S z)} + \text{c.c.}) \tag{8.6-15}$$

where E_L and E_S are the field amplitudes, ω_L and ω_S are frequencies, and k_L and k_S are wavenumbers. Using Equation (8.6-1), the amplitudes of the nonlinear polarization oscillating at frequencies ω_L and ω_S can be written

$$P_{\mathrm{NL}}^{(\omega_L)} = E_L(6\chi_{R1}^{(3)} E_S E_S^* + 3\chi_1^{(3)} E_L E_L^*) \tag{8.6-16}$$

$$P_{\mathrm{NL}}^{(\omega_S)} = E_S(3\chi_2^{(3)} E_S E_S^* + 6\chi_{R2}^{(3)} E_L E_L^*) \tag{8.6-17}$$

where we use different effective third-order nonlinear susceptibilities to account for the possible dependence on the frequencies. By examining the above equations, we note that the $\chi_1^{(3)}$ and $\chi_2^{(3)}$ terms in the nonlinear polarization only act to modify the dielectric constant (or index of refraction). They are responsible for the optical Kerr effect discussed earlier. These terms have no direct effect on stimulated Raman scattering and thus will be neglected in the

following discussion. The $\chi_{R1}^{(3)}$ and $\chi_{R2}^{(3)}$ terms in the nonlinear polarization, on the other hand, effectively couple E_L and E_S and can cause energy exchange between these two waves. These two nonlinear susceptibilities are responsible for the stimulated Raman process and are often called Raman susceptibilities. Using quantum mechanical theory, it is possible to show that [19]

$$\chi_{R1}^{(3)} = \chi_{R2}^{*(3)} \tag{8.6-18}$$

In other words, these two Raman susceptibilities are complex conjugates of each other.

Following the procedure leading to the coupled equations discussed in Section 8.3, we obtain

$$\frac{dE_L}{dz} = -\frac{1}{2}\alpha_L E_L + i\frac{6\mu_0\omega_L^2}{k_L}\chi_{R1}^{(3)}|E_S|^2 E_L \tag{8.6-19}$$

$$\frac{dE_S}{dz} = -\frac{1}{2}\alpha_S E_S + i\frac{6\mu_0\omega_S^2}{k_S}\chi_{R2}^{(3)}|E_L|^2 E_S \tag{8.6-20}$$

where we have added terms (α_L and α_S) to account for the linear attenuation due to absorption and scattering. These are the equations that govern the energy coupling between the laser (pump) beam and the Stokes beam.

To understand the Raman gain, first we consider a simplified problem in which the laser amplitude E_L is assumed constant with z. This applies when the energy depletion from the laser field (at ω_L) by the nonlinear interaction is small compared to the laser beam energy. This makes it possible to integrate Equation (8.6-20) for the Stokes field (at ω_S). We obtain, according to Equation (8.6-20),

$$|E_S|^2(z) = |E_S|^2(0)\exp(G_R z - \alpha_S z) \tag{8.6-21}$$

where the Raman gain coefficient G_R is given by

$$G_R = \frac{12\mu_0\omega_S^2}{k_S}|E_L|^2\,\mathrm{Im}(\chi_{R1}^{(3)}) = -\frac{12\mu_0\omega_S^2}{k_S}|E_L|^2\,\mathrm{Im}(\chi_{R2}^{(3)}) \tag{8.6-22}$$

We note that the Stokes beam is amplified with an exponential gain coefficient G_R that is proportional to the laser beam intensity. Thus a Raman medium pumped by a laser beam can be employed as an optical amplifier for the spectral regime around the Stokes line.

In stimulated Raman scattering (SRS), there's no input Stokes beam. The laser beam is focused onto a Raman medium in which the laser beam is trapped in the form of a thin filament with a diameter of a few microns. This self-induced trapping is due to a combination of the optical Kerr effect described earlier and the Gaussian intensity profile of the laser beam. This effect gives rise to an extremely large laser beam intensity and hence a very large Raman gain coefficient at the Stokes line. This gain is so large that any electromagnetic radiation at the Stokes line is amplified significantly in a single pass, leading to stimulated Raman scattering.

As we indicated, the Raman scattering involves the interaction of the input laser beam with the molecular vibration. The scattering process creates the Stokes wave at a lower frequency. Once the Stokes wave is generated, it beats with the input laser wave at a beat frequency, $\omega_L - \omega_S$, that is automatically equal to that of the molecular vibration frequency. Therefore a positive feedback loop is created as the beating term acts as a driving force that further increases the amplitude of the molecular vibration, which in turn increases the amplitude of the Raman scattered wave.

In nonlinear optical interaction, where photons of different frequencies are involved, it is convenient to describe the energy exchange in terms of photon flux, which is a measure of the number of photons per unit area per unit time. The photon flux for a beam of light with an electric field amplitude E is defined as

$$\phi = \frac{I}{\hbar\omega} = \frac{I}{\hbar\omega}\frac{k}{2\mu_0\omega}|E|^2 \tag{8.6-23}$$

where I is the beam intensity. Using photon fluxes, the coupled equations can be written

$$\frac{d}{dz}\phi_L = -\alpha_L\phi_L - G'_R\phi_L\phi_S \tag{8.6-24}$$

$$\frac{d}{dz}\phi_S = -\alpha_S\phi_S + G'_R\phi_L\phi_S \tag{8.6-25}$$

where the Raman gain coefficient for photon fluxes is given by

$$G'_R = \frac{12\hbar\mu_0^2\omega_L^2\omega_S^2}{k_Lk_S}\,\mathrm{Im}(\chi^{(3)}_{R1}) = -\frac{12\hbar\mu_0^2\omega_L^2\omega_S^2}{k_Lk_S}\,\mathrm{Im}(\chi^{(3)}_{R2}) \tag{8.6-26}$$

We note that Equations (8.6-24) and (8.6-25) are consistent with the conservation of photon flux in the event when the linear attenuation coefficients vanish ($\alpha_L = \alpha_S = 0$). In other words, one laser photon (at ω_L) is converted into one Stokes photon (at ω_S).

In the event when the difference between the two linear attenuation coefficients can be neglected (i.e., $\alpha_L = \alpha_S = \alpha$), exact solutions of the coupled equations (8.6-24) and (8.6-25) are available. They are written

$$\phi_L(z) = \phi_L(0)\frac{1 + m^{-1}}{1 + m^{-1}\exp\left(\dfrac{\gamma}{\alpha}(1 - e^{-\alpha z})\right)}\exp(-\alpha z) \tag{8.6-27}$$

$$\phi_S(z) = \phi_S(0)\frac{1 + m}{1 + m\exp\left(-\dfrac{\gamma}{\alpha}(1 - e^{-\alpha z})\right)}\exp(-\alpha z) \tag{8.6-28}$$

where the exponential gain coefficient γ is given by

$$\gamma = G'_R[\phi_S(0) + \phi_L(0)] \tag{8.6-29}$$

and m is the photon flux ratio at input ($z = 0$)

$$m = \phi_L(0)/\phi_S(0) \tag{8.6-30}$$

In the event when the linear attenuation coefficient becomes zero ($\alpha = 0$), the solutions reduce to

$$\phi_L(z) = \phi_L(0)\frac{1 + m^{-1}}{1 + m^{-1}\exp(\gamma z)} \tag{8.6-31}$$

$$\phi_S(z) = \phi_S(0)\frac{1 + m}{1 + m\exp(-\gamma z)} \tag{8.6-32}$$

In typical stimulated Raman scattering, the initial Stokes photon flux is very small, $\phi_S(0) \ll \phi_L(0)$ (or $1 \ll m$). According to Equation (8.6-32), we note that the Stokes photon flux increases exponential initially and then reaches saturation as the laser beam is depleted.

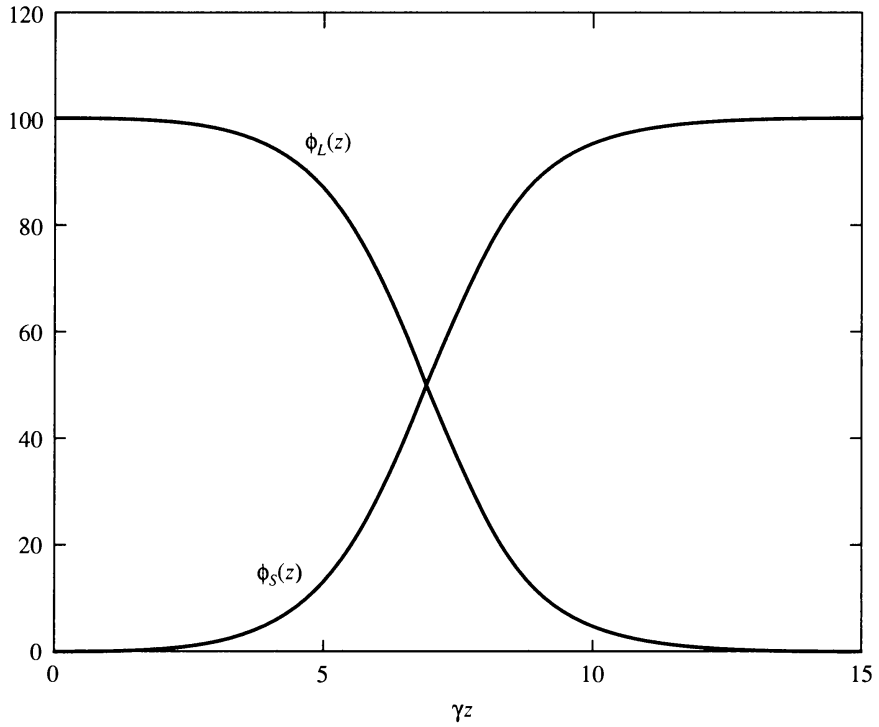


Figure 8.9 Theoretical plot of the photon fluxes of the laser beam and the Stokes beam as a function of interaction distance γz for the case when $\alpha = 0$.

Figure 8.9 shows the energy exchange in terms of the photon flux as a function of interaction distance.

Based on the quantum mechanical energy diagrams shown in Figure 8.8, the peak of Raman emission occurs when the frequency shift equals the molecular vibration frequency. In other words,

$$\omega_L - \omega_S = \omega_v \tag{8.6-33}$$

where ω_v is the molecular vibration frequency.

We conclude by noting that the nonlinear polarization (8.6-17) at the Stokes line (with frequency ω_S) has a space–time dependence of $\exp(i\omega_S t - ik_S z)$, which is exactly the same as the Stokes field. This explains why the "phase-matching" conditions that had to be satisfied in parametric interactions are automatically satisfied in stimulated Raman scattering. Table 8.6 provides a list of Raman properties for selected materials.

TABLE 8.6 Raman Properties for Selected Materials

Material	Raman Shift (THz)	Linewidth (GHz)	Raman Gain (10^{-3} cm/MW)
Liquid oxygen, O_2	46.6	3.5	14.5 ± 4
Liquid nitrogen, N_2	69.8	2.01	17 ± 5
Benzene, C_6H_6	29.8	64.5	2.8
CS_2	19.7	15	24
Nitrobenzene	40.4	198	2.1
$LiNbO_3$	7.74	210	28.7
$LiTaO_3$	6.45	360	10
SiO_2	14		0.8

From Reference [19].

EXAMPLE: RAMAN GAIN IN SiO_2

Consider a laser beam power of 10 mW in a single-mode fiber with an effective core area of 100 μm^2. The intensity is thus $I = 10^8$ W/m^2. Using the gain coefficient of 0.8×10^{-3} cm/MW, we obtain a Raman gain coefficient of 8×10^{-4} m^{-1}.

8.7 STIMULATED BRILLOUIN SCATTERING

In Raman scattering, incoming photons are interacting with the molecules, causing molecular vibrations. The scattered photons are downshifted in frequency by an amount that equals the internal vibration frequency of the molecules. The frequency shift is in the range of 10^{12}–10^{13} Hz. Similar inelastic scatterings can occur in which the frequency shift is caused by the physical motion of the atoms or molecules. For example, spontaneous Brillouin scattering involves the interaction with thermally excited acoustic waves. Such a scattering of light was considered as early as 1922 by L. Brillouin. The frequency shift in Brillouin scattering is in the range of 10^{10} Hz.

Similar to SRS, stimulated Brillouin scattering (SBS) occurs when the pump laser beam is intense enough. The theory of the nonlinear interaction in SBS follows the formalism of SRS given in Section 8.6. In general, stimulated Brillouin scattering can occur in all directions. Here we consider the most important case of contradirectional Brillouin emission. Let the electric field (assume x-polarized) be written

$$E_x(t) = \text{Re}(E_L e^{i(\omega_L t - k_L z)} + E_B e^{i(\omega_B t + k_B z)}) = \tfrac{1}{2}(E_L e^{i(\omega_L t - k_L z)} + E_B e^{i(\omega_B t + k_B z)} + \text{c.c.}) \tag{8.7-1}$$

where E_L and E_B are the field amplitudes, ω_L and ω_B are frequencies, and k_L and k_B are wavenumbers. Note that the Brillouin beam is propagating backward. Using Equation (8.6-1), the amplitudes of the nonlinear polarization oscillating at frequencies ω_L and ω_B can be written

$$P_{\text{NL}}^{(\omega_L)} = E_L(6\chi_B^{(3)} E_B E_B^* + 3\chi_1^{(3)} E_L E_L^*) \tag{8.7-2}$$

$$P_{\text{NL}}^{(\omega_B)} = E_B(3\chi_2^{(3)} E_B E_B^* + 6\chi_B^{(3)} E_L E_L^*) \tag{8.7-3}$$

where we use different effective third-order nonlinear susceptibilities to account for the possible dependence on the frequencies. By examining the above equations, we note that the $\chi_1^{(3)}$ and $\chi_2^{(3)}$ terms in the nonlinear polarization only act to modify the dielectric constant (or index of refraction). They are responsible for the optical Kerr effect discussed earlier. These terms have no direct effect on stimulated Brillouin scattering and thus will be neglected in the following discussion. The $\chi_B^{(3)}$ terms in the nonlinear polarization, on the other hand, effectively couple E_L and E_B and can cause energy exchange between these two waves. These two nonlinear susceptibilities are responsible for the stimulated Brillouin process and are often called Brillouin susceptibilities.

Following the procedure leading to the coupled equations of SRS, we obtain

$$\frac{dE_L}{dz} = -\frac{1}{2}\alpha E_L + i\frac{6\mu_0\omega_L^2}{k_L}\chi_B^{(3)}|E_B|^2 E_L \tag{8.7-4}$$

$$\frac{dE_B^*}{dz} = +\frac{1}{2}\alpha E_B^* + i\frac{6\mu_0\omega_B^2}{k_B}\chi_B^{(3)}|E_L|^2 E_B^* \tag{8.7-5}$$

where α accounts for the linear attenuation due to absorption and other scattering loss. These are the coupled equations governing the energy exchange between the laser beam and the Brillouin beam.

To understand the Brillouin gain, first we consider a simplified problem in which the laser amplitude E_L is assumed constant with z. This applies when the energy depletion from the laser field (at ω_L) by the nonlinear interaction is small compared to the laser beam energy. This makes it possible to integrate Equation (8.7-5) for the Brillouin field (at ω_B). We obtain, according to Equation (8.7-5),

$$|E_B|^2(z) = |E_B|^2(L) \exp[(G_B - \alpha)(L - z)] \tag{8.7-6}$$

or

$$\frac{|E_B|^2(0)}{|E_B|^2(L)} = \exp[(G_B L - \alpha L)]$$

where the Brillouin gain coefficient G_B is given by

$$G_B = \frac{12\mu_0\omega_B^2}{k_B}|E_L|^2 \operatorname{Im}(\chi_B^{(3)}) \tag{8.7-7}$$

We note that the Brillouin beam is amplified with an exponential gain coefficient G_B, which is proportional to the laser beam intensity. Thus a Brillouin medium pumped by a laser beam can be employed as an optical amplifier for the spectral regime around the Brillouin line. The gain bandwidth, however, is relatively narrow.

In practice, it is legitimate to ignore the small difference between ω_L and ω_B. From the coupled equations for the field amplitudes, we obtain

$$\frac{d}{dz}|E_L|^2 = -\alpha|E_L|^2 - G_B|E_L|^2|E_B|^2 \tag{8.7-8}$$

$$\frac{d}{dz}|E_B|^2 = +\alpha|E_B|^2 - G_B|E_L|^2|E_B|^2 \tag{8.7-9}$$

or equivalently, in terms of beam intensities,

$$\frac{d}{dz}I_L = -\alpha I_L - gI_L I_B \tag{8.7-10}$$

$$\frac{d}{dz}I_B = +\alpha I_B - gI_L I_B \tag{8.7-11}$$

where g is the Brillouin gain coefficient in units of cm/MW.

The coupled equations for the beam intensities can be solved analytically in the absence of linear attenuation ($\alpha = 0$). The solutions can be written

$$I_L(z) = \frac{c_0}{1 - \rho \exp(-gc_0 z)} \tag{8.7-12}$$

$$I_B(z) = \frac{c_0}{1 - \rho \exp(-gc_0 z)}\rho \exp(-gc_0 z) \tag{8.7-13}$$

where c_0 and ρ are constants related to the incident laser beam intensity $I_L(0)$ at $z = 0$, and the noise Brillouin seed intensity $I_B(0)$ at $z = L$,

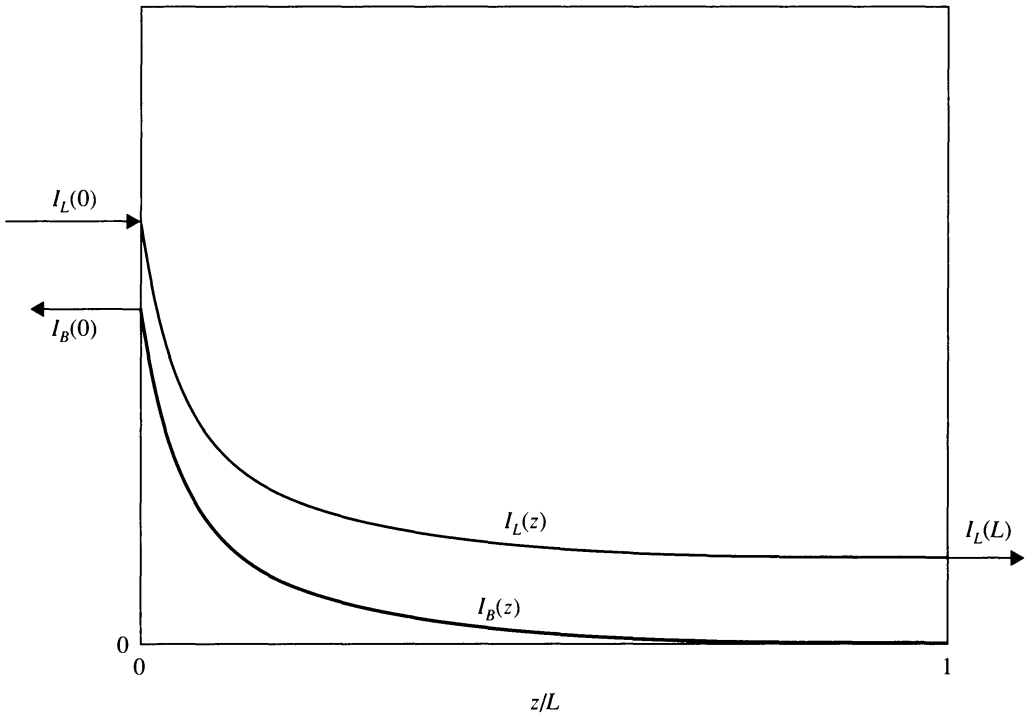


Figure 8.10 Theoretical plot of beam intensity in SBS in the absence of linear attenuation. Note that the SBS beam intensity grows from a very low noise intensity $I_B(L)$ at $z = L$ to a significant value of $I_B(0)$ at $z = 0$.

$$\frac{I_B(L)}{c_0 + I_B(L)} = \frac{I_L(0) - c_0}{I_L(0)} \exp(-gc_0 L) \tag{8.7-14}$$

$$\rho = \frac{I_L(0) - c_0}{I_L(0)} = \frac{I_B(0)}{I_L(0)} \tag{8.7-15}$$

The parameter ρ is actually the fraction of input beam power converted into the SBS beam. Figure 8.10 shows the beam intensities as a function of interaction distance z.

For transmission in optical networks, we are particularly interested in the relationship between the incident intensity (or power) $I_L(0)$ and the transmitted intensity (or power) $I_L(L)$. Using Equations (8.7-12)–(8.7-15), we obtain

$$I_L(0) = I_L(L) \frac{I_L(L) - I_B(L)}{I_L(L) - I_B(L) \exp\{g[I_L(L) - I_B(L)]L\}} \tag{8.7-16}$$

where $I_B(L)$ is an infinitesimal SBS seed intensity due to noise at $z = L$.

Although it may be difficult to invert this equation to obtain an expression of the output intensity as a function of the input intensity, it is still quite useful. Figures 8.11 and 8.12 are plots of the input–output relationship. We note that as the input intensity increases, the nonlinear SBS interaction also increases. The nonlinear interaction leads to a stronger SBS beam, and thus the transmitted beam is a nonlinear function of the input intensity. At low intensities, the transmitted beam initially increases linearly with the input intensity. At high intensities (e.g., $I_L(0) > 15$ kW/cm^2), the transmitted intensity increases only logarithmically (almost flat). In other words, the transmitted intensity does not increase proportionally with the input intensity. SBS is a nonlinear optical process. It is interesting to examine the transmitted laser beam intensity $I_L(L)$ as a function of the input laser beam intensity $I_L(0)$. This is particularly important in single-mode fibers, where the interaction length can be many kilometers, and the cumulative nonlinear optical effect can be quite strong.

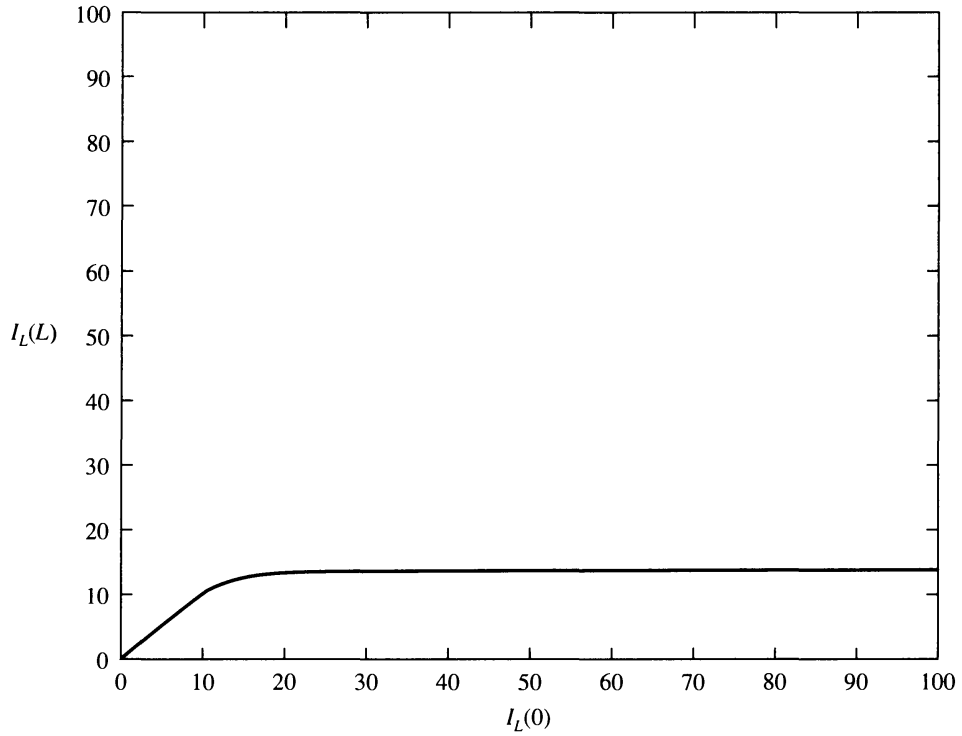


Figure 8.11 Theoretical plot of transmitted intensity versus input intensity in the SBS process (in units of kW/cm^2). We assume $gL = 0.00135$ cm^2/W and a seed intensity of $I_B(L) = 0.0001$ kW/cm^2.

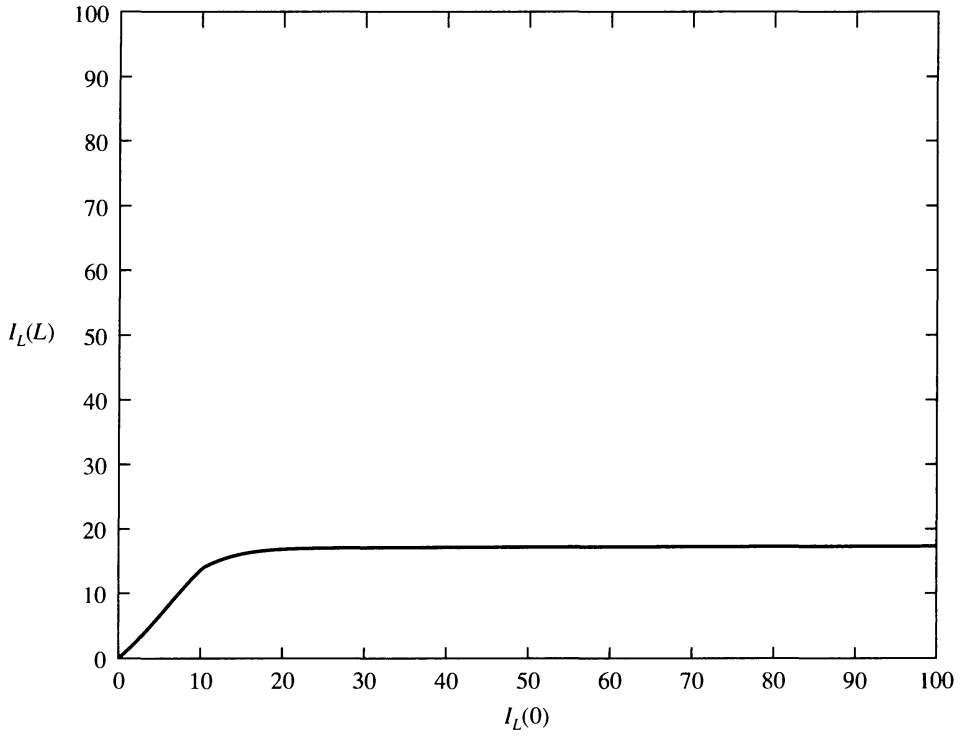


Figure 8.12 Transmitted intensity $I_L(L)$ versus input intensity $I_L(0)$ in SBS process (in units of kW/cm^2). We assume $gL = 0.00135$ cm^2/W and a seed intensity of $I_B(L) = 0.000001$ kW/cm^2.

As we indicated, Brillouin scattering involves interaction with an acoustic wave in the nonlinear medium. The scattering process must conserve both energy and momentum, so the frequencies and wavevectors of the three waves are related by

$$\begin{aligned} \Omega_A &= \omega_L - \omega_B \\ K_A &= k_L - (-k_B) \end{aligned} \tag{8.7-17}$$

where we note that the Brillouin wave is propagating backward, ω_L and ω_B are frequencies, and k_L and k_B are the wavenumbers of the laser and Brillouin waves, respectively. An electrostrictive force $\propto \chi_B^{(3)} E_L E_B^*$ is created by the simultaneous presence of the laser beam and the Brillouin wave. This force drives the acoustic wave at frequency ($\Omega_A = \omega_L - \omega_B$). The acoustic wave modulates the dielectric constant ε and creates a sideband (the Brillouin wave) at $\omega_L - (\omega_L - \omega_B) = \omega_B$. A positive feedback loop is created as the electrostrictive force increases the amplitude of the acoustic wave, which in turn increases the amplitude of the Brillouin wave. The frequency Ω_A and wavevector K_A of the acoustic wave satisfy the dispersion relation

$$\Omega_A = v_A K_A \tag{8.7-18}$$

where v_A is the velocity of acoustic wave in the medium. Since k_L and k_B are almost equal, the Brillouin frequency shift can be written

$$\Omega_A = 2 v_A k_L = \frac{4\pi n}{\lambda_L} v_A \tag{8.7-19}$$

or equivalently,

$$f_A = \frac{2n}{\lambda_L} v_A \quad \text{(in Hz)} \tag{8.7-20}$$

Table 8.7 provides a list of Brillouin properties for selected materials, including gain, frequency shift, and bandwidth.

EXAMPLE: SBS GAIN IN SiO_2

According to Table 8.7, the Brillouin gain coefficient is 0.0045 cm/MW. Assuming 10 mW of power in a core of 100 μm^2, the beam intensity is 10^8 W/m^2. The Brillouin gain is thus given by $g_B I_L = 0.0045$ cm/MW $\times$ 10^8 W/m^2 $= 4.5 \times 10^{-3}$ m^{-1}.

TABLE 8.7 Brillouin Gain, Frequency Shift, and Gain Bandwidth for Selected Materials

Material	Acoustic Velocity (km/s)	Frequency Shift (GHz)	Bandwidth (MHz)	Gain (cm/MW)
CS_2		5.85	52.3	0.13
Acetone		4.6	224	0.02
CCl_4		4.39	520	0.006
Benzene		6.47	289	0.018
Water	1.5	5.69	317	0.0048
Optical glasses		11–16	10–106	0.004–0.025
SiO_2		23	26	0.0045
Silica	5.95	11.2		
GaAs	5.15	23.3		
$LiNbO_3$	6.57	18.7		
$LiTaO_3$	6.19	17.6		
α-Al_2O_3	11.15	25.3		

8.8 FOUR-WAVE MIXING AND PHASE CONJUGATION

We have so far discussed several special cases of third-order nonlinear optical phenomena, including the Kerr effect, SRS, and SBS. Generally, a third-order nonlinear process involves the participation of four optical waves. Such a general third-order process can be viewed as the generation of a fourth optical wave with an input of three optical waves. Let the frequencies and wavevectors of the three input waves be $(\omega_1, \mathbf{k}_1)$, $(\omega_2, \mathbf{k}_2)$, and $(\omega_3, \mathbf{k}_3)$; then the frequencies and wavevectors of the fourth optical wave $(\omega_4, \mathbf{k}_4)$ can be written

$$\omega_4 = \omega_1 \pm \omega_2 \pm \omega_3 \tag{8.8-1}$$

$$\mathbf{k}_4 = \mathbf{k}_1 \pm \mathbf{k}_2 \pm \mathbf{k}_3 \tag{8.8-2}$$

These two equations can be viewed as the conservation of energy and momentum. Since the wavevector $(\mathbf{k}_4)$ is related to the frequency (ω_4) by the dispersion relationship, these two equations may not be satisfied simultaneously. Thus nonlinear optical four-wave mixing occurs only in some particular input configurations when the above two equations are satisfied simultaneously. In this section, we will discuss a very interesting and useful scheme of nonlinear optical four-wave mixings in which the phase-matching condition is satisfied. In particular, we will be discussing the degenerate (or near-degenerate) four-wave mixing process, which has important applications, including phase conjugation and spectral reversal.

We now consider a special case of four-wave mixing in which three input optical waves with frequencies ω_1, ω_2, and ω_3 mix and produce a fourth optical wave at $\omega_1 + \omega_2 - \omega_3$. To be specific, we assume that a nonlinear optical medium is irradiated simultaneously by three input optical fields:

$$\begin{aligned}
\mathbf{E}_1(\mathbf{r}, t) &= \tfrac{1}{2}\mathbf{A}_1(\mathbf{r})e^{i(\omega_1 t - \mathbf{k}_1\cdot\mathbf{r})} + \text{c.c.} \\
\mathbf{E}_2(\mathbf{r}, t) &= \tfrac{1}{2}\mathbf{A}_2(\mathbf{r})e^{i(\omega_2 t - \mathbf{k}_2\cdot\mathbf{r})} + \text{c.c.} \\
\mathbf{E}_3(\mathbf{r}, t) &= \tfrac{1}{2}\mathbf{A}_3(\mathbf{r})e^{i(\omega_3 t - \mathbf{k}_3\cdot\mathbf{r})} + \text{c.c.}
\end{aligned} \tag{8.8-3}$$

where $\mathbf{A}_1$, $\mathbf{A}_2$, and $\mathbf{A}_3$ are amplitudes of the waves. According to the third-order term in Equation (8.1-1), a nonlinear (NL) polarization oscillating at $\omega_1 + \omega_2 - \omega_3$ is generated:

$$\begin{aligned}
P_i^{(\text{NL})}(\mathbf{r}, t) &= 3\chi_{ijkl}^{(3)} A_{1j} A_{2k} A_{3l}^*(\mathbf{r}) e^{i[(\omega_1+\omega_2-\omega_3)t-(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k}_3)\cdot\mathbf{r}]} + \text{c.c.} \\
&= \text{Re}[6\chi_{ijkl}^{(3)} A_{1j} A_{2k} A_{3l}^*(\mathbf{r}) e^{i[(\omega_1+\omega_2-\omega_3)t-(\mathbf{k}_1+\mathbf{k}_2-\mathbf{k}_3)\cdot\mathbf{r}]}]
\end{aligned} \tag{8.8-4}$$

where i, j, k, l refer to Cartesian coordinates, and summations over repeated indices are assumed. $\chi_{ijkl}^{(3)}$ is a fourth-rank tensor characteristic of the medium [18] that depends on the input frequencies ω_1, ω_2, ω_3. If we apply to (8.8-4) the argument leading to Equation (8.2-8), we can convince ourselves that, unlike the phenomenon of second-harmonic generation (SHG), the third-order optical effects considered here exist in all media, including centrosymmetric crystals and isotropic media. The form of χ_{ijkl} but not its magnitude is determined by the symmetry properties of the media that are discussed and tabulated in References [18, 19]. The nonlinear polarization is actually a volume distribution of oscillating dipoles, which can radiate a new wave at the frequency $\omega_4 = \omega_1 + \omega_2 - \omega_3$, provided the wavevector of the new wave satisfies the momentum conservation condition $\mathbf{k}_4 = \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3$.

Here we consider a special case of interest. In this case, all frequencies of the three input waves are identical. In other words, $\omega_1 = \omega_2 = \omega_3 \equiv \omega_0$ and $\omega_4 = \omega_0$. Assuming the medium

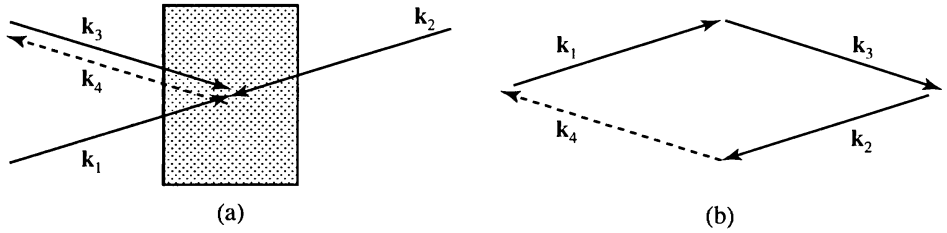


Figure 8.13 (a) Schematic drawing of degenerate four-wave mixing in an isotropic nonlinear medium. (b) Vector diagram of the conservation of momentum.

is optically isotropic, then all four wavevectors are of equal length. Thus conservation of momentum requires that the wavevectors form an equilateral parallelogram (see Figure 8.13b). In a typical degenerate four-wave mixing (shown in Figure 8.13a), two of the input beams ($\mathbf{k}_1$, $\mathbf{k}_2$) are counterpropagating, so that $\mathbf{k}_1 + \mathbf{k}_2 = 0$. A third input wave arrives at the interaction region. The fourth wave is emitted in a direction opposite to the third one. If we view the third wave as a signal wave, then the newly generated wave (the fourth wave) is always counterpropagating with respect to the signal wave. In other words, the generated wave (ω_4, $\mathbf{k}_4$) is retracing the path of the third input wave (ω_3, $\mathbf{k}_3$). The retroreflection nature of the generated wave relative to the third input wave is responsible for the phenomenon of phase conjugation.

A list of the nonlinear coefficients of some optical materials is included in Table 8.8. We also include in the table a listing of the Kerr constant n_2 [18]. This constant, as discussed in Section 8.6, describes the dependence of the index of refraction of an isotropic medium on the optical intensity I according to

$$n = n_0 + n_2 I \tag{8.8-5}$$

TABLE 8.8 Nonlinear Optical Coefficient $\chi^{(3)}_{ijkl}(-\omega, \omega, \omega, -\omega)$ of Selected Materials

Material	λ (μm)	Index	$\chi^{(3)}_{ijkl}$ (esu)	$\chi^{(3)}_{ijkl}$ (MKS)	n_2 (m^2/W)
CS_2	1.06	1.594	$\chi_{1111} = 4.65 \times 10^{-13}$	$\chi_{1111} = 5.7 \times 10^{-32}$	2.9×10^{-18}
	0.694	1.612	$\chi_{1221} = 3.6 \times 10^{-13}$	$\chi_{1221} = 4.4 \times 10^{-32}$	
YAG	1.06	1.83	$\chi_{1111} = 0.15 \times 10^{-13}$	$\chi_{1111} = 19 \times 10^{-34}$	7.2×10^{-20}
	0.694	1.829	$\chi_{1221} = 0.06 \times 10^{-13}$	$\chi_{1221} = 7.4 \times 10^{-34}$	
CCl_4	0.694	1.454	$\chi_{1221} = 0.05 \times 10^{-13}$	$\chi_{1221} = 6.2 \times 10^{-34}$	
Fused silica	0.694	1.455	$\chi_{1111} = 4 \times 10^{-15}$	$\chi_{1111} = 5 \times 10^{-34}$	3×10^{-20}
			$\chi_{1221} = 1.2 \times 10^{-15}$	$\chi_{1221} = 1.5 \times 10^{-34}$	
Ruby	1.06	1.76	$\chi_{1111} = 6.9 \times 10^{-15}$	$\chi_{1111} = 8.5 \times 10^{-34}$	3.5×10^{-20}
SF_6 glass	1.06	1.77	$\chi_{1111} = 4.2 \times 10^{-14}$	$\chi_{1111} = 5.2 \times 10^{-33}$	2.1×10^{-19}
Lucite	1.06	1.49	$\chi_{1111} = 1.1 \times 10^{-14}$	$\chi_{1111} = 1.3 \times 10^{-33}$	7.7×10^{-20}
2-Methyl-4-nitroaniline (MNA)		1.8	$\chi_{1111} = 1.19 \times 10^{-11}$	$\chi_{1111} = 1.5 \times 10^{-30}$	5.8×10^{-17}
PTS polydiacetylene		1.88	$\chi_{1111} = 4.0 \times 10^{-11}$	$\chi_{1111} = 4.9 \times 10^{-30}$	1.8×10^{-16}

From Reference [18].

It follows directly from Equation (8.6-1) that (see also Equation (8.6-7))

$$n_2 = \frac{3}{n_0^2 \varepsilon_0}\sqrt{\frac{\mu}{\varepsilon_0}}\chi_{1111}(\text{MKS}) = \frac{3}{cn_0^2\varepsilon_0^2}\chi_{1111}(\text{MKS}) \tag{8.8-6}$$

where the last equality holds for nonmagnetic materials. The relatively large value of χ_{1111} in MNA is due to a large charge separation (on the order of 30 Å) and hence large induced dipoles that can obtain in certain organic molecules.

Many of the published results on third-order nonlinear coefficients are expressed in units of esu. Based on the definition, these two units can be converted as follows:

$$\chi^{(3)}_{ijkl}(\text{MKS}) = \frac{4\pi\varepsilon_0}{9\times 10^8}\chi^{(3)}_{ijkl}(\text{esu}) \quad \text{(based on } P_i = \chi^{(3)}_{ijkl}E_jE_kE_l\text{, in this book)} \tag{8.8-7}$$

Some authors define $P_i = \varepsilon_0\chi^{(3)}_{ijkl}E_jE_kE_l$. In this case, the nonlinear coefficients $\chi^{(3)}_{ijkl}$ are in units of m^2/V^2 and the conversion formula is as follows:

$$\varepsilon_0\chi^{(3)}_{ijkl}(m^2/V^2) = \frac{4\pi\varepsilon_0}{9\times 10^8}\chi^{(3)}_{ijkl}(\text{esu}) \quad \text{(based on the definition of } P_i = \varepsilon_0\chi^{(3)}_{ijkl}E_jE_kE_l) \tag{8.8-8}$$

Coupled-Mode Formulation of Degenerate Four-Wave Mixing for Phase Conjugation

Referring to Figure 8.14, we consider the four-wave mixing process in a nonlinear medium between $z = 0$ and $z = L$. We will find that the result is a new wave generated in and radiated by the nonlinear medium. For traditional reasons, waves 1 and 2 are called pump beams, and wave 4 is an input beam. The new wave generated by the nonlinear polarization is labeled as wave 3. Let the electric fields of the waves be written

$$\begin{aligned}
\mathbf{E}_1(\mathbf{r}, t) &= \tfrac{1}{2}\mathbf{A}_1(\mathbf{r})e^{i(\omega t - \mathbf{k}_1\cdot\mathbf{r})} + \text{c.c.}\\
\mathbf{E}_2(\mathbf{r}, t) &= \tfrac{1}{2}\mathbf{A}_2(\mathbf{r})e^{i(\omega t - \mathbf{k}_2\cdot\mathbf{r})} + \text{c.c.}\\
\mathbf{E}_3(\mathbf{r}, t) &= \tfrac{1}{2}\mathbf{A}_3(z)e^{i(\omega t + kz)} + \text{c.c.}\\
\mathbf{E}_4(\mathbf{r}, t) &= \tfrac{1}{2}\mathbf{A}_4(z)e^{i(\omega t - kz)} + \text{c.c.}\\
k^2 &\equiv \omega^2\mu\varepsilon
\end{aligned} \tag{8.8-9}$$

Waves 1 and 2 propagate along the directions $\mathbf{k}_1$ and $\mathbf{k}_2$, respectively. In the analysis that follows, these two waves correspond to the pump beams and their amplitudes $|A_1|$ and $|A_2|$

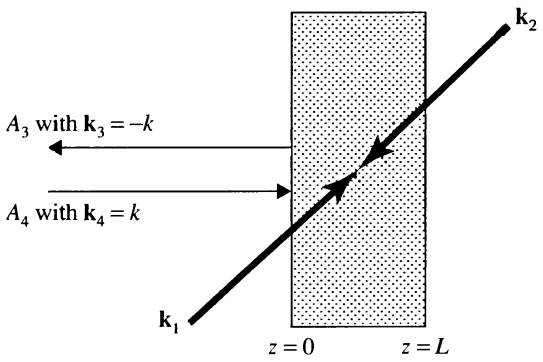


Figure 8.14 The "canonical" geometry of phase conjugation by four-wave mixing.

will be taken as much larger than $|A_3|$ and $|A_4|$ and thus will be scarcely affected by the interaction so that they will be taken as constants throughout the interaction volume. We will further take $|A_1| = |A_2|$. This causes the effect of each one of these two waves on the phase velocity of the other to be the same so that $k_2 = k_1$ [20]. Furthermore, the pump waves 1 and 2 are made to propagate through the nonlinear medium in opposite directions so that

$$\mathbf{k}_1 + \mathbf{k}_2 = 0 \tag{8.8-10}$$

As we will see later, this condition is essential for the generation of phase-conjugated waves.

Wave 4 corresponds to the input beam and in what follows wave 3, which is generated by the interaction of beams 1, 2, and 4, is the (desired) phase-conjugate replica of wave 4.

We now examine the nonlinear polarizations that are relevant to the interaction between waves 3 and 4. The third-order nonlinear polarization is given by Equation (8.6-1), where E is the sum of the electric fields of the four waves (8.8-9). There are a total of 512 terms. The choice of the relevant terms in $P_i^{(\mathrm{NL})}$ requires some judicious reasoning. First, since we assumed that the amplitudes A_3 and A_4 are not time dependent, $P_i^{(\mathrm{NL})}$ must contain the exponential time factor $\exp(i\omega t)$ to radiate efficiently into these two waves. Second, in order that $A_3(z)$ and $A_4(z)$ not vary significantly on the scale of one optical wavelength (i.e., "slow" variation), $P_i^{(\mathrm{NL})}$ must also include the factors $\exp(-ikz)$ and $\exp(+ikz)$ so that we must look for nonlinear polarization terms that contain the wave factors $\exp[i(\omega t - kz)]$ and $\exp[i(\omega t + kz)]$.

Recalling that $\mathbf{k}_1 + \mathbf{k}_2 = 0$, it follows from Equation (8.6-1) that the only third-order products of the fields $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$, and $\mathbf{E}_4$ that contain the factors $\exp[\pm i(\omega t - kz)]$ and $\exp[\pm i(\omega t + kz)]$ are (see Problem 8.12)

$$\begin{aligned} P_i^{(\mathrm{NL})}(\mathbf{r}, t) = {} & \tfrac{1}{2}\chi_{ijkl}^{(3)} \times \{6A_{1j}A_{2k}A_{3l}^* + 6A_{1j}A_{1k}^*A_{4l} + 6A_{2j}A_{2k}^*A_{4l} + 6A_{3j}A_{3k}^*A_{4l} \\ & + 3A_{4j}A_{4k}^*A_{4l}\} \exp[i(\omega t - kz)] + \tfrac{1}{2}\chi_{ijkl}^{(3)} \times \{6A_{1j}A_{2k}A_{4l}^* + 6A_{1j}A_{1k}^*A_{3l} + 6A_{2j}A_{2k}^*A_{3l} \\ & + 6A_{4j}A_{4k}^*A_{3l} + 3A_{3j}A_{3k}^*A_{3l}\} \exp[i(\omega t + kz)] + \text{c.c.} \end{aligned} \tag{8.8-11}$$

where summations over repeated indices are assumed, and we have used the cyclic symmetry of the third-order nonlinear coefficients,

$$\chi_{ijkl}^{(3)} = \chi_{[ijkl]}^{(3)} \tag{8.8-12}$$

with $[ijkl]$ being any permutation of $(ijkl)$.

For the purpose of clarity in introducing the concept, we will consider a simple case of importance. This simple case allows us to drop the tensorial notation and limit ourselves to cases where a single χ_{ijkl} (χ_{1111}) is involved. In this case, we assume that all waves are polarized as s-waves (perpendicular to the plane of incidence). By taking 1 as the axis perpendicular to the plane of incidence, the nonlinear polarization can be written

$$\begin{aligned} P_i^{(\mathrm{NL})}(\mathbf{r}, t) = {} & \tfrac{1}{2}\chi_{1111}^{(3)} \times \{6A_1A_2A_3^* + 6A_1A_1^*A_4 + 6A_2A_2^*A_4 + 6A_3A_3^*A_4 + 3A_4A_4^*A_4\} \exp[i(\omega t - kz)] \\ & + \tfrac{1}{2}\chi_{1111}^{(3)} \times \{6A_1A_2A_4^* + 6A_1A_1^*A_3 + 6A_2A_2^*A_3 + 6A_4A_4^*A_3 + 3A_3A_3^*A_3\} \exp[i(\omega t + kz)] + \text{c.c} \end{aligned} \tag{8.8-13}$$

Furthermore, we will neglect the small terms on the right side of Equation (8.8-13). This leads to

$$\begin{aligned} P_i^{(\mathrm{NL})}(\mathbf{r}, t) = {} & 3\chi_{1111}^{(3)} \times \{A_1A_2A_3^* + A_1A_1^*A_4 + A_2A_2^*A_4\} \exp[i(\omega t - kz)] \\ & + 3\chi_{1111}^{(3)} \times \{A_1A_2A_4^* + A_1A_1^*A_3 + A_2A_2^*A_3\} \exp[i(\omega t + kz)] + \text{c.c.} \end{aligned} \tag{8.8-14}$$

since $|A_3|, |A_4| \ll |A_1|, |A_2|$. Using Equation (8.8-14) in the wave equation (8.3-2) results in

$$\frac{dA_4}{dz} = -i\frac{\omega}{2}\sqrt{\frac{\mu}{\varepsilon}}\,\chi^{(3)}(|A_1|^2 + |A_2|^2)A_4 - i\frac{\omega}{2}\sqrt{\frac{\mu}{\varepsilon}}\,\chi^{(3)}A_1A_2A_3^* \tag{8.8-15}$$

$$\frac{dA_3}{dz} = +i\frac{\omega}{2}\sqrt{\frac{\mu}{\varepsilon}}\,\chi^{(3)}(|A_1|^2 + |A_2|^2)A_3 + i\frac{\omega}{2}\sqrt{\frac{\mu}{\varepsilon}}\,\chi^{(3)}A_1A_2A_4^* \tag{8.8-16}$$

where $\chi^{(3)} = 6\chi_{1111}$.

We note that the first term on the right side of Equations (8.8-15) and (8.8-16), acting alone, merely modifies the propagation constant of waves 3 and 4 from k to $k + (\omega/2)\sqrt{\mu/\varepsilon}\,\chi^{(3)}(|A_1|^2 + |A_2|^2)$. This is the optical Kerr effect discussed earlier. Using the modified wavenumbers, we can simplify the analysis by removing the first term on the right side of Equations (8.8-15) and (8.8-16). This leads to the following coupled equations:

$$\frac{dA_4}{dz} = -i\frac{\omega}{2}\sqrt{\frac{\mu}{\varepsilon}}\,\chi^{(3)}A_1A_2A_3^* \tag{8.8-17}$$

$$\frac{dA_3^*}{dz} = -i\frac{\omega}{2}\sqrt{\frac{\mu}{\varepsilon}}\,\chi^{(3)*}A_1^*A_2^*A_4 \tag{8.8-18}$$

Defining

$$\kappa^* \equiv \frac{\omega}{2}\sqrt{\frac{\mu}{\varepsilon}}\,\chi^{(3)}A_1A_2 \tag{8.8-19}$$

and taking the complex conjugate of Equations (8.8-17) and (8.8-18) results in our final form of the coupled-mode equations for phase-conjugate optics [21]:

$$\begin{aligned}\frac{dA_4^*}{dz} &= i\kappa A_3\\ \frac{dA_3}{dz} &= i\kappa^* A_4^*\end{aligned} \tag{8.8-20}$$

The student is urged to ponder at this point how a relatively complex physical experiment involving four optical beams interacting through the nonlinear optical response of a material medium can be described by equations as simple as (8.8-20). This is possible through a "ruthless," but justifiable, elimination of mathematical terms whose effects are physically negligible but whose inclusion will have rendered the analysis intractable. This is an *essential* difference between mathematics and physics.

Since wave 4 propagates in the $+z$ direction, while wave 3 propagates in the $-z$ direction, we can specify their complex amplitudes at their respective input planes $z = 0$ and $z = L$ (see Figure 8.14). These are taken as $A_4(0)$ and $A_3(L)$. Subject to these boundary conditions, the solutions of (8.8-20) are

$$\begin{aligned}A_3(z) &= \frac{\cos|\kappa|z}{\cos|\kappa|L}A_3(L) + i\frac{\kappa^*\sin|\kappa|(z-L)}{|\kappa|\cos|\kappa|L}A_4^*(0)\\ A_4(z) &= -i\frac{|\kappa|\sin|\kappa|z}{\kappa\cos|\kappa|L}A_3^*(L) + \frac{\cos|\kappa|(z-L)}{\cos|\kappa|L}A_4(0)\end{aligned} \tag{8.8-21}$$

In the basic phase-conjugate experiments, there is but a single input, $A_4(0)$ (the "pump" beams A_1 and A_2 are considered here as part of the apparatus and are lumped in our analysis into the coupling constant κ). Putting $A_3(L) = 0$, we obtain, from (8.8-21) for the reflected wave at the input,

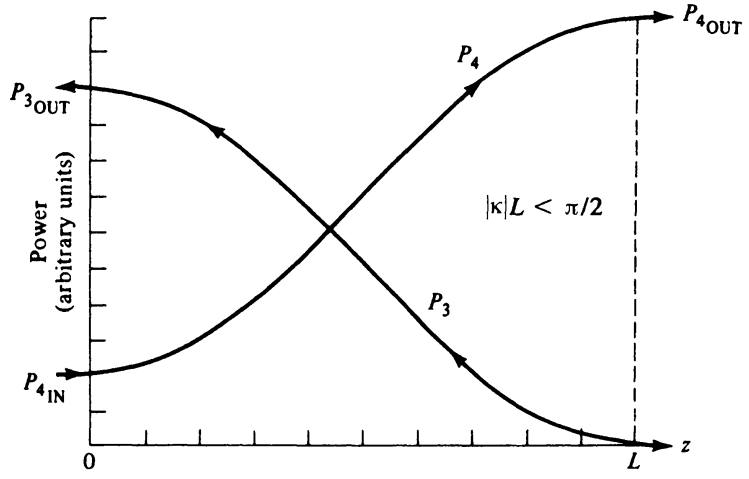


Figure 8.15 Intensity distribution inside the interaction region corresponding to the amplifier case $\pi/4 < |\kappa|L < \pi/2$.

$$A_3(0) = -i\left(\frac{\kappa^*}{|\kappa|}\tan|\kappa|L\right)A_4^*(0) \tag{8.8-22}$$

while at the output ($z = L$) we have

$$A_4(L) = \frac{A_4(0)}{\cos|\kappa|L} \tag{8.8-23}$$

We note that $|A_4(L)| > |A_4(0)|$; that is, the device acts as a phase *coherent optical amplifier* with a gain of $|(\cos|\kappa|L)^{-1}|$. When

$$\frac{\pi}{4} \le |\kappa|L \le \frac{3\pi}{4} \tag{8.8-24}$$

the result $|A_3(0)| > |A_4(0)|$ obtains, so that *the reflectivity of the phase conjugate mirror exceeds unity*. The intensity distribution of the two waves inside the nonlinear medium for a value of $|\kappa|L$ satisfying Equation (8.8-24) is shown in Figure 8.15.

Of particular interest is the condition $|\kappa|L = \pi/2$. In this case,

$$\frac{A_3(0)}{A_4(0)} = \infty \quad \frac{A_4(L)}{A_4(0)} = \infty \tag{8.8-25}$$

that is, both the transmission gain $[A_4(L)/A_4(0)]$ and the reflection gain $[A_3(0)/A_4(0)]$ become infinite so that finite outputs $A_3(0)$ and $A_4(L)$ can result even when the input $A_4(0)$ is zero. This corresponds to *oscillation*. This oscillation takes place *without the benefit of mirror feedback*. The feedback process that is essential to oscillation is provided by the fact that waves 3 and 4 propagate in opposite directions, so that $A_4(z_1)$, for example, is influenced by $A_4(z_2)$ even when $z_2 > z_1$, the information being carried from z_2 to z_1 by the backward-going wave 3. The intensity distribution corresponding to the oscillation condition is shown in Figure 8.16.

Another point of physical interest is that of the source of the power. Since energy is conserved, it follows that the increase in the output powers of beams 3 and 4 relative to their input values must come at the expense of the pump beams 1 and 2. A more exact analysis that does not neglect the spatial dependence of beams 1 and 2 shows that this indeed is the case. A quantum mechanical description of this process [20] shows that on the atomic scale the basic process is one where, simultaneously, two photons, one from beam 1 and one from beam 2,

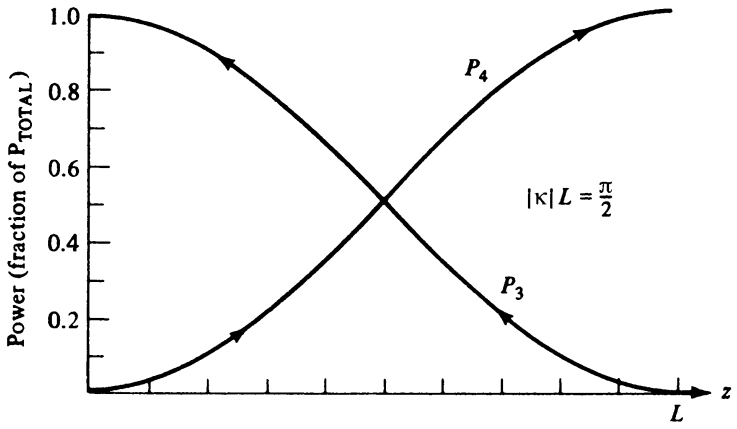


Figure 8.16 Intensity distribution inside the interaction region when the oscillation condition $|\kappa|L = \pi/2$ is satisfied.

are annihilated while two photons are created—one of these photons is added to beam 3 and the other to beam 4.

We have strayed somewhat from our main purpose, which is to show that the four-wave mixing geometry of Figure 8.14 is capable of rendering in real time the phase-conjugate (reflected) replica of an input beam. Returning to our basic plane wave result (8.8-22),

$$A_3(0) = -i\left(\frac{\kappa^*}{|\kappa|}\tan|\kappa|L\right)A_4^*(0) \tag{8.8-26}$$

It follows that the amplitude A_3 of the backward wave 3 at the input plane $z = 0$ is proportional to the complex amplitude $A_4^*(0)$ of the input wave 4 at the same plane. It follows directly that since an input wave with an arbitrarily complex wavefront $A_4(x, y, z)$ can be expanded in terms of plane wave components (which in the paraxial limit adopted here will span a small solid angle centered about the $+z$ axis), we can extend Equation (8.8-26) to each plane wave component individually [21] to obtain

$$A_3(x, y, z < 0) = -i\left(\frac{\kappa^*}{|\kappa|}\tan|\kappa|L\right)A_4^*(x, y, z < 0) \tag{8.8-27}$$

This is the basic result of phase conjugation by four-wave mixing. It shows that the reflected beam $A_3(\mathbf{r})$ to the left of the nonlinear medium ($z < 0$) is the phase conjugate of the input beam $A_4(\mathbf{r})$ (see Figure 8.17).

As in the rest of this book, our analysis of phase conjugation employs the MKS system of units. Much of the research literature, unfortunately, uses the esu system. The relations described in Equations (8.8-7) and (8.8-8) earlier in this section should facilitate the translation from one system to another.

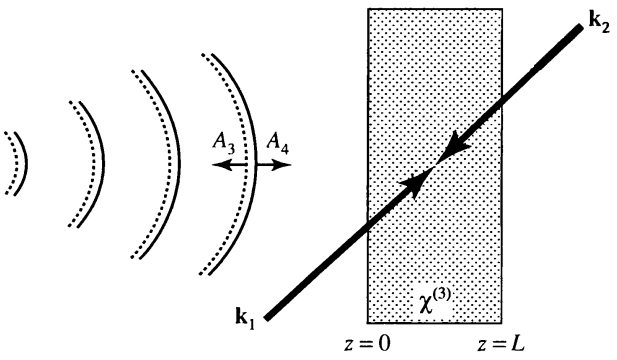


Figure 8.17 Optical phase conjugation by four-wave mixing. Solid curves represent wavefronts of incident wave A_4, and dotted curves are for the phase-conjugated wave A_3.

EXAMPLE: PHASE CONJUGATION IN CS_2

Many of the first experiments in phase-conjugate optics were performed in carbon disulfide, CS_2. In the case when all are polarized along the same direction that we will call x, the relevant nonlinear coefficient is χ_{1111}, which in MKS units has the value (Table 8.8)

$$\chi_{1111} = 5.7 \times 10^{-32}$$

The coefficient $\chi^{(3)}$ used in our analysis, we recall, is according to Equation (8.8-16) equal to $6\chi_{1111}$. We will further assume that the experiment is carried out with waves at $\lambda = 1.06\ \mu m$ and that the two pump beams are of equal intensities with $I_1 = I_2 = 5 \times 10^{10}\ W/m^2$ (i.e., $5 \times 10^6\ W/cm^2$). The index of refraction of CS_2 is $n = 1.594$. Using these data, we obtain

$$A_1 = A_2 = \sqrt{\frac{2I_1}{\varepsilon_0 cn}} = 4.86 \times 10^6\ V/m$$

and from Equation (8.8-22), $|\kappa| = 1.1\ m^{-1}$.

It follows that in CS_2 we need to use pump beams with intensities on the order of megawatts per square centimeter with path lengths on the order of 1 m in order to satisfy the condition $|\kappa|L \sim 1$ needed according to Equation (8.8-22) for appreciable (~1) phase-conjugate reflectivities.

8.9 FREQUENCY TUNING IN PARAMETRIC OSCILLATION

We have shown earlier in this chapter that the pair of signal (ω_1) and idler (ω_2) frequencies that are caused to oscillate by parametric pumping at ω_3 satisfy the condition $k_3 = k_1 + k_2$. Using $k_i = \omega_i n_i / c$, we can write it as

$$\omega_3 n_3 = \omega_1 n_1 + \omega_2 n_2 \tag{8.9-1}$$

where $n_1 = n(\omega_1)$, $n_2 = n(\omega_2)$, and $n_3 = n(\omega_3)$ are the indices of refraction. This is the phase-matching condition. In a crystal the indices of refraction generally depend, as shown in Section 8.3, on the frequency, crystal orientation (if the wave is extraordinary), electric field (in electro-optic crystals), and temperature. If, as an example, we change the crystal orientation in the oscillator shown in Figure 8.6, the oscillation frequencies ω_1 and ω_2 will change so as to compensate for the change in indices, and thus condition (8.9-1) will be satisfied at the new frequencies. To be specific, we consider the case of a parametric oscillator pumped by an extraordinary beam at a fixed frequency ω_3. The signal (ω_1) and the idler (ω_2) are ordinary waves. At some crystal orientation θ_0 (angle between the c axis and the direction of propagation), the oscillation takes place at frequencies ω_{10} and ω_{20}. Let the indices of refraction at ω_{10}, ω_{20}, and ω_3 under those conditions be n_{10}, n_{20}, and n_{30}, respectively. We want to find the change in ω_1 and ω_2 due to a small change $\Delta\theta$ in the crystal orientation.

From Equation (8.9-1) we have, at $\theta = \theta_0$,

$$\omega_3 n_{30} = \omega_{10} n_{10} + \omega_{20} n_{20} \tag{8.9-2}$$

After the crystal orientation has been changed from θ_0 to $\theta_0 + \Delta\theta$, the following changes occur:

$$n_{30} \rightarrow n_{30} + \Delta n_3$$
$$n_{10} \rightarrow n_{10} + \Delta n_1$$
$$n_{20} \rightarrow n_{20} + \Delta n_2$$
$$\omega_{10} \rightarrow \omega_{10} + \Delta\omega_1$$

Since $\omega_1 + \omega_2 = \omega_3 = \text{constant}$,

$$\omega_{20} \rightarrow \omega_{20} + \Delta\omega_2 = \omega_{20} - \Delta\omega_1$$

that is, $\Delta\omega_2 = -\Delta\omega_2$. Furthermore, since Equation (8.9-1) must be satisfied at $\theta_0 = \theta_0 + \Delta\theta$, we have

$$\omega_3(n_{30} + \Delta n_3) = (\omega_{10} + \Delta\omega_1)(n_{10} + \Delta n_1) + (\omega_{20} - \Delta\omega_1)(n_{20} + \Delta n_2)$$

Neglecting the second-order terms $\Delta n_1\, \Delta\omega_1$ and $\Delta n_2\, \Delta\omega_1$ and using Equation (8.9-2), we obtain

$$\Delta\omega_1 \Big|_{\substack{\omega_1 \approx \omega_{10} \\ \omega_2 \approx \omega_{20}}} = \frac{\omega_3\, \Delta n_3 - \omega_{10}\, \Delta n_1 - \omega_{20}\, \Delta n_2}{n_{10} - n_{20}} \tag{8.9-3}$$

According to our starting hypotheses the pump is an extraordinary ray; therefore, according to Equation (8.3-32), its index depends on the orientation θ, giving

$$\Delta n_3 = \frac{\partial n_3}{\partial \theta}\bigg|_{\theta_0} \Delta\theta \tag{8.9-4}$$

The signal and idler are ordinary rays, so their indices depend on the frequencies but not on the direction. It follows that

$$\begin{aligned} \Delta n_1 &= \frac{\partial n_1}{\partial \omega_1}\bigg|_{\omega_{10}} \Delta\omega_1 \\ \Delta n_2 &= \frac{\partial n_2}{\partial \omega_2}\bigg|_{\omega_{20}} \Delta\omega_2 \end{aligned} \tag{8.9-5}$$

Using the last two equations in (8.9-3) results in

$$\frac{\partial \omega_1}{\partial \theta} = \frac{\omega_3(\partial n_3/\partial\theta)}{(n_{10} - n_{20}) + [\omega_{10}(\partial n_1/\partial\omega_1) - \omega_{20}(\partial n_2/\partial\omega_2)]} \tag{8.9-6}$$

for the rate of change of the oscillation frequency with respect to the crystal orientation. Using Equation (8.3-32) and the relation $d(1/x^2) = -(2/x^3)dx$, we obtain

$$\frac{\partial n_3}{\partial \theta} = -\frac{n_3^3}{2} \sin(2\theta) \left[\left(\frac{1}{n_e^{\omega_3}} \right)^2 - \left(\frac{1}{n_o^{\omega_3}} \right)^2 \right]$$

which, when substituted in Equation (8.9-6), gives

$$\frac{\partial \omega_1}{\partial \theta} = \frac{-\frac{1}{2}\omega_3 n_{30}^3 \left[\left(\frac{1}{n_e^{\omega_3}} \right)^2 - \left(\frac{1}{n_o^{\omega_3}} \right)^2 \right] \sin(2\theta_0)}{(n_{10} - n_{20}) + \left(\omega_{10} \frac{\partial n_1}{\partial \omega_1} - \omega_{20} \frac{\partial n_2}{\partial \omega_2} \right)} \tag{8.9-7}$$

An experimental curve showing the dependence of the signal and idler frequencies on θ in $NH_4H_2PO_4$ (ADP) is shown in Figure 8.18. Also shown is a theoretical curve based on a quadratic approximation of Equation (8.9-7), which was plotted using the dispersion (i.e., n versus ω) data of ADP; see Reference [22]. Reasoning similar to that used to derive the angle-tuning expression (8.9-7) can be applied to determine the dependence of the oscillation frequency on temperature. Here we need to know the dependence of the various indices on temperature. This is discussed further in Problem 8.6. An experimental temperature-tuning curve is shown in Figure 8.19.

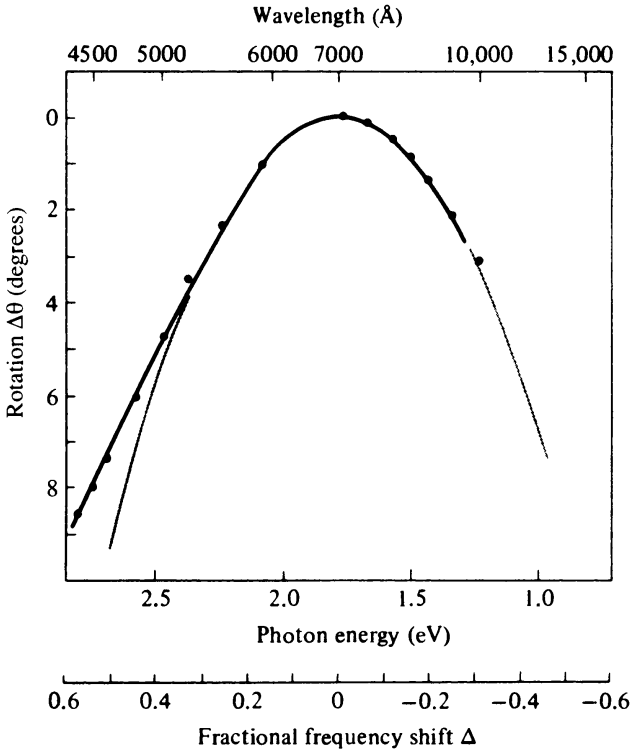


Figure 8.18 Dependence of the signal (ω_1) frequency on the angle between the pump propagation direction and the optic axis of the ADP crystal. The angle θ is measured with respect to the angle for which $\omega_1 = \omega_1/2$. $\Delta \equiv (\omega_1 - \omega_3/2)/(\omega_3/2)$. (After Reference [22].)

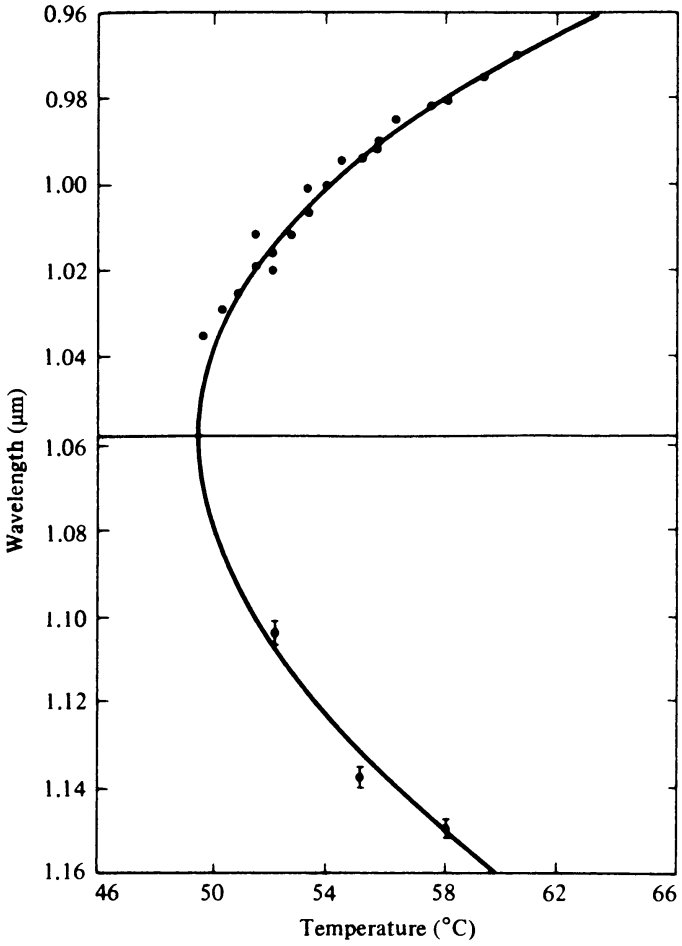


Figure 8.19 Signal and idler wavelength as a function of the temperature of the oscillator crystal. (After Reference [23].)

PROBLEMS

8.1 Show that if θ_m is the phase-matching angle for an ordinary wave at ω and an extraordinary wave at 2ω, then

$$\Delta k(\theta) l\big|_{\theta\approx\theta_m} = -\frac{2\omega l}{c_0}\sin(2\theta_m)\frac{(n_e^{2\omega})^{-2} - (n_o^{2\omega})^{-2}}{2(n_o^{\omega})^{-3}}(\theta - \theta_m)$$

8.2 Derive the expression for the phase-matching angle of a parametric amplifier using KDP, in which two of the waves are extraordinary while the third is ordinary. Which of the three waves (i.e., signal, idler, or pump) would you choose as ordinary? Can this type of phase matching be accomplished with $\omega_3 = 10{,}000\ \text{cm}^{-1}$ and $\omega_1 = \omega_2 = 5000\ \text{cm}^{-1}$? If so, what is θ_m?

8.3 Show that Equations (8.4-12) are consistent with the fact that the increases in the photon flux at ω_1 and ω_2 are identical—that is, $A_1^*(z)A_1(z) - A_1^*(0)A_1(0) = A_2^*(z)A_2(z) - A_2^*(0)A_2(0)$. Derive the same result directly from the coupled equations (8.4-10).

8.4 Complete the missing steps in the derivation of Equation (8.4-21).

8.5 Show, using the arguments of Section 8.1, that third-order optical effects as defined by Equation (8.8-4) can exist in all homogeneous media.

8.6 Consider a parametric oscillator setup such as that shown in Figure 8.19. The crystal orientation angle is θ, its temperature is T, and the signal and idler frequencies are ω_{10} and ω_{20}, respectively, with $\omega_{10} + \omega_{20} = \omega_3$. Show that a small temperature change ΔT causes the signal frequency to change by

$$\Delta\omega_1 = \Delta T \times \left\{\omega_3\left[\cos^2\theta\left(\frac{n_e^{\omega_3}(\theta)}{n_o^{\omega_3}}\right)^3\frac{\partial n_o^{\omega_3}}{\partial T} + \sin^2\theta\left(\frac{n_e^{\omega_3}(\theta)}{n_e^{\omega_3}}\right)^3\frac{\partial n_e^{\omega_3}}{\partial T}\right] - \omega_{10}\frac{\partial n_o^{\omega_1}}{\partial T} - \omega_{20}\frac{\partial n_o^{\omega_2}}{\partial T}\right\} \times \frac{1}{n_{10} - n_{20}}$$

The pump is taken as an extraordinary ray, whereas the signal and idler are ordinary. [Hint: The starting point is Equation (8.9-3), which is valid regardless of the nature of the perturbation.]

8.7 Using the published dispersion data of proustite (Reference [12]), calculate the maximum angular deviation of the input beam at ν_1 (from parallelism with the pump beam at ν_2) that results in a reduction by a factor of 2 in the up-conversion efficiency. Take $\lambda_1 = 10.6\ \mu\text{m}$, $\lambda_2 = 1.06\ \mu\text{m}$, and $\lambda_3 = 0.964\ \mu\text{m}$. [Hint: A proper choice must be made for the polarizations at ω_1, ω_2, and ω_3 so that phase matching can be achieved along some angle.] The maximum angular deviation is that for which

$$\frac{\sin^2[\Delta k(\theta) l/2]}{[\Delta k(\theta) l/2]^2} = \frac{1}{2}$$

where, at the phase-matching angle θ_m, $\Delta k(\theta_m) = 0$. Approximate the dispersion data by a Taylor-series expansion about the nominal ($\Delta k = 0$) frequencies.

8.8 Using the dispersion data of Reference [12], discuss what happens to phase matching in an up-conversion experiment due to a deviation of the input frequency from the nominal ($\Delta k = 0$) ν_{10} value. Derive an expression for the spectral width of the output in the case where the input spectral density (power per unit frequency) in the vicinity of ν_{10} is uniform. [Hint: Use a Taylor-series expansion of the dispersion data about the phase-matching ($\Delta k = 0$) frequencies to obtain an expression for $\Delta k(\nu_3)$.] Define the output spectral width as twice the frequency deviation at which the output is one-half its maximum ($\Delta k = 0$) value.

8.9 Derive the coupled-mode equations in a manner similar to that leading to (8.8-15) and (8.8-16) for the case where the frequency of the incident wave ω_4 is detuned from that of the pump beams (ω) by

$$\omega_4 = \omega - \delta$$

(a) Show that the reflected wave frequency is $\omega_3 = \omega + \delta$.
(b) Solve the coupled-mode equations for the reflection coefficient $|A_3(0)/A_4(0)|^2$. Plot it as a function of the frequency offset δ.

8.10
(a) Solve the degenerate ($\omega_1 = \omega_2 = \omega_3 = \omega_4$) coupled-mode equations (8.8-15) and (8.8-16) as modified for (ordinary) optical losses. The new equations are

$$\frac{dA_4^*}{dz} = i\kappa e^{-\alpha L}A_3 - \alpha A_4^*$$

$$\frac{dA_3}{dz} = i\kappa^* e^{-\alpha L}A_4^* + \alpha A_3$$

$$\kappa^* = \frac{\omega}{2}\sqrt{\frac{\mu}{\varepsilon}}\,\chi^{(3)}A_1(L)A_2(0)$$

where α is the optical amplitude loss coefficient (assumed the same for all four beams).

(b) Plot the reflection coefficient $|A_3(0)/A_4(0)|^2$ as a function of κL for $\alpha L = 0.1, 0.5, 1, 2$. Discuss qualitatively the effect of the losses.

8.11

(a) Derive Equations (8.3-8) and (8.3-9).

(b) Show that for phase-matched SHG in KDP, $d = d_{36} \sin \theta_m$.

8.12 Starting from the definition of third order and the presence of four waves given by Equation (8.8-9), show that the nonlinear polarization with spatial–temporal dependence of $\exp[i(\omega t - kz)]$ is given by

$$\begin{aligned}P_i^{(\mathrm{NL})}(\mathbf{r}, t) = \tfrac{1}{2}\chi_{ijkl}^{(3)} \times \{&(A_{1j}A_{2k}A_{3l}^* + A_{1j}A_{2l}A_{3k}^* + A_{1k}A_{2j}A_{3l}^* \\ &+ A_{1k}A_{2l}A_{3j}^* + A_{1l}A_{2k}A_{3j}^* + A_{1l}A_{2j}A_{3k}^*) \\ &+ (A_{1j}A_{1k}^*A_{4l} + A_{1k}A_{1j}^*A_{4l} + A_{1j}A_{1l}^*A_{4k} + A_{1l}A_{1j}^*A_{4k} \\ &+ A_{1l}A_{1k}^*A_{4j} + A_{1k}A_{1l}^*A_{4j}) + (A_{2j}A_{2k}^*A_{4l} + A_{2k}A_{2j}^*A_{4l} \\ &+ A_{2j}A_{2l}^*A_{4k} + A_{2l}A_{2j}^*A_{4k} + A_{2l}A_{2k}^*A_{4j} + A_{2k}A_{2l}^*A_{4j}) \\ &+ (A_{3j}A_{3k}^*A_{4l} + A_{3k}A_{3j}^*A_{4l} + A_{3j}A_{3l}^*A_{4k} + A_{3l}A_{3j}^*A_{4k} \\ &+ A_{3l}A_{3k}^*A_{4j} + A_{3k}A_{3l}^*A_{4j}) + (A_{4j}A_{4k}^*A_{4l} + A_{4j}A_{4l}^*A_{4k} \\ &+ A_{4l}A_{4j}^*A_{4k})\} \exp[i(\omega t - kz)]\end{aligned}$$

[Hint: Use the symmetric relation $\chi_{ijkl}^{(3)} = \chi_{[ijkl]}^{(3)}$, where $[ijkl]$ is any permutation of $(ijkl)$.] Derive a similar expression with $\exp[i(\omega t + kz)]$.

8.13 The polarization including linear and third-order electric susceptibility as well as the refractive index can be written as in Table P8.13.

TABLE P8.13

MKS	esu (CGS)
$P = \varepsilon_0 \chi E + \chi^{(3)} E^3$	$P = \chi E + \chi^{(3)} E^3$
$n^2 = 1 + \chi$	$n^2 = 1 + 4\pi\chi$

(a) Treat the contribution from third-order susceptibility as a correction to Δn^2. Show that

$$\frac{1}{\varepsilon_0}\chi_{\mathrm{MKS}}^{(3)} E_{\mathrm{MKS}}^2 = 4\pi\chi_{\mathrm{esu}}^{(3)} E_{\mathrm{esu}}^2$$

(b) $E_{\mathrm{esu}} = 1$ corresponds to $E_{\mathrm{MKS}} = 3 \times 10^4$ V/m. Show that

$$\chi_{\mathrm{MKS}}^{(3)} = \frac{4\pi\varepsilon_0}{9 \times 10^8}\chi_{\mathrm{esu}}^{(3)}$$

8.14

(a) Start from Equations (8.6-11). Show that for unpolarized light, the third terms average out to zero. Derive Equation (8.6-13).

(b) For circularly polarized light, let $E_{0x}^{\omega} = A$ and $E_{0y}^{\omega} = Ae^{i\delta}$, where A and $\delta = \pi/2$ are real numbers. Show that

$$P_{\mathrm{NL}x}^{(\omega)} = 4\chi_{xxxx}|A|^2 E_{0x}^{(\omega)} \quad \text{and} \quad P_{\mathrm{NL}y}^{(\omega)} = 4\chi_{xxxx}|A|^2 E_{0y}^{(\omega)}$$

We note that the third-order nonlinear polarization is parallel to the electric field vector. This ensures that the polarization state will remain the same. Derive Equation (8.6-14).

(c) For a general linear polarized light, $E_{0x}^{\omega} = A$ and $E_{0y}^{\omega} = B$, where A and B are real numbers. Show that the third-order nonlinear polarization is parallel to the electric field vector. This is consistent with the isotropic nature of the medium.

(d) Consider the situation of a beam of elliptically polarized light. Without loss of generality, we can choose the x axis parallel to the long axis of the polarization ellipse at input. Let $E_{0x}^{\omega} = A$ and $E_{0y}^{\omega} = Be^{i\delta}$, where A, B, and $\delta = \pi/2$ are real numbers. Show that

$$P_{\mathrm{NL}x}^{(\omega)} = \chi_{xxxx}(3A^2 + B^2)E_{0x}^{(\omega)}$$

$$P_{\mathrm{NL}y}^{(\omega)} = \chi_{xxxx}(3B^2 + A^2)E_{0y}^{(\omega)}$$

Note that the third-order nonlinear polarization is not parallel to the electric field vector. As a result, a phase retardation exists between the two components as the wave propagates in the Kerr medium.

(e) Show, with the aid of a Poincaré sphere, that the polarization ellipse will rotate while maintaining the ellipticity as the wave propagates in the Kerr medium. Show that the rate of rotation per unit distance of propagation is given by

$$\frac{d\psi}{dz} = -\frac{1}{3}\frac{2AB}{A^2 + B^2}\frac{2\pi}{\lambda}n_2 I$$

where I is the intensity of the beam. The sense of rotation is the same as the handedness of the elliptical polarization state. The rate of rotation is zero for the special case of linearly polarized light (A or $B = 0$).

8.15 Consider a nonlinear medium with $P_i = \varepsilon_0\chi_{ij}E_j + 2d_{ijk}E_jE_k$. According to Problem 1.6, the work done per unit volume by applying an electric field in a dielectric medium is $W = \int \mathbf{E} \cdot d\mathbf{P}$, where $\mathbf{P}$ is the polarization.

Carry out the integration from (0, 0) to (E_{10}, E_{20}) in a two-dimensional case by using two different paths:

Path A: First integrate from (0, 0) to $(E_{10}, 0)$, and then from $(E_{10}, 0)$ to (E_{10}, E_{20}).
Path B: First integrate from (0, 0) to $(0, E_{20})$, and then from $(0, E_{20})$ to (E_{10}, E_{20}).

(a) Show that for path A, the work done due to the nonlinear polarization is

$$W = \tfrac{4}{3}d_{111}E_{10}^3 + \tfrac{4}{3}d_{222}E_{20}^3 + 4d_{112}E_{10}^2E_{20} + 2d_{122}E_{10}E_{20}^2 + 2d_{212}E_{10}E_{20}^2$$

(b) Show that for path B, the work done due to the nonlinear polarization is

$$W = \tfrac{4}{3}d_{111}E_{10}^3 + \tfrac{4}{3}d_{222}E_{20}^3 + 2d_{112}E_{10}^2E_{20} + 2d_{211}E_{20}E_{10}^2 + 4d_{212}E_{10}E_{20}^2$$

In a lossless medium, the integration should be independent of the path. This leads to $d_{122} = d_{212}$ and $d_{121} = d_{211}$. A similar analysis in the three-dimensional case will yield $d_{ijk} = d_{jik}$. Combining the result here with Equations (8.1-2), we obtain $d_{ijk} = d_{[ijk]}$, where $[ijk]$ is any permutation of the subscript indices.

REFERENCES

1. Franken, P. A., A. E. Hill, C. W. Peters, and G. Weinreich, Generation of optical harmonics. *Phys. Rev. Lett.* **7**:118 (1961).
2. Zemike, F., Jr., Refractive indices of ammonium dihydrogen phosphate and potassium dihydrogen phosphate between 2000 Å and 1.5 μ. *J. Opt. Soc. Am.* **54**:1215 (1964).
3. Maker, P. D., R. W. Terhune, M. Nisenoff, and C. M. Savage, Effects of dispersion and focusing on the production of optical harmonics. *Phys. Rev. Lett.* **8**:21 (1962).
4. Giordmaine, J. A., Mixing of light beams in crystals. *Phys. Rev. Lett.* **8**:19 (1962).
5. Ashkin, A., G. D. Boyd, and J. M. Dziedzic, Observation of continuous second harmonic generation with gas lasers. *Phys. Rev. Lett.* **11**:14 (1963).
6. Seka, W., S. D. Jacobs, J. E. Rizzo, R. Boni, and R. S. Craxton, Demonstration of high efficiency third harmonic conversion of high power Nd–glass laser radiation. *Opt. Commun.* **34**:469 (1980). Also see R. S. Craxton, High efficiency frequency tripling schemes for high power Nd:glass lasers. *IEEE J. Quantum Electron.* **17**:177 (1981). (Additional articles on doubling and frequency conversion are to be found in the same issue.)
7. Geusic, J. E., H. J. Levinstein, S. Singh, R. G. Smith, and L. G. Van Uitert, Continuous 0.53 μm solid-state source using $Ba_2NaNb_5O_{15}$. *IEEE J. Quantum Electron.* **4**:352 (1968).
8. Thorsos, E. I. (ed.), *Laboratory for Laser Analytics Review*, Vol. II, p. 7, 1979–1980.
9. Johnson, F. M., and J. A. Durado, Frequency up-conversion. *Laser Focus* **3**:31 (1967).
10. Midwinter, J. E., and J. Warner, Up-conversion of near infrared to visible radiation in lithium–meta-niobate. *J. Appl. Phys.* **38**:519 (1967).
11. Warner, J., Photomultiplier detection of 10.6 μm radiation using optical up-conversion in proustite. *Appl. Phys. Lett.* **12**:222 (1968).
12. Hulme, K. F., O. Jones, P. H. Davies, and M. V. Hobden, Synthetic proustite (Ag_3AsS_3): a new material for optical mixing. *Appl. Phys. Lett.* **10**:133 (1967).
13. Somekh, S., and A. Yariv, Phase matching by periodic modulation of the nonlinear optical properties. *Opt. Commun.* **6**(3):301 (1972).
14. Bonz, M. L., *Quasi Phase Matched Optical Frequency Conversion in Lithium Niobate Waveguides*. Ph.D. thesis, Stanford University, 1994.
15. Jackel, J. L., C. E. Rice, and J. J. Vesseka, Proton exchange for high index waveguides in $LiNbO_3$. *Appl. Phys. Lett.* **41**:607 (1982).
16. Fiesst, A., and P. Koidl, Current induced periodic ferroelectric domain reversal in $LiNbO_3$ for efficient second harmonic generation. *Appl. Phys. Lett.* **47**:1125 (1985).

17. Yamada, M., N. Nada, M. Saito, and K. Watanabe, First order quasi phase-matched $LiNbO_3$ waveguide periodically poled by applying an external field for efficient blue second harmonic generation. *Appl. Phys. Lett.* **62**:435 (1993).
18. Hellwarth, R. W., Third order susceptibilities of liquids and gases. *Progr. Quantum Electron.* **5**:1 (1977).
19. See, for example, Y. R. Shen, *Principles of Nonlinear Optics.* Wiley, New York, 1984, Chapter 10.
20. Fisher, R. A. (ed.), *Optical Phase Conjugation.* Academic Press, New York, 1983. A comprehensive collection of material covering many aspects of phase conjugation.
21. Yariv, A., and D. M. Pepper, Amplified reflection, phase conjugation, and oscillation in degenerate four-wave mixing. *Opt. Lett.* **1**:16 (1977).
22. Magde, D., and H. Mahr, Study in ammonium dihydrogen phosphate of spontaneous parametric interaction tunable from 4400 to 16000 Å. *Phys. Rev. Lett.* **18**:905 (1967).
23. The first demonstration of optical parametric oscillation is that of J. A. Giordmaine and R. C. Miller Tunable optical parametric oscillation in $LiNbO_3$ at optical frequencies. *Phys. Rev. Lett.* **14**:973 (1965).