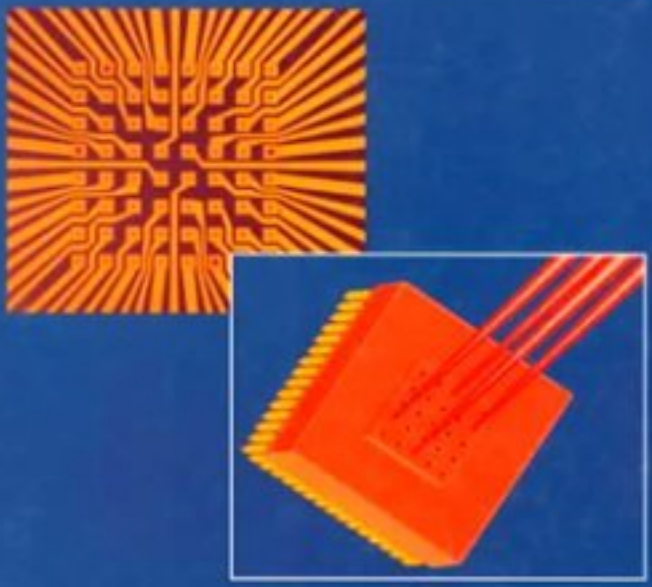
LASER ELECTRONICS

THIRD EDITION



Joseph T. Verdeyen

PRENTICE HALL SERIES IN SOLID STATE PHYSICAL ELECTRONICS
Nick Holonyak, Jr., Series Editor

Laser Electronics

THIRD EDITION

JOSEPH T. VERDEYEN

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, Urbana, Illinois

PRENTICE HALL SERIES IN SOLID STATE PHYSICAL ELECTRONICS
Nick Holonyak, Jr., Series Editor



PRENTICE HALL
Englewood Cliffs, New Jersey 07632

Library of Congress Cataloging-in-Publication Data

Verdeyen, Joseph Thomas
Laser electronics. / Joseph T. Verdeyen. – 3rd ed.
p. cm. — (Prentice Hall series in solid state physical electronics)
Includes bibliographical references and index.
ISBN 0-13-706666-X
1. Lasers. 2. Semiconductor lasers. I. Title. II. Series.
TA1675.V47 1995
621.36'61—dc20 93-2184
CIP

Acquisitions editor: ***Alan Apt***
Production editor: ***Irwin Zucker***
Copy editor: ***Michael Schwartz***
Production coordinator: ***Linda Behrens***
Supplements editor: ***Alice Dworkin***
Cover design: ***Design Solutions***
Cover illustration: ***Dr. R. P. Bryan of Photonics Research***
Editorial assistant: ***Shirley McGuire***



© 1995, 1989, 1981 by Prentice-Hall, Inc.
A Paramount Communications Company
Englewood Cliffs, New Jersey 07632

The author and publisher of this book have used their best efforts in preparing this book. These efforts include the development, research, and testing of the theories and programs to determine their effectiveness. The author and publisher make no warranty of any kind, expressed or implied, with regard to these programs or the documentation contained in this book. The author and publisher shall not be liable in any event for incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of these programs.

All rights reserved. No part of this book may be
reproduced, in any form or by any means,
without permission in writing from the publisher.

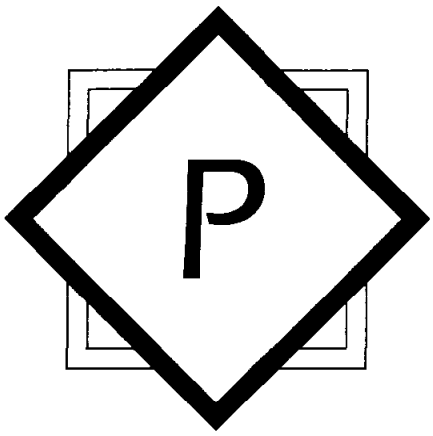
Printed in the United States of America

10 9 8 7 6 5 4 3 2

ISBN 0-13-706666-X

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

This book is dedicated to Katie,
my wife, constant companion, and best friend
for 40 years of marriage and courtship.
She is the loving mother of my children Mary, Joe, Jean, and Mike,
an exciting grandmother to their children,
and an understanding mother-in-law to Dennis, Pam, Jim, and Tammy.
She has demonstrated incredible patience and understanding
with the rather painful process of revising this book while maintaining
a most pleasant, cheerful and comforting home.
From my perspective, our marriage has had a storybook characteristic to it
with my love for her increasing daily.
With her enthusiasm, example, and love, it is easy to learn
to love God, to love our neighbors, and to keep His commandments.
Thank you honey for my life!



Preface

The underlying philosophy of this third edition of *Laser Electronics* is the same as in the previous two: lasers are very simple devices and are far simpler than the very complicated high frequency RF or microwave transistor circuits. The main purpose of the book is to convince the student of this fact. In one sense, lasers are a simple movement of the decimal point on the frequency scale three to five places to the right, but much of the terminology and all of the insight developed by the earlier pioneers of radio have been translated to the optical domain.

The potential of the many applications of lasers and optical phenomena has necessitated the formation of a new word to describe the field: *photonics*. One would be hard pressed to define all of its ramifications since new ideas, devices, and applications are frequently being added. In a very loose sort of way, the early history of radio is being repeated in the optical frequency domain, and this is a theme that will be employed throughout the book. Although both have a common basis in electromagnetic theory, there are special phenomena peculiar to the optical wavelengths.

For instance, a wave intensity of 10^{15}–10^{19} watts/m^2 would have been incomprehensible in 1960, but is now attainable with rather common lasers and comparatively cheap optics. Similarly, a 50 femtoseconds (50×10^{-15} s) pulse requires more frequency bandwidth for transmission than that which was installed in all of the telecommunications networks of 1960. Yet such a pulse is rather common with optical techniques.

The ability to generate such short pulses and transmit them over significant distances (many hundreds of kilometers) by using low loss fibers and erbium-doped fiber

amplifiers (EDFA) was a major impetus for the revisions incorporated into this third edition.

Chapter 4 has been changed to emphasize some of the more sophisticated aspects of guided wave propagation, such as dispersion in fibers, solitons, and perturbation theory. By necessity, the chapter is an introduction intended to encourage further investigation. While those are important topics for a communication system, they may be too involved for a first course in lasers. Thus, the entire chapter can be skipped if the focus of the course is on the *generation* portion of photonics.

Chapter 9 has been rewritten and reorganized to emphasize the dynamics of the laser: the approach to CW oscillation, Q switching, and various aspects of mode locking. The latter has been greatly expanded, but, even so, there are important topics not included.

Various additions have been included in Chapter 10 on specific laser systems. The example of a semiconductor laser pumping a YAG system was carried through in some detail so as to emphasize the application of the theoretical tools developed in the previous chapters and to indicate a significant application of the semiconductor laser. The *erbium doped fiber amplifier* (EDFA) is also discussed here, and a fairly long-winded simplified "problem" (with answers) is given to emphasize some of the unique considerations of the topic and to encourage further investigation of the literature. The multiplicity of levels of the EDFA serves as an introduction to gain/absorption between bands and to tunable vibronic lasers such as alexandrite, Ti:sapphire, and dye lasers.

Much of the expansion in photonics is being led by the improvements in the semiconductor laser, which has become the dominant laser for communication and control. Its use as a pump for the fiber amplifiers and solid-state lasers has also become most important. Chapter 11 has been expanded somewhat but is still intended to be an introduction to a course devoted entirely to that laser.

Most students have a fair grasp of the beauty and elegance of electromagnetic theory but have the mistaken view that the word *photon* somehow weakens its applicability. That is unfortunate. The lowest power laser generates literally billions of photons per second, and thus the classical field description of it is quite adequate. Even when the photon flux becomes small—say 10 to 100 s^{-1}, the classical field description will handle the practical cases. Many of the advances in semiconductor lasers, in particular, can be traced to classical electromagnetic theory of guidance of the modes by the heterostructures. Chapter 12 is included to introduce the student to some of the more advanced topics, possibly to be studied in a second course.

Chapters 13 and 14 are aimed at the student who wants a gradual transition to a quantum theory of the laser while the simple theory is fresh. Chapter 14 is an attempt to provide a bridge between the simple rate equation description of a laser and the more formal quantum theory using the density matrix. The two approaches agree, precisely, for the case of a CW two-level system, but the former is much easier and more akin to the student's background. The latter will handle the transient cases, scattering, two-photon phenomena, etc., at the expense of considerably more mathematics. The serious student should become aware of the transition between the two approaches, have confidence in both, and be aware of the pitfalls and limitations, again in a second course. One of the main conclusions is that

a simple rate equation of laser phenomena is quite adequate and accurate most of the time. A few cases that do not follow this rule are included.

Many more problems are included in this third edition with the primary purpose of convincing the student of the transparent simplicity of the rate equation approach. Rate equations are no more difficult than coupled circuit equations (or the differential equation describing the student's finances): There is always a source (a salary) and a loss (expenses) that may or may not be in steady state equilibrium.

ACKNOWLEDGMENTS

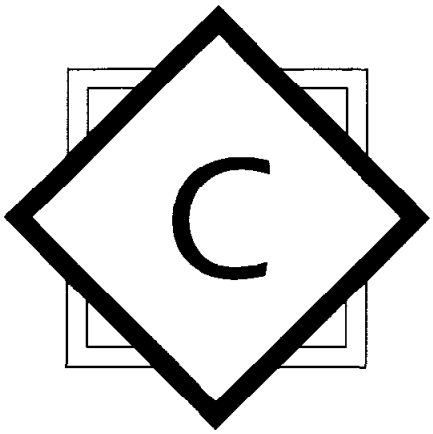
It is a pleasure to acknowledge my present and former colleagues at the University of Illinois for their help, encouragement, and many discussions of the topics included here. I am particularly grateful to: N. Holonyak, Jr., for his ability and patience in communicating his masterful insight into semiconductor electronics; to J.J. Coleman for the initial encouragement to write the book and general discussions on semiconductor materials; to T.A. DeTemple, who has been most patient and helpful with my attempt to simplify some of the topics included here; to S. Bishop for his leadership as the Director of the Microelectronics Laboratory; and to P.D. Coleman who had a significant impact on my view of electrodynamics.

I would also like to thank the reviewers: Jorge Rocca of Colorado State University, Daniel Elliott of Purdue University, Raymond Rostuk of the University of Arizona, and Sally Stevens-Tammens of the University of Illinois at Urbana-Champaign.

I especially wish to thank the many students who have helped "write" and modify this book while keeping their good humor. Their enthusiasm for photonics has really been an inspiration to me. I hope that I have taught them as well as they have educated me. I am also grateful to Ms. Galena Smirnov who patiently checked much of the new material.

I am particularly grateful to Dr. Robert Bryan of Photonics Research, Inc. for his permission to use some of the figures on the cover.

Joseph T. Verdeyen



Contents

List of Symbols xx

0 Preliminary Comments 1

Note to the students 3

References 6

1 Review of Electromagnetic Theory 8

1.1 Introduction 8

1.2 Maxwell's Equations 9

1.3 Wave Equation for Free Space 10

1.4 Algebraic Form of Maxwell's Equations 11

1.5 Waves in Dielectrics 12

1.6 The Uncertainty Relationships 13

1.7 Spreading of an Electromagnetic Beam 15

1.8 Wave Propagation in Anisotropic Media 16

1.9 Elementary Boundary Value Problems in Optics 20

1.9.1 Snell's Law, 20
1.9.2 Brewster's Angle, 21

1.10 Coherent Electromagnetic Radiation 23

1.11 Example of Coherence Effects 28

Problems 31

References and Suggested Readings 34

2 Ray Tracing in an Optical System 35

2.1 Introduction 35

2.2 Ray Matrix 35

2.3 Some Common Ray Matrices 37

2.4 Applications of Ray Tracing: Optical Cavities 39

2.5 Stability: Stability Diagram 42

2.6 The Unstable Region 44

2.7 Example of Ray Tracing in a Stable Cavity 44

2.8 Repetitive Ray Paths 47

2.9 Initial Conditions: Stable Cavities 48

2.10 Initial Conditions: Unstable Cavities 49

2.11 Astigmatism 50

2.12 Continuous Lens-Like Media 51

2.12.1 Propagation of a Ray in an Inhomogeneous Medium, 53
2.12.2 Ray Matrix for a Continuous Lens, 54

2.13 Wave Transformation by a Lens 56

Problems 57

References and Suggested Readings 62

3 Gaussian Beams 63

3.1 Introduction 63

3.2 Preliminary Ideas: TEM Waves 63

3.3 Lowest-Order $TEM_{0,0}$ Mode 66

3.4 Physical Description of $TEM_{0,0}$ Mode 70

3.4.1 Amplitude of the Field, 70
3.4.2 Longitudinal Phase Factor, 71
3.4.3 Radial Phase Factor, 72

3.5 Higher-Order Modes 73

3.6 $ABCD$ law for Gaussian beams 76

3.7 Divergence of the Higher-Order Modes: Spatial Coherence 79

Problems 80

References and Suggested Readings 84

4 Guided Optical Beams 86

4.1 Introduction 86

4.2 Optical Fibers and Heterostructures: A Slab Waveguide Model 87

4.2.1 Zig-Zag Analysis, 87
4.2.2 Numerical Aperture, 89

4.3 Modes in a Step-Index Fiber (or a Heterojunction Laser): Wave Equation Approach 90

4.3.1 TE Mode ($E_z = 0$), 92
4.3.2 TM Modes ($H_z = 0$), 94
4.3.3 Graphic Solution for the Propagation Constant: "R" and "V" Parameters, 95

4.4 Gaussian Beams in Graded Index (GRIN) Fibers and Lenses 96

4.5 Perturbation Theory 102

4.6 Dispersion and Loss in Fibers: Data 105

4.7 Pulse Propagation in Dispersive Media: Theory 109

4.8 Optical Solitons 116

Problems 122

References and Suggested Readings 127

5 Optical Cavities 130

5.1 Introduction 130

5.2 Gaussian Beams in Simple Stable Resonators 130

5.3 Application of the $ABCD$ Law to Cavities 133

5.4 Mode Volume in Stable Resonators 137

Problems 139

References and Suggested Readings 142

6 Resonant Optical Cavities 144

6.1 General Cavity Concepts 144

6.2 Resonance 144

6.3 Sharpness of Resonance: Q and Finesse 148

6.4 Photon Lifetime 151

6.5 Resonance of the Hermite-Gaussian Modes 154

6.6 Diffraction Losses 156

6.7 Cavity With Gain: An Example 157

Problems 159

References and Suggested Readings 170

7 Atomic Radiation 172

7.1 Introduction and Preliminary Ideas 172

7.2 Blackbody Radiation Theory 173

7.3 Einstein's Approach: *A* and *B* Coefficients 179

7.3.1 Definition of Radiative Processes, 179
7.3.2 Relationship Between the Coefficients, 181

7.4 Line Shape 183

7.5 Amplification by an Atomic System 187

7.6 Broadening of Spectral Lines 191

7.6.1 Homogeneous broadening mechanisms, 191
7.6.2 Inhomogeneous Broadening, 196
7.6.3 General Comments on the Line Shape, 200

7.7 Review 200

Problems 201

References and Suggested Readings 205

8 Laser Oscillation and Amplification 207

8.1 Introduction: Threshold Condition for Oscillation 207

8.2 Laser Oscillation and Amplification in a Homogeneous Broadened Transition 208

8.3 Gain Saturation in a Homogeneous Broadened Transition 212

8.4 Laser Oscillation in an Inhomogeneous System 223

8.5 Multimode Oscillation 229

8.6 Gain Saturation in Doppler-Broadened Transition: Mathematical Treatment 230

8.7 Amplified Spontaneous Emission (ASE) 234

8.8 Laser Oscillation: A Different Viewpoint 238

Problems 242

References and Suggested Readings 258

9 General Characteristics of Lasers 260

9.0 Introduction 260

9.1 Limiting Efficiency 260

9.1.1 Factors in the efficiency, 260
9.1.2 Two, 3, 4 : : :, n level lasers, 261

9.2 CW Laser 263

9.2.1 Traveling Wave Ring Laser, 264
9.2.2 Optimum Coupling, 267
9.2.3 Standing Wave Lasers, 269

9.3 Laser Dynamics 274

9.3.1 Introduction and model, 274
9.3.2 Case a: A sub-threshold system, 276
9.3.3 Case b: A CW laser: threshold conditions, 276
9.3.4 Case c: A sinusoidal modulated pump, 277
9.3.5 Case d. A sudden "step" change in excitation rate, 280
9.3.6 Case e: Pulsed excitation → gain switching, 282

9.4 Q Switching, Q Spoiling, or Giant Pulse Lasers 284

9.5 Mode Locking 296

9.5.1 Preliminary considerations, 296
9.5.2 Mode locking in an inhomogeneous broadened laser, 298
9.5.3 Active mode locking, 304

9.6 Pulse Propagation in Saturable Amplifiers or Absorbers 311

9.7 Saturable Absorber (Colliding Pulse) Mode Locking 317

9.8 Additive-Pulse Mode Locking 322

Problems 324

References and Suggested Readings 344

10 Laser Excitation **347**

10.1 Introduction 347

10.2 Three- and Four-Level Lasers 348

10.3 Ruby Lasers 351

10.4 Rare Earth Lasers and Amplifiers 358

10.4.1 General Considerations, 358
10.4.2 Nd:YAG lasers: Data, 359
10.4.3 Nd:YAG Pumped by a Semiconductor Laser, 362
10.4.4 Neodymium-Glass Lasers, 369
10.4.6 Erbium-Doped-Fiber-Amplifiers, 371

10.5 Broad-Band Optical Gain 376

10.5.1 Band-to-Band Emission and Absorption, 376
10.5.2 Theory of Band-to-Band Emission and Absorption, 377

10.6 Tunable Lasers 385

10.6.1 General Considerations, 385
10.6.2 Dye Lasers, 386
10.6.3 Tunable Solid State Lasers, 391
10.6.4 Cavities for Tunable Lasers, 395

10.7 Gaseous-Discharge Lasers 396

10.7.1 Overview, 396
10.7.2 Helium-Neon Laser, 397
10.7.3 Ion Lasers, 403
10.7.4 CO_2 Lasers, 405

10.8 Excimer Lasers: General Considerations 411

10.8.1 Formation of the Excimer State, 412
10.8.2 Excitation of the Rare Gas–Halogen Excimer Lasers, 415

10.9 Free Electron Laser 417

Problems 423

References and Suggested Readings 434

11 Semiconductor Lasers **440**

11.1 Introduction 440

11.1.1 Overview, 440
11.1.2 Populations in Semiconductor Laser, 442

11.2 Review of Elementary Semiconductor Theory 444

11.2.1 Density of States, 445

11.3 Occupation Probability: Quasi-Fermi Levels 449

11.4 Optical Absorption and Gain in a Semiconductor 450

11.4.1 Gain Coefficient in a Semiconductor, 454
11.4.2 Spontaneous Emission Profile, 459
11.4.3 An Example of an Inverted Semiconductor, 460

11.5 Diode Laser 464

11.5.1 Homojunction Laser, 464
11.5.2 Heterojunction Lasers, 467

11.6 Quantum Size Effects 470

11.6.1 Infinite Barriers, 470
11.6.2 Finite Barriers: An Example, 476

11.7 Vertical Cavity Surface Emitting Lasers 482

11.8 Modulation of Semiconductor Lasers 486

11.8.1 Static Characteristics, 488
11.8.2 Frequency Response of Diode Lasers, 489

Problems 492

References and Suggested Readings 499

12 Advanced Topics in Laser Electromagnetics 502

12.1 Introduction 502

12.2 Semiconductor Cavities 503

12.2.1 TE Modes ($E_z = 0$), 505
12.2.2 TM Modes ($H_z = 0$), 507
12.2.3 Polarization of TE and TM Modes, 508

12.3 Gain Guiding: An Example 509

12.4 Optical Confinement and Effective Index 516

12.5 Distributed Feedback and Bragg Reflectors 517

12.5.1 Introduction, 517
12.5.2 Coupled Mode Analysis, 520
12.5.3 Distributed Bragg Reflector, 524
12.5.4 A Quarter-Wave Bandpass Filter, 525

12.5.5 Distributed Feedback Lasers (Active Mirrors), 528
12.5.6 Tunable Semiconductor Lasers, 531

12.6 Unstable Resonators 534

12.6.1 General Considerations, 534
12.6.2 Unstable Confocal Resonator, 540

12.7 Integral Equation Approach to Cavities 543

12.7.1 Mathematical Formulation, 543
12.7.2 Fox and Li Results, 547
12.7.3 Stable Confocal Resonator, 550

12.8 Field Analysis of Unstable Cavities 555

12.9 *ABCD* Law for "Tapered Mirror" Cavities 562

12.10 Laser Arrays 568

12.10.1 System Considerations, 568
12.10.2 Semiconductor Laser Array: Physical Picture, 568
12.10.3 Supermodes of the Array, 570
12.10.4 Radiation Pattern, 574

Problems 574

References and Suggested Readings 585

13 Maxwell's Equations and the "Classical" Atom 589

13.1 Introduction 589

13.2 Polarization Current 590

13.3 Wave Propagation With Active Atoms 592

13.4 The Classical A_{21} Coefficient 596

13.5 (Slater) Modes of a Laser 597

13.5.1 Slater Modes of a Lossless Cavity, 598
13.5.2 Lossy Cavity With a Source, 600

13.6 Dynamics of the Fields 602

13.6.1 Excitation Clamped to Zero, 602
13.6.2 Time Evolution of the Field, 603

13.7 Summary 609

Problems 610

References and Suggested Readings 615

14 Quantum Theory of the Field-Atom Interaction 616

14.1 Introduction 616

14.2 Schrödinger Description 617

14.3 Derivation of the Einstein Coefficients 621

14.4 Dynamics of an Isolated Atom 624

14.5 Density Matrix Approach 627

14.5.1 Introduction, 627
14.5.2 Definition, 628

14.6 Equation of Motion for the Density Matrix 633

14.7 Two-Level System 635

14.8 Steady State Polarization Current 639

14.9 Multilevel or Multiphoton Phenomena 643

14.10 Raman Effects 651

14.10.1 Phenomena, 651
14.10.2 A Classical Analysis of the Raman Effect., 654
14.10.3 Density Matrix Description of the Raman Effect, 660

14.11 Propagation of Pulses: Self-Induced Transparency 665

14.11.1 Motivation for the Analysis, 665
14.11.2 A Self-Consistent Analysis of the Field-Atom Interaction, 666
14.11.3 "Area" Theorem, 670
14.11.4 Pulse Solution, 673

Problems 676

References and Suggested Readings 679

15 Spectroscopy of Common Lasers 681

15.1 Introduction 681

15.2 Atomic Notation 681

15.2.1 Energy Levels, 681
15.2.2 Transitions: Selection Rules, 682

15.3 Molecular Structure: Diatomic Molecules 684

15.3.1 Preliminary Comments, 684
15.3.2 Rotational Structure and Transitions, 685
15.3.3 Thermal Distribution of the Population in Rotational States, 686
15.3.4 Vibrational Structure, 687

15.3.5 Vibration-Rotational Transitions, 688
15.3.6 Relative Gain on P and R Branches: Partial and Total Inversions, 689

15.4 Electronic States in Molecules 691

15.4.1 Notation, 691
15.4.2 The Franck–Condon Principle, 692
15.4.3 Molecular Nitrogen Lasers, 692*

Problems 693

References and Suggested Readings 695

16 Detection of Optical Radiation 697

16.1 Introduction 697

16.2 Quantum Detectors 697

16.2.1 Vacuum Photodiode, 698
16.2.2 Photomultiplier, 699

16.3 Solid-State Quantum Detectors 701

16.3.1 Photoconductor, 701
16.3.2 Junction Photodiode, 703
16.3.3 p-i-n Diode, 706
16.3.4 Avalanche Photodiode, 707

16.4 Noise Considerations 707

16.5 Mathematics of Noise 709

16.6 Sources of Noise 713

16.1.1 Shot Noise, 713
16.6.2 Thermal Noise, 714
16.6.3 Noise Figure of Video Amplifiers, 716
16.6.4 Background Radiation, 717

16.7 Limits of Detection Systems 718

16.7.1 Video Detection of Photons, 718
16.7.2 Heterodyne System, 722

Problems 725

References and Suggested Readings 728

17 Gas-Discharge Phenomena 729

17.1 Introduction 729

17.2 Terminal Characteristics 731

17.3 Spatial Characteristics 732

17.4 Electron Gas 734

17.4.1 Background, 734
17.4.2 "Average" or "Typical" Electron, 734
17.4.3 Electron Distribution Function, 741
17.4.4 Computation of Rates, 743
17.4.5 Computation of a Flux, 745

17.5 Ionization Balance 746

17.6 Example of Gas-Discharge Excitation of a CO_2 Laser 748

17.6.1 Preliminary Information, 748
17.6.2 Experimental Detail and Results, 748
17.6.3 Theoretical Calculations, 750
17.6.4 Correlation Between Experiment and Theory, 753
17.6.5 Laser-Level Excitation, 756

17.7 Electron Beam Sustained Operation 758

Problems 761

References and Suggested Readings 764

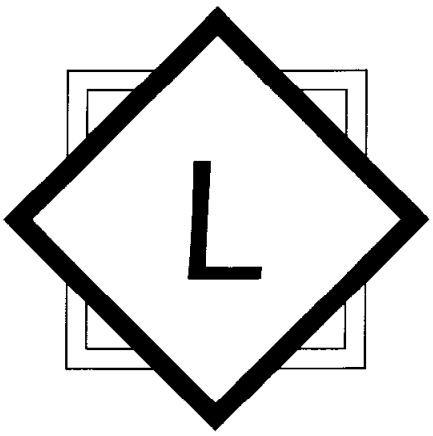
Appendices

I An Introduction to Scattering Matrices 765

II Detailed Balancing or Microscopic Reversibility 770

III The Kramers–Kronig Relations 774

Index 779



List of Symbols with Typical Dimensions

(Q in Coulombs; M mass in kg; L in length (cm or m); T in seconds; W in Watts; E in Joules; V in Volts; Temperature in K)

Roman Symbols

a	Attachment rate per unit of drift (L^{-1}), radius of a fiber (L)
$\mathbf{a}_n$	Unit vector in direction of n
a_w	Wiggler parameter (dimensionless)
$[A]$	Density of A (L^{-3})
A^*	Complex conjugate of A
$\langle A \rangle$	Expectation value of A
A_{21}	Einstein coefficient for spontaneous emission (T^{-1})
A, B, C, D	Components of ray matrix
$\mathbf{b}, \mathbf{B}$	Magnetic induction vector (Tesla = (Volt-sec)/L^2)
B_e	Rotational constant (always in cm^{-1})
B_v	$B_e - \alpha_v(v + \frac{1}{2})$, rotational constant within a band (always in cm^{-1})
B_{21}	Einstein coefficient for stimulated emmission (L^3-Energy^{-2}-T^{-2})
B_{12}	Einstein coefficient for absorption $B_{12} = g_2 B_{21}/g_1$

c		Velocity of light in a vacuum ($\sim 3 \times 10^{10}$ cm/s)
c'		c/n, phase velocity of light in a material with index n
$c_m(t)$		Probability of occupation of state m
d, **D**		Displacement vector (Q-L^2)
D		Delay dispersion (ps/km/nm)
$D_{n(p)}$		Diffusion coefficients for electrons (holes) (L^{-2}-T^{-1} usually in cm^2/s)
D_T		Transverse diffusion coefficient (L^{-2}-T^{-1} usually in cm^2/s)
e		Electronic charge (1.602018×10^{-19} coulombs)
e, **E**		Electric field intensity (Volts/L; usually V/m or V/cm)
$\bar{e}^2$, $\bar{i}^2$		Equivalent noise generators (Volts^2 or Amps^2)
E		Energy (in Joules)
$E_{c(v)}$		Energy of the conduction (valence) band (in Joules)
E_F		Fermi energy (in Joules)
f		Focal lengths (L)
	or	Frequency (T^{-1})
	or	Laser cavity fill factor (dimensionless)
$f(E)$		Fermi function (dimensionless)
$f(v)$		Distribution function for speed (L^{-3}-velocity^{-3})
$f(v_x, v_y, v_z)$		Distribution function (L^{-3}-velocity^{-3})
$f(z)$		Wave propagating in the $+z$ direction (Volts/L)
f_{12}		Absorption oscillation strength (dimensionless)
F		Finesse = FSR/($\Delta\nu_{1/2}$) (dimensionless)
$F(\epsilon)$		Electron distribution function per unit of energy (Energy^{-1})
$F(J)$		Rotation energy (cm^{-1})
$F_{n(p)}$		Quasi-Fermi level for electrons (holes) (Energy in Joules)
FSR		Free spectral range = $c/2d$ (frequency in Hertz (T^{-1}))
FWHM		Full width at half maximum
g		$\int \gamma dz \approx \gamma l_g$, the line integrated gain (dimensionless)
$g(v)$		Lineshape (frequency^{-1} or T)
$g_{1,2}$		$(1 - d/R_{1,2})$, the g parameter of a cavity
	or	$2J_{1,2} + 1$, the degeneracy of quantum states (1, 2)
$\bar{g}(v)$		Lineshape normalized to unity at line center, i.e., $\bar{g}(v_0) = 1$ (dimensionless)
G		Power gain ($P_{\text{out}}/P_{\text{in}}$)
$G(r, r_0)$		Green's function
$G(v)$		Energy of vibration state v (always in cm^{-1})
G_0		Small signal power gain
h		Planck's constant (6.626076×10^{-34} Joule-second)

$\hbar$		Planck's constant divided by 2π (1.05457×10^{-34} Joule-second)
h, H		Magnetic field intensity (Ampere/L)
H		Operator corresponding to the Hamiltonian or total energy
$H_n(u)$		Hermite polynomial of order n argument u
I_ν		Intensity at a frequency ν (Watts/area)
$I(\nu)$		Intensity per unit of frequency at ν (Watts-Frequency^{-1}-L^{-2} = Watts-T-L^{-2})
$Im(\ )$		Imaginary part of the quantity ( )
j		The imaginary number $(-1)^{1/2}$
j, J		*Conduction current* (Amperes/area)
J		Angular momentum quantum number
k		$\omega n/c = 2\pi n/\lambda_0$ with $k_0 = \omega/c = 2\pi/\lambda_0$ (L^{-1})
k		Wave vector $= k\mathbf{a}_n$ (L^{-1})
K.E.		Kinetic energy (in Joules)
l_g		Length of gain medium (L)
ln( )		Natural log of ( )
L		Laser intensity normalized to a saturation value
$\mathcal{L}(\ )$		Laplace transform of ( )
$m^*_{c(v)}$		Effective mass in the conduction (valence) band (M in kg)
m_0		Free electron rest mass (9.1094×10^{-31} kg)
M		Magnification of a beam (dimensionless)
n		Population difference $(N_2 - N_1) \cdot$ Vol. (dimensionless)
	or	Index of refraction $(\epsilon_r)^{1/2}$ (dimensionless)
$n_c(E)$		Density of electrons in conduction band per unit of energy (L^{-3}-Energy^{-1})
n_e		Electron density (L^{-3})
n_g		Group index
$n(\nu)$ or $n(\lambda)$		Frequency- or wavelength-dependent refractive index (dimensionless)
n_{th}		Threshold value of population difference $[N_2 - (g_2/g_1)N_1] \cdot$ Volume
N		Fresnel number (dimensionless)
N_{eq}		Equivalent Fresnel number (dimensionless)
NLS		Nonlinear Schrödinger equation
N_p		Number of photons in the laser cavity
N_{tr}		Density of electron/holes at optical transparency (L^{-3})
N_ν		Number of modes in a volume V between 0 and ν
$N_{2,1}$		Density of states 2, 1 (L^{-3})
$N_2(v)$		Nitrogen in a vibrational state v
p		Hole density (L^{-3})
	or	Mode index
	or	Number of modes (or states) per unit of volume (L^{-3})

	or	Power per unit of volume (Watts-L^{-3})
	or	Pressure (Newtons-L^{-2})
	or	$[\kappa^2 + (g - j\delta)^2]^{1/2}$; the coupled mode phase constant (L^{-1})
$p(t)$		Instantaneous power (averaged over a few optical cycles) (in Watts)
$\mathbf{p}_a$, $\mathbf{P}_a$		Polarization vector of the active atoms (Q-L^{-2})
p_{el}		Power into electron gas (Watts-L^{-3})
$p_v(E)$		Density of holes in the valence band per unit of energy (L^{-3}-energy^{-1})
$p(\nu)$		Mode density (per unit of frequency) at ν (L^{-3}-frequency^{-1})
P		Optical power normalized to a saturation value
	or	Pumping rate (L^{-3}-T^{-1})
$P_{1,2}$		Apparent source point for the limited extent spherical wave (dimensionless)
$\langle P \rangle$		Average power (averaged over many cycles) (Watts)
P_f		Fluorescence power (Watts)
$\mathcal{P}_s$		Probability of belonging to a class s
q		Axial mode number of a resonant mode in a cavity
	or	The number of half-wavelengths between the mirrors
	or	Index of a sub-band in a semiconductor
$q(z)$		Complex beam parameter (L) with
$1/q(z)$		$= 1/R(z) - j\lambda/[\pi n w^2(z)]$ (L^{-1})
Q		Quality factor $= \omega W/(-dW/dt)$ (dimensionless)
QSE		Quantum size effect
$\mathbf{r}$		Position vector $= x\mathbf{a}_x + y\mathbf{a}_y + z\mathbf{a}_z$
r_{ij}		Rate connecting states i and j
$r_{1,2}$		Fraction of the distance between the mirrors $M_{1;2}$ to the points $P_{1;2}$
$r_{\min}$		Equilibrium spacing in a stable molecule (Å)
$r(z)$		Wave in the reverse or negative z direction (Volts/L)
R		Resistance (Ω)
R		If the numerical value > 1, radius of curvature (L)
	or	If the numerical value < 1, power reflectivity (dimensionless)
$Re(\,)$		Real part of the quantity ( )
$\Re(\omega)$		Raman line shape ((radian frequency)$^{-1}$ or T)
RWA		Rotating wave approximation
$R(z)$		Radius of curvature of phase front (L^{-1})
$R(\nu)$		Recombination spectra from a semiconductor (Watts-L^{-3}-frequency^{-1})
s		Fraction of the filed surviving a round trip $= S^{1/2}$
	or	Laplace transform variable (T^{-1})
S		Fraction of the photons surviving a round trip (dimensionless)
	or	Poynting vector (Watts/area)

S_{ij}		Elements of the scattering matrix (dimensionless)
S/N		Signal to noise ratio
$S_T(\nu)$		Power per unit of frequency at ν (Watts-frequency^{-1})
t		Field transmission coefficient (dimensionless)
T		Temperature (K)
T_e		Electronic term energy (in cm^{-1} or eV)
TE		Transverse electric mode
TEM		Transverse electric and magnetic mode
TM		Transverse magnetic mode
T		Ray matrix
T_1		Lifetime of the inversion ($\rho_{22} - \rho_{11}$) (T)
T_2		Mean time between dephasing collisions (T)
v		Speed or velocity (L-T^{-1})
	or	Vibrational quantum number (dimensionless)
	or	Perturbation of the potential energy (Joules)
v_g		Group velocity (L-T^{-1})
v_p		Phase velocity $= \omega/\beta$ (L-T^{-1})
v_z		Velocity in z direction (L-T^{-1})
V		Voltage (Volts)
$V_{m,n}$		Volume of the TEM$_{m,n}$ mode (L^3)
Vol.		Volume (L^3)
VSWR		Voltage standing wave ratio (dimensionless)
w		Energy per unit of area (volume) (Joules-L^{-2} or (Joules-L^{-3}))
w_d		Drift velocity (L-T^{-1})
w_e		Energy of the electron gas (Joules-L^{-3})
w_0		Minimum spot size (L)
w_s		Saturation energy (per unit of area) (Joules/area)
$w(z)$		Spot size as a function of z (L)
W		Energy as a dependent variable (Joules)
z_0		Characteristic length parameter of a Gaussian beam
Z		Impedance (Ω)
Z_0		Characteristic impedance of a transmission line (Ω)

Greek Symbols

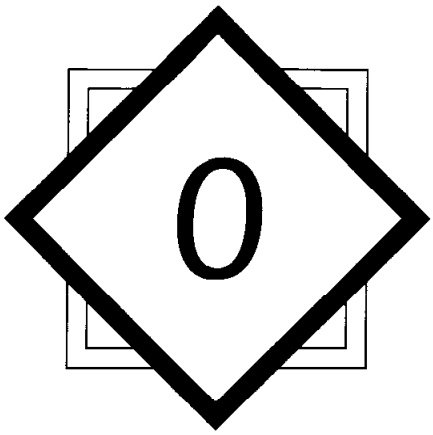
α		Absorption coefficient (loss per length) (cm^{-1})
	or	Townsend ionization coefficient (ionization rate per unit of drift) (cm^{-1})
α_e		Correction to the rotational constant due to vibration (dimensionless)
β		Phase constant of a guided wave (rad/length)

β_0		Unperturbed phase constant
$\beta_m = \pi/\Lambda_m$		Phase constant satisfying the Bragg condition
$\gamma(v)$		Intensity-dependent gain coefficient (L^{-1})
$\gamma_0(v)$		Small signal gain coefficient (L^{-1})
Γ		Field reflection coefficient (dimensionless)
	or	Optical confinement factor (dimensionless)
	or	Electric field reflection coefficient (dimensionless)
Γ_p		Photon flux ($I/h\nu$) (L^{-2}-T^{-1})
δ		Secondary emission ratio (dimensionless)
	or	Fraction of the electron's excess energy lost in an elastic collsion (dimensionless)
$\Delta t_{1/2}$		Pulse width (FWHM) (T)
$\Delta\nu_D$		Doppler line width (Hz)
$\Delta\nu_h$		Homogeneous line width (Hz)
$\Delta\nu_H$		Hole line width (Hz)
$\Delta\nu_n$		Natural line width (Hz)
$\Delta\omega$		Line width in radian frequency units (radians-T^{-1})
ϵ		Electron energy (Joules)
ϵ'		Real part of the relative dielectric constant (dimensionless)
ϵ''		Imaginary part of the relative dielectric constant (dimensionless)
ϵ_0		Permittivity of free space (8.85×10^{-12} F/m)
ϵ_A		Characteristic energy of atoms or molecules (Volts)
ϵ_k		Characteristic energy of electrons $= D_T/\mu$ (Volts)
ϵ_r		Relative dielectric constant (n^2)
η_0		Wave impedance of free space $(\mu_0/\epsilon_0)^{1/2} = 377\ \Omega$
η_{cpl}		Coupling efficiency
η_{qe}		Quantum efficiency
η_{xtn}		Extraction efficiency
λ		Wavelength λ_0/n
λ_0		Free-space wavelength
λ_q		Wavelength of the $\text{TEM}_{m,m,q}$ mode
Λ		Characteristic length in a periodic structure (L)
μ		Mobility (cm^2/(Volt-s))
μ_0		Permeability of free space ($4\pi \times 10^{-7}$ H/m)
μ_{21}		Electric dipole moment (Q-L)
ν, f		Frequency (Hz = T^{-1})
$\tilde{\nu}$		Wave number (number of wavelengths per centimeter (always in cm^{-1}))
ν_0 or ν_{21}		Line center of $2 \rightarrow 1$ transition (frequency or T^{-1})
ν_c		Collision frequency (T^{-1})

ν_i	Ionization rate (per electron) (T^{-1})
$\rho(\nu)$	Energy per unit of volume per unit of frequency at ν (Joules-T-L^{-3})
ρ_ν	Energy per unit of volume at a frequency ν (Joules-L^{-3})
ρ_{11}	Density matrix element corresponding to the fraction of excited atoms in state 1
ρ_{21}	Density matrix element related to the induced polarization
ρ_{22}	Density matrix corresponding to the fraction of excited atoms in state 2
$\rho_{jnt}(h\nu)$	Joint density of states per unit of energy (L^{-3}-Energy^{-1})
$\rho_{jnt}(\nu)$	Joint density of states per unit of frequency (L^{-3}-T)
$\sigma(\upsilon)$	Stimulated emission cross-section (area = L^2)
$\sigma_{abs}(\nu)$	Absorption cross-section (area)
$\sigma_c(E)$	Collision cross-section (L^2)
$\sigma_i(E)$	Ionization cross-section (L^2)
τ_1	Lifetime of state 1 (T)
τ_2	Lifetime of state 2 (T)
$1/\tau_{1,2}$	Decay rate of state 1 (or 2) (T^{-1})
τ_{21}^{-1}	Decay rate of state 2 into state 1 (T^{-1})
τ_p	Photon lifetime (T)
τ_r	Radiative lifetime (T)
τ_{RT}	Time for a round trip (T)
ϕ, θ	Phase shift or geometric angles
ϕ_{12}	Branching ratio (dimensionless)
χ_a	Electric susceptibility of the active atoms (dimensionless)
χ_a'	Real part of susceptibility of the active atoms
χ_a''	Imaginary part of susceptibility of the active atoms
Ψ	Wave function
ω	Radian frequency = $2\pi\nu$ (radians-T^{-1})
Ω	Rabi frequency ($\mu_{21}E/2\hbar$) (radians-T^{-1})

Other Mathematical Symbols

$\triangleq$	By definition
$\approx$ or $\cong$	Approximately equal to
$\equiv$	Identical to
$\lvert(\,)\rvert$	Absolute value of ( )
∇	Gradient (L^{-1})
$\nabla\cdot$	Divergence (L^{-1})
$\nabla\times$	Curl (L^{-1})
∇^2	Laplacian
∂	Partial deriviative



Preliminary Comments

Before the 1960s, optics formed the basis for a relatively small industry involving rather sedate and mature topics such as optical instruments, cameras, microscopes, and scientific applications. Then the laser came on the scene, first the solid-state (ruby) laser, then the gas laser, then the semiconductor injection laser. Now optics form the basis for many more functions, products, and services.

At first, the standard joke was, "The laser is a solution in search of a problem." More seriously, almost everybody recognized the potential of the laser in communications; in data processing, storage, and retrieval; and even in eye surgery. The question to be answered was, Could the laser do things that had not been done before, and could it do things better and more economically than had previous devices and technologies? It is interesting to observe that the initial applications of the laser have not been in the rather obvious fields just listed but in new applications by ingenious people who understood the principles of the laser and who understood the problems to be solved. Hence, we have laser transits, laser pattern cutting, laser cutting of steel, and laser fusion, and we are starting to make inroads in optical communications. The history of the laser in the field of communications illustrates the point about "obvious" applications.

The frequency of the first laser, ruby, at $\lambda = 694.3$ nm is 4.32×10^{14} Hz, a quantity of interest to any communication engineer. If only 1% of this carrier frequency is used for the information bandwidth, then we have a communication channel that has two to three orders of magnitude (10^2 to 10^3) more capacity than the widest band channel in existence. Some

of the microwave radio-relay links used by the telephone company have channel widths as large as 10% of the carrier frequency. Consequently, one laser beam should be able to carry a huge number of telephone conversations (bandwidth required per telephone conversation, 4 kHz) and many television programs (bandwidth $\sim$ 5 MHz) simultaneously.

There are, however, a few problems. We do not know how to modulate this carrier at a 4×10^{12} Hz rate; nor does the technology exist to demodulate at this rate nor can our terminal equipment handle information at this rate. If that is not enough, we are not overly confident of being able to transmit the information from point A to a distant point B with the reliability afforded by microwave links. Finally, there was some doubt as to the reliability of the laser. Consequently, communications by lasers with the same degree of sophistication as is done at microwave frequencies lies in the future, but some inroads are being made.

The invention of glass fibers exhibiting very low loss has made laser communications a viable alternative to wired links. After all, the world has practically exhausted high-grade copper ore supplies, but we have not really touched the primary ingredient of glass, SiO_2 (i.e., sand). Thus the first "obvious" application of the laser, communications, had to wait until 1977 for trial runs over short-haul links.

Now, fiber optic links are the dominant communication channel being installed. A significant question is "When will fiber arrive at the home?"

The point to be made is that obvious applications are not so straightforward and simple. Most often, it is the materials that are the major impediment, but this should not stop us from looking for other uses. For instance, a first major use of the ruby laser was in the "trimming" of solid-state circuit components. In that case, we use the ability to focus the energy of the laser onto a very small spot so as to vaporize the excess material.

Recently, there has been the marriage of xerography, word processing with computers, and the ability to modulate, deflect, and focus a laser beam to produce manuscript with nearly the quality of offset printing, but at a fraction of the cost in capital equipment.

In view of this, then, we will not look at a laser from an applications standpoint. Rather, we will try to introduce the elementary and *simple* principles of the laser itself, the propagation of its radiation, and the elements of the detection problem. The word "simple" was italicized to emphasize the goal of this textbook: to make the reader feel comfortable with the following issues:*

1. What physical principles are involved in *generation* of the laser radiation? (2)
2. What peculiarities can be anticipated in the *transmission* of laser beams? (1)
3. What are the characteristics and limitations of common detectors? (3)

Once the reader feels comfortable with an initial understanding, he or she can then read the more advanced texts and current literature to obtain the finer points of the field of quantum electronics.

One final comment should be made about the material. Lasers are quantum devices. There is no avoiding that fact. However, it is not necessary to be familiar with every rule,

*The numbers in parentheses represent the order of the topics covered in this book.

regulation, philosophy, and theorem of quantum mechanics to develop a good understanding of lasers. Lasers could have been invented in 1908 before the vacuum triode tube.

All the necessary quantum theory was available by 1930—indeed the formula for the amplification or gain coefficient can be found in reference 1 of Chapter 7 (first published in 1934)—and thus the laser could have been invented then. Why did scientists take so long (about 30 years) to obtain an oscillator at optical frequencies? The answer is, of course, speculation, but the question truly does generate wonderful philosophy. According to Townes [1] one of the Nobel Prize winners for the laser, the time lag can be attributed to physicists' knowing the atom-field interaction but being less than dexterous with the feedback requirement for an oscillator, whereas the electrical engineers had the opposite strength and weakness. (Refs. [1] to [4] are papers published for the twenty-fifth anniversary of the laser and are highly recommended.) Thus it was not until between 1950 and 1960 that the field of quantum electronics boomed.

Much of the credit for this boom must be given to the microwave industry, a technology that was developed during and after World War II (and is still going strong). Many of the concepts can be taken over into the laser field en masse; and after complicated microwave tubes, lasers are, in some respects, quite simple.

NOTE TO THE STUDENTS

The object of the first part of this book is to provide sufficient tools to understand the generation and propagation of coherent optical radiation without all of the mystic and wonderful theorems associated with quantum mechanics. Most of the initial work is merely an extension of lower frequency concepts with the proviso that one moves the decimal point four to five places to the right on the frequency scale. This requires a few approximations to justify this extension, but those are usually quite transparent and palatable to the most casual observer. They lead to a simpler description of a generator of coherent waves than is possible at RF or microwave frequencies but there are differences.

For instance, most optical amplifiers and components are bilateral (i.e., they transmit the same relative value when driven from the right or left terminals, a feature considerably different from a simple transistor amplifier). An optical amplifier is naturally bilateral, and techniques for the suppression or encouragement of oscillation are easily identified. It is very difficult to predict the performance of the transistor amplifier outside of its Class A or linear regime, but it is trivial to consider an optical amplifier at any level until the optical signal becomes so large that the amplifier self-destructs (i.e., melts or explodes) or multiphoton effects occur. Other contrasting simplifications will become apparent later.

Electromagnetic theory is most important, and all of the phenomena discussed in elementary courses (for RF and microwave frequencies) can be taken over to the optical domain with the shift in decimal point mentioned earlier. Not until one is quite advanced in laser phenomena is the quantization of the field needed. The only issue needed that is not explicit in Maxwell's equations is that energy comes in discrete packages, a multiple of $h\nu$. But most of our cases involve "billions" of photons so ± 1 (more or less) is not significant. Even in the extreme of a few number of photons, classical electromagnetic

theory will handle most cases, even for the cases that are given as "examples" of the necessity for the photonic character of light, such as the photoelectric effect. The atoms or the material must be quantized, but we presume that someone else has done the hard work of measuring the pertinent coefficients and our job is to understand the results and use them. This approach is called the "semiclassical" quantum theory and is discussed in greater detail in later chapters. The major shift in your thinking is that *voltage* and *current* are no longer measurable quantities. Power (or intensity) is always the ultimate goal for all predictions.

After teaching this material for a number of years and to about 500 students, I have compiled an advice list:

1. Dust off your electromagnetic theory. With a minimal amount of change in your thinking, you can handle much of laser electronics.
2. Do not be a slave to the SI (or MKSA) system of units. This book will use those units in the theory, but "practical units" will be used for evaluation. Much of the published literature uses the cgs system of units (for reasons that escape me), but they usually give the results in transparent practical units that are easily converted to a "pure" SI format. However, doing such a conversion is usually an unnecessary step that introduces a finite probability of error. For instance, you will see that the product of density (L^{-3}), a cross-sectional area (L^2), and a length (L) plays a critical role in laser theory. Since the product is dimensionless, why go through the exercise of converting length in centimeters to meters three times? Only if ϵ_0 or μ_0 appears in the solution—separately—is it necessary to make the conversion to SI units for the rest of the expression.
3. Become familiar with different ways of specifying fundamental quantities. For instance, the frequency ν is properly specified in (Hz), but equally informative is wavelength [in nm (10^{-9} m), Å (10^{-10} m), or μm (10^{-6} m)] with $\lambda_0 = c/\nu$, wavenumber units $\bar{\nu} = 1/\lambda$ (always expressed in cm^{-1} units), or in energy units of $h\nu$ (Joules) or $h\nu/e$ (volts). There is no accepted way of specifying this information in the literature so you must be familiar with all of them.
4. Dust off your differential equation background. The most common and important ones are those that are linear and of first order in the derivatives, and they are usually coupled. Laplace transforms can be used if you wish. The rare cases of second-order differential equations encountered here are usually solved either by the trigonometric or the hyperbolic functions. It will save you considerable work if you recall that all arguments of mathematical functions must be dimensionless, and hence a distance variable z will always be multiplied by quantities with a dimension L^{-1} and time t will be divided by a characteristic time constant.
5. Learn the dimensions of all symbols. Develop the habit of including the dimensions of all calculations even to the point of including the word *dimensionless* if appropriate. It does no good to use obscure dimensions such as coulombs per second when the common ampere should be familiar to all.

6. Do not depend upon the canned symbolic math programs to do the analytic work for you. The thought process expended in obtaining an analytic solution is most valuable. Reserve the computer for cases that are nonlinear and/or too complicated.
7. Be willing to consider extremes. For instance, an infinite intensity will be considered—obviously an impossibility—but we can come close. Quite often an examination of the extremes yields the upper and lower bounds to the phenomena.
8. Finally, memorize the following equation for the necessary condition for any oscillator.

$$\boxed{\text{the net round trip gain} \geq 1} \tag{0.1}$$

The net round trip gain is the product of all amplification and attenuation ratios as the wave makes a round trip. Thus if G_0 is the small signal power gain per pass through the amplifier, S = the fraction of the power surviving each pass due to imperfect passive components (thus $1 - S = L$, the fraction of the power lost per pass), then for the simple helium/neon laser shown in Fig. 0.1,

$$G_0^2 S^2 > 1 \qquad \text{or} \qquad G > 1/S$$

We will see the application of this equation many, many times. It is too simple not to commit to memory.

Consider the simple He/Ne gas laser operating at 632.8 nm shown in Fig. 0.1. If R represents the power reflectivity of the mirrors, L the loss per pass through the windows, and G the power gain through the tube per pass, the laser will oscillate provided that

$$G(1 - L_2)R_2(1 - L_2)G(1 - L_1)R_1(1 - L_1) \geq 1$$

In writing this equation, we have broken the "loop" at the right of window 1 and followed a wave around the path. The equation is trivial and transparent.

Some of the interesting problems are (1) How do we excite the system to get the gain G? (2) Are there any special techniques to construct the mirrors? (3) Why use curved mirrors? (4) Why orient the windows as shown? (5) What is the beam spread? (6) How much power do we obtain? (Obviously, it must be less than we put into the system.)

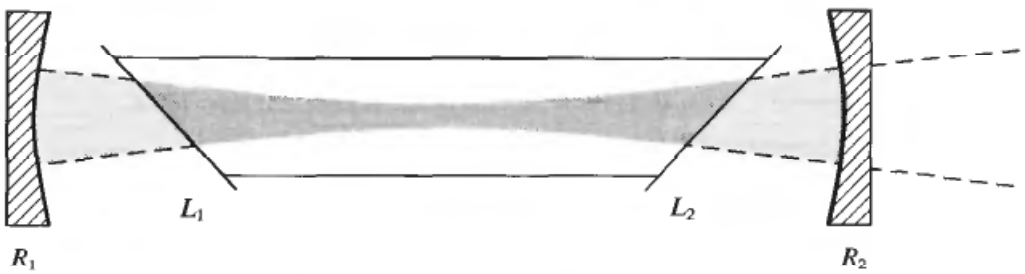


FIGURE 0.1. Schematic of a simple laser.

Interestingly enough, quantum theory enters only in the choice of the gases involved, helium and neon, and then only to provide two energy states separated by

$$E = h\nu \qquad \nu = \frac{c}{\lambda}, \qquad \lambda = 632.8 \text{ nm} \tag{0.2}$$

Then a few relatively simple equations relate the gain to the number of atoms in each of these two states. All the other problems listed can be discussed to an unusual degree of precision without once invoking the quantum nature of the device.

Most readers will be familiar with the theory of the simple pn junction for rectification of AC signals and as an integral part of transistors and other solid-state devices. These are the "complicated" applications of semiconductor electronics that depend, to a major degree, on the differences between "forward" and "reverse" bias. The semiconductor injection laser uses this same pn junction in the forward direction to promote the stimulated recombination of the electrons and holes.

$$e + h \rightarrow h\nu \tag{0.3}$$

The basic physics is quite simple. The technology has benefited from some rather ingenious thinking, so that now the semiconductor laser is the overwhelming choice for low-power communication and control applications.

However, the point remains: lasers are quantum devices, a fact that we accept, live with, enjoy, and frequently ignore.

REFERENCES

See the Centennial Issue of IEEE J. Quant. Electron. *QE-20,* 1984

1. C. H. Townes, "Ideas and Stumbling Blocks in Quantum Electronics," *IEEE J. Quant. Electron. QE-20*, 547, No. 6, 1984.
2. W. E. Lamb, Jr., "Laser Theory and Doppler Effect," *IEEE J. Quant. Electron. QE-20*, 551, 1984.
3. N. Bloembergen, "Non-linear Optics," *IEEE J. Quant. Electron. QE-20*, 556, 1984.
4. A. L. Schawlow, "Lasers in Historical Perspective," *IEEE J. Quant. Electron. QE-20*, 558, 1984.
5. See also the historical section of IEEE J. Quant. Electron. *QE-20*, 1987—25th Anniversary of Semiconductor Lasers.
6. The five-volume set entitled *The Laser Handbook* (New York: North-Holland Publishing Company:

   Volume 1, Eds. F. T. Arecchi and E. O. Schulz-Dubois, 1972.

   Volume 2, Eds. F. T. Arecchi and E. O. Schulz-Dubois, 1972.

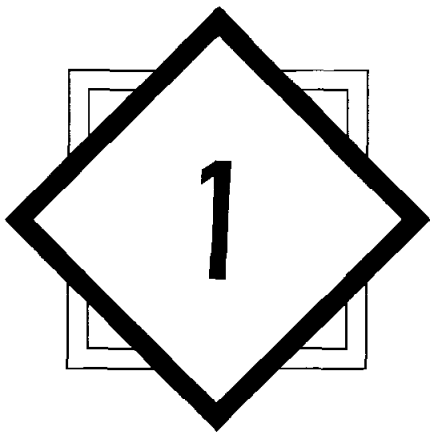
   Volume 3, Ed. M. L. Stitch, 1974.

   Volume 4, Eds. M. L. Stitch and M. Bass, 1979.

   Volume 5, Eds. M. Bass and M. L. Stitch, 1985.
7. R. J. Pressley, Editor in-Chief, *Handbook of Lasers* (Cleveland, Ohio: Chemical Rubber Co.), 1971.

8. There are "thousands" of known lasers spanning the wavelength range of far infrared to the UV and x-ray portion of the spectrum. A reasonably complete listing for gases is given by R. Beck, W. Englisch, and K. Gürs, *Table of Laser Lines in Gases and Vapors*, Springer Series in Optical Sciences, 3rd ed. (New York: Springer-Verlag, 1980).

References [1] to [4] are very easy to read and can give a sense of the historical perspective about a field that is exploding. Reference [5] is a volume of the IEEE Transaction on Quantum Electronics, which commemorated the twenty-fifth anniversary of the very important semiconductor laser. The last three are general handbooks that are useful for physical properties of optical materials, laser wavelengths, and specialized phenomena.



Review of Electromagnetic Theory

1.1 INTRODUCTION

We will be dealing with electromagnetic waves in that part of the spectrum where optical techniques have played a historical role. Lenses are used to focus the radiation, mirrors to direct it, and free space to transmit it. Yet it is still electromagnetic radiation, it obeys Maxwell's equations, and all the laws studied at low frequencies apply at the "optical" portion of the spectrum.

The major difference lies in the size of the components used. For instance, a 1-cm-diameter capacitor used at 1 MHz is less than 10^{-4} of a free-space wavelength ($\lambda = 300$ m), whereas a 1-cm-diameter "contact lens" for your eyes is greater than 10^4 wavelengths for visible radiation. The small size of the capacitor compared to a wavelength is a requirement for the validity of circuit theory; however, the large size of the lens makes life easy for the more exact field theory.

Before we go into field theory at optical frequencies, let us mention the question of units. The *rationalized* SI (or MKSA) system will be used throughout in analytical developments. However, numerical answers will almost always be expressed in cm, cm^{-3}, or cm/sec. This does not mean we are using a CGS system of units but merely that we are expressing an answer in a more convenient and intuitively comfortable form, as well as conforming to most modern and traditional literature. Only if ϵ_0, the permittivity of free space, or μ_0, the permeability of free space, appears in the equation must we go through

the exercise of converting centimeters to meters. Most of the time, the product appears (i.e., $\mu_0\epsilon_0$), which is, of course, equal to $1/c^2$. In that case, we can keep c as $\sim 3 \times 10^{10}$ cm/sec in all the equations, provided that the other quantities are also measured in centimeters.

1.2 MAXWELL'S EQUATIONS

To describe an electromagnetic wave, we need two field-intensity vectors, **e** and **h**, which are related to each other by

$$\nabla \times \mathbf{h} = \mathbf{j} + \epsilon_0 \frac{\partial \mathbf{e}}{\partial t} + \frac{\partial \mathbf{p}}{\partial t} \tag{1.2.1a}$$

$$\nabla \times \mathbf{e} = -\mu_0 \frac{\partial \mathbf{h}}{\partial t} \tag{1.2.1b}$$

where **p** is the polarization current induced by the electric field. (A term of the form $\partial \mathbf{m}/\partial t$ can be added to (1.2.1b) but will be ignored for now.) We use lowercase letters to represent vectors that are explicit functions of time t and the three spatial coordinates x, y, and z. Most of the time we will be talking about sinusoidal variations of the field and use the phasor representation

$$\begin{aligned} \mathbf{e}(\mathbf{r}, t) &= Re\left[\mathbf{E}(\mathbf{r})e^{j\omega t}\right] \qquad & \mathbf{h}(\mathbf{r}, t) &= Re\left[\mathbf{H}(\mathbf{r})e^{j\omega t}\right] \\ \mathbf{j}(\mathbf{r}, t) &= Re\left[\mathbf{J}(\mathbf{r})e^{j\omega t}\right] \qquad & \mathbf{p}(\mathbf{r}, t) &= Re\left[\mathbf{P}(\mathbf{r})e^{j\omega t}\right] \end{aligned} \tag{1.2.2}$$

where Re is real part, $\mathbf{r} = x\mathbf{a}_x + y\mathbf{a}_y + z\mathbf{a}_z$, $\mathbf{a}_i$ is the unit vector in the ith direction, and the capital letters **E** and **H** are complex vector quantities depending on space coordinates but not on time. We recognize that if we want the complete field, we must take the real part of the product **E** exp $(j\omega t)$.

If we substitute (1.2.2) into (1.2.1) we obtain the time-independent form of Maxwell's equations:

$$\begin{aligned} \nabla \times \mathbf{H} &= \mathbf{J} + j\omega\epsilon_0\mathbf{E} + j\omega\mathbf{P} = \mathbf{J} + j\omega\mathbf{D} \\ \nabla \times \mathbf{E} &= -j\omega\mu_0\mathbf{H} \end{aligned} \tag{1.2.3}$$

with

$$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P}$$

where the common factor of exp $(j\omega t)$ has been canceled from each side of the equation. The polarization term is related to the electric field by a constitutive relation:

$$\mathbf{P} = \epsilon_0\chi\mathbf{E} \tag{1.2.4}$$

where the term χ is the complex susceptibility of the medium through which the wave is propagating.

After we have become familiar with the simple approach to lasers, we will find that the atoms enter Maxwell's equations via an "equation of motion" for **P**; but for now we assume that the coefficient χ is a given parameter of the medium. For instance, the form of

the polarization given by (1.2.4) suggests that it and the vacuum displacement term can be combined into a single term:

$$\begin{aligned}\mathbf{D} &= \epsilon_0\mathbf{E} + \mathbf{P} \\ &= \epsilon_0(1+\chi)\mathbf{E} \\ &= \epsilon_0\epsilon_\tau\mathbf{E} = \epsilon_0 n^2\mathbf{E}\end{aligned} \tag{1.2.5}$$

Thus the relative dielectric constant ϵ_τ is related to the susceptibility by $1+\chi$, and it in turn is equal to the square of the index of refraction. In the interest of simplicity, **P** was assumed to be in the same direction as **E**, but this is not true for many of the interesting electro-optic materials.

Actually, we have done something very important in going from (1.2.1) to (1.2.3, 1.2.4, and 1.2.5). We have gone from the time domain to the angular-frequency domain, ω, by the application of the Fourier transform, defined by

$$F(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t}\,dt \tag{1.2.6}$$

If we follow the prescription of (1.2.6) as applied to Maxwell's equations, we obtain (1.2.3) directly. Consequently, **E**, **H**, **P**, **D**, and **B** are also spectral representations of the respective quantities and should be written as $\mathbf{E}(r, \omega)$ $\mathbf{H}(r, \omega)$, etc. However, since we are usually dealing with a single frequency, we are often lazy and do not bother to show that dependence. Unfortunately, this laziness will return to haunt us unless we are forewarned.

1.3 WAVE EQUATION FOR FREE SPACE

Let us consider free space, so that the conduction current **J** is zero. If we take the curl of (1.2.1b) and eliminate **h** by the use of (1.2.1a) we obtain

$$\left.\begin{aligned}\nabla\times\nabla\times\mathbf{e} &= \mu_0\frac{\partial}{\partial t}(\nabla\times\mathbf{h}) = -\mu_0\epsilon_0\frac{\partial^2\mathbf{e}}{\partial t^2} \\ \text{or}\qquad \nabla^2\mathbf{e} - \frac{1}{c^2}\frac{\partial^2\mathbf{e}}{\partial t^2} &= 0\end{aligned}\right\} \tag{1.3.1a}$$

where $c^2 = 1/\mu_0\epsilon_0$ is the square of the velocity of light. If the procedure is reversed to eliminate **e**, we obtain the same equation with **h** substituted for **e** in (1.3.1a):

$$\nabla^2\mathbf{h} - \frac{1}{c^2}\frac{\partial^2\mathbf{h}}{\partial t^2} = 0 \tag{1.3.1b}$$

It is most important to realize that *any function of the form* $f(t - \mathbf{a}_n\cdot\mathbf{r}/c)$ *is a solution*, where $\mathbf{a}_n$ is a unit vector. It is easy to show this in one dimension and only slightly more complicated to do so for the general case. Physically, it merely means that the wave propagates in the direction of $\mathbf{a}_n$ with a velocity of c.

For sinusoidal representation, we say that there is a phase change as the wave propagates along the direction described by $\mathbf{a}_n$.

$$\mathbf{e}(\mathbf{r}, t) = Re\left\{[\mathbf{E}(\omega, \mathbf{k}_0)] \exp\left[j\omega\left(t - \frac{\mathbf{a}_n \cdot \mathbf{r}}{c}\right)\right]\right\} \tag{1.3.2a}$$

$$\mathbf{e}(\mathbf{r}, t) = Re\left\{\left[\mathbf{E}(\omega, \mathbf{k}_0)\right] \exp(j\omega t)\,(-j\mathbf{k}_0 \cdot \mathbf{r})\right\} \tag{1.3.2b}$$

$$|\mathbf{k}_0| = \frac{\omega}{c} = \frac{2\pi}{\lambda_0} \tag{1.3.3}$$

where λ_0 is the wavelength in free space.

In writing (1.3.2) we took the functional form of (1.2.2) and, in every place that t appeared, we replaced it by $t - \mathbf{a}_n \cdot \mathbf{r}/c$, just as the solution to the wave equation demanded. We also combined ω, c, and the unit vector $\mathbf{a}_n$ into a new vector $\mathbf{k}_0$. Obviously, the equation for $\mathbf{h}$ is modified in the same manner. We will use $\mathbf{k}_0$ to denote the wave vector in free space and $\mathbf{k}$ (without the subscript) to indicate it in a dielectric medium.

We could have been more formal in our approach and started with $\mathbf{k}_0$ as a three-dimensional Fourier transform variable with respect to the three spatial coordinates. Again, we tend to be somewhat lazy and not bother to state that E is now a function of $\mathbf{k}_0$ in addition to being a function of ω. Thus, most of the time, we say that we are representing a wave of constant amplitude E propagating along the direction $\mathbf{k}_0$.

1.4 ALGEBRAIC FORM OF MAXWELL'S EQUATIONS

If we take (1.3.2) and the corresponding one for $\mathbf{h}$ and insert them into Maxwell's equation for free space, we obtain the algebraic form of Maxwell's equations:

$$\begin{Bmatrix} \mathbf{e} \\ \mathbf{h} \end{Bmatrix} = \begin{Bmatrix} \mathbf{E} \\ \mathbf{H} \end{Bmatrix} \exp(j\omega t) \exp(-j\mathbf{k}_0 \cdot \mathbf{r}) \tag{1.4.1}$$

where

$$\begin{aligned} \mathbf{k}_0 \cdot \mathbf{r} &= (k_x\mathbf{a}_x + k_y\mathbf{a}_y + k_z\mathbf{a}_z) \cdot (x\mathbf{a}_x + y\mathbf{a}_y + z\mathbf{a}_z) \\ &= k_x x + k_y y + k_z z \end{aligned}$$

In (1.4.1), $\mathbf{E}$ and $\mathbf{H}$ are not functions of x, y, or z. Some fortitude and patience with the rules for the curl operation yield

$$\begin{aligned} \mathbf{k}_0 \times \mathbf{E} &= +\omega\mu_0\mathbf{H} \\ \mathbf{k}_0 \times \mathbf{H} &= -\omega\epsilon_0\mathbf{E} \end{aligned} \tag{1.4.2}$$

The main utility of (1.4.2) lies in the geometric interpretation. For instance, $\mathbf{H}$ is obviously perpendicular to both $\mathbf{k}_0$ and $\mathbf{E}$ from the first one, and $\mathbf{E}$ is also perpendicular to $\mathbf{H}$ and $\mathbf{k}_0$ by the second. This geometric arrangement is shown in Fig. 1.1. The vectors $\mathbf{E}$ and $\mathbf{H}$ are related to each other (and are obviously in phase, since j is absent from the equations).

$$\frac{|\mathbf{E}|}{|\mathbf{H}|} = \frac{\omega\mu_0}{|\mathbf{k}_0|} = \frac{|\mathbf{k}_0|}{\omega\epsilon_0} = \eta_0 = \left(\frac{\mu_0}{\epsilon_0}\right)^{1/2} \cong 377\ \Omega \tag{1.4.3}$$

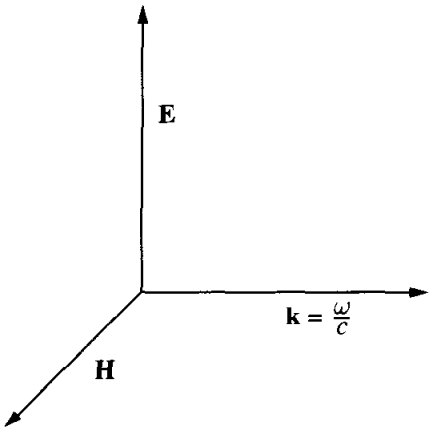


FIGURE 1.1. Geometric orientation of the vectors **E**, **H**, and **k** according to (1.4.2).

where: $\eta_0 = \left(\frac{\mu_0}{\epsilon_0}\right)^{1/2}$ = wave impedance of free space

For free space, the Poynting vector $\mathbf{S} = \frac{1}{2}\mathbf{E} \times \mathbf{H}^*$ points in the same direction as $\mathbf{k}_0$.

$$\mathbf{S} = \frac{1}{2}\mathbf{E} \times \mathbf{H}^* = \frac{1}{2}\mathbf{E} \times \frac{(\mathbf{k}_0 \times \mathbf{E})^*}{\omega\mu_0} = \frac{1}{2}\mathbf{E} \cdot \mathbf{E}^* \frac{\mathbf{k}_0}{\omega\mu_0} \tag{1.4.4}$$

since $\mathbf{k}_0 \cdot \mathbf{E} = 0$.

1.5 WAVES IN DIELECTRICS

Let us examine (1.2.1) in more detail for cases that are commonly encountered in solid-state lasers or electro-optic materials. For instance, the active atoms in a ruby laser are chromium, which is added (doped) to a level of 5% into the Al_2O_3 host crystal, the details of which are covered in Chapter 10. The point to be made here is that both Al_2O_3 lattice and active atoms contribute to the polarization, and that it is useful to separate their effects on the propagation of waves. Accordingly, we rewrite Maxwell's equations (with $\mathbf{j} = 0$):

$$\nabla \times \mathbf{h} = \epsilon_0 \frac{\partial \mathbf{e}}{\partial t} + \frac{\partial \mathbf{p}_l}{\partial t} + \frac{\partial \mathbf{p}_a}{\partial t} = \epsilon_0 n^2 \frac{\partial \mathbf{e}}{\partial t} + \frac{\partial \mathbf{p}_a}{\partial t} \tag{1.5.1a}$$

$$\nabla \times \mathbf{e} = -\mu_0 \frac{\partial \mathbf{h}}{\partial t} \tag{1.5.1b}$$

where we have combined the lattice polarization term $\mathbf{p}_l$ with the vacuum displacement $\epsilon_0\mathbf{e}$ with the aid of (1.2.5) to obtain the term involving the square of the index of refraction n^2. Now we repeat the mathematics used to derive the homogeneous wave equation of Sec. 1.3: take the curl of (1.5.1b)

$$\nabla \times \nabla \times \mathbf{e} = -\mu_0 \frac{\partial[\nabla \times \mathbf{h}]}{\partial t}$$

$$\nabla(\nabla \cdot \mathbf{e}) - \nabla^2\mathbf{e} = -\mu_0\epsilon_0 n^2 \frac{\partial^2 \mathbf{e}}{\partial t^2} - \mu_0 \frac{\partial^2 \mathbf{p}_a}{\partial t^2}$$

$$\nabla^2\mathbf{e} - \left(\frac{n}{c}\right)^2 \frac{\partial^2\mathbf{e}}{\partial t^2} = \mu_0 \frac{\partial^2\mathbf{p}_a}{\partial t^2} \tag{1.5.2}$$

This differs from (1.3.1) in two ways: (1) the velocity of the propagation is c/n, a result which could be anticipated from elementary electromagnetic theory; and (2) the right-hand side is no longer zero and that the equation is now an inhomogeneous one, with the source (i.e., the right-hand side) being the time derivative of the polarization contributed by the active atoms. In other words, the active atoms are the *source* for the optical *fields*. This is the proper approach, but it is also somewhat tedious and thus will be postponed until Chapters 13 and 14.

1.6 THE UNCERTAINTY RELATIONSHIPS

It was mentioned in Sec. 1.2 that the representation of a sinusoidal function by the real part of a complex phasor is a shorthand way of taking the Fourier transform of the time function. This is very important from many standpoints. First, it is the formal way of handling nonsinusoidal functions and paves the way for a general transient analysis. But most of all, it leads most naturally into the concept of minimum beam spread from a given aperture.

To appreciate this, let us recall the "uncertainty" relation as it pertains to communication:

$$\Delta\omega\Delta t \geq \tfrac{1}{2} \tag{1.6.1}$$

In communications, this theorem says that a minimum bandwidth $\Delta\omega$ is required to pass a pulse with a rise time Δt. If we multiply both sides of the equation by $\hbar = h/2\pi$, we obtain formally a relation equivalent to the Heisenberg* uncertainty principle:

$$\Delta E\Delta t \geq \frac{h}{4\pi} \tag{1.6.2}$$

It is not a very interesting exercise in transform theory to prove that any two conjugate variables (such as ω and t), which are related by the Fourier transform, obey (1.6.1). The genius of Heisenberg was in relating a physical problem to a mathematical abstraction.

Let us now turn to other conjugate variables. For instance, k_x is the Fourier transform variable with its conjugate x, k_y with y, and k_z with z. Once (1.6.1) is accepted, the same theory of Fourier transforms yields

$$\begin{aligned} \Delta k_x\Delta x &\geq \tfrac{1}{2} \\ \Delta k_y\Delta y &\geq \tfrac{1}{2} \\ \Delta k_z\Delta z &\geq \tfrac{1}{2} \end{aligned} \tag{1.6.3}$$

If we again multiply $\hbar = h/2\pi$ and identify $\hbar\mathbf{k}$ as the momentum, we obtain the conventional form of Heisenberg's uncertainty relations. These relationships are summarized in

*Whether the factor in (1.6.1) should be 1, $\frac{1}{2}$, or some other number close to 1 depends on how $\Delta\omega$ and Δt are defined.

TABLE 1.1

Item	Physical	Conjugate variable	Relation
ω	Angular frequency	t (time)	$\Delta\omega\Delta t \geq \frac{1}{2}$
k_x	Propagation along x	x	$\Delta k_x \Delta x \geq \frac{1}{2}$
k_y	Propagation along y	y	$\Delta k_y \Delta y \geq \frac{1}{2}$
k_z	Propagation along z	z	$\Delta k_z \Delta z \geq \frac{1}{2}$
E	$\hbar\omega$ = energy	t	$\Delta E \Delta t \geq h/4\pi$
p_x	Momentum along x	x	$\Delta p_x \Delta x \geq h/4\pi$
p_y	Momentum along y	y	$\Delta p_y \Delta y \geq h/4\pi$
p_z	Momentum along z	z	$\Delta p_z \Delta z \geq h/4\pi$

Table 1.1 Note that the uncertainty principle says nothing whatsoever about the relation between nonconjugate variables.

Before we leave this topic, it is worthwhile to have a more precise definition of the term "uncertainty": it is the rms value of the deviation of the parameter from its average value. For instance, if the transverse variation of the electric field of an optical beam were given by

$$E(y) = E_0 \exp\left[-\left(\frac{y}{w_0}\right)^2\right] \tag{1.6.4}$$

then the average location of the field is at $y = 0$ and the "uncertainty" Δy is found from

$$(\Delta y)^2 = \frac{\int_{-\infty}^{+\infty} (y - 0)^2 E^2(y) dy}{\int_{-\infty}^{+\infty} E^2(y) dy} \tag{1.6.5}$$

In other words, the mathematical formula for the field can also be interpreted as a probability function. The Fourier transform (in k_y space) is given by

$$E(k_y) = \pi^{1/2} w_0 E_0 \exp\left[-\left(\frac{k_y w_0}{2}\right)^2\right] \tag{1.6.6}$$

Thus there is a distribution of k_y wave vectors around $k_y = 0$ and thus the "uncertainty" of k_y is

$$(\Delta k_y)^2 = \frac{\int_{-\infty}^{+\infty} (k_y - 0)^2 E^2(k_y) dk_y}{\int_{-\infty}^{+\infty} E^2(k_y) dk_y} \tag{1.6.7}$$

It is left for a problem to show that this particular field distribution has the minimum value permitted: $\Delta y \cdot \Delta k_y = 1/2$.

1.7 SPREADING OF AN ELECTROMAGNETIC BEAM

Let us use the uncertainty relationships to predict the spread of a beam of light energy. Now we know that this beam is traveling more or less at the velocity of light, c; hence, the wave vector k_z is very well defined at $k_z = \omega/c$ (and, sure enough, the beam is almost everywhere along the z axis). But if this is a "beam," its extent in the transverse dimension is limited to the beam diameter, as shown in Fig. 1.2.

If we assume that this "beam" has a smooth "Gaussian-like" spatial extent in the y direction of the form given by (1.6.4)

$$E(y) = E_0 \exp\left[\left(-\frac{y}{w_0}\right)^2\right]$$

then we must also allow for a spread in wave vectors centered around $k_y = 0$:

$$E(k_y) = \pi^{1/2} w_0 E_0 \exp\left[-\left(\frac{k_y w_0}{2}\right)^2\right]$$

This interrelationship is sketched in Fig. 1.3 on page 16.

Although a Gaussian spatial envelope is unique in the sense that it is also a Gaussian in k space, the conclusions are the same irrespective of what is chosen for $E(y)$.

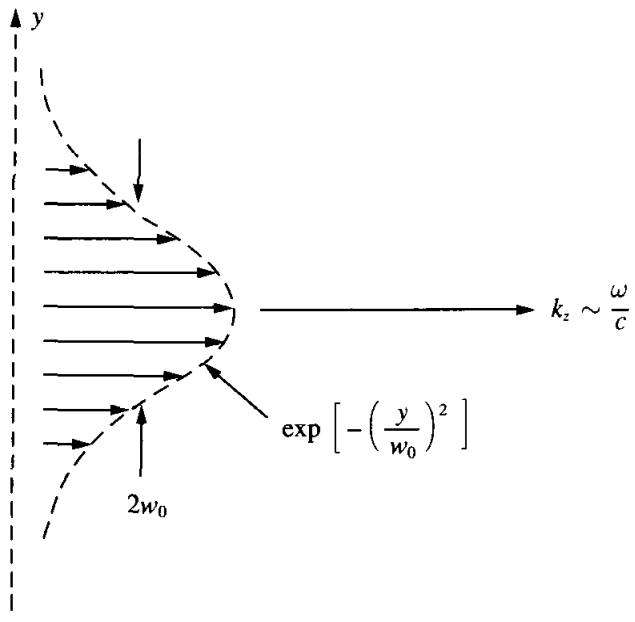


FIGURE 1.2. Beam of light diameter $2w_0$ passing the surface $z = 0$. We will use the uncertainty relations to predict the beam diameter along the propagation path.

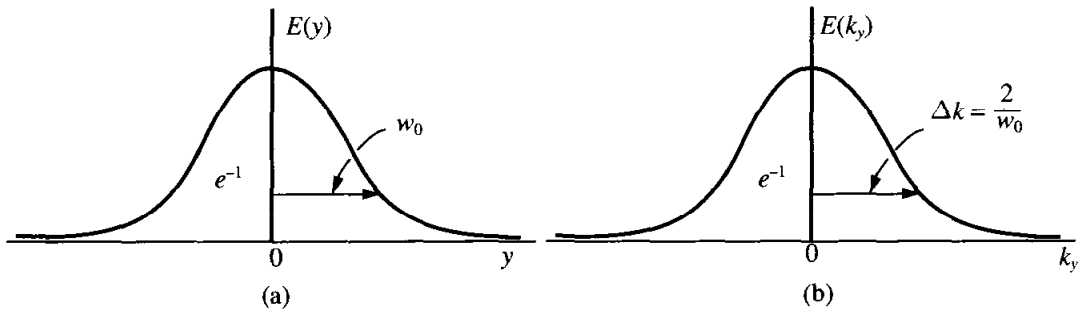


FIGURE 1.3. Interrelationship between (a) the spatial extent of a beam and (b) the wave number k_y.

Thus, we can construct a diagram for the propagation vectors k_y and k_z as shown in Fig. 1.4. It is obvious that the angle $\theta_0/2$ is given by

$$\left.\begin{aligned} \frac{\theta_0}{2} &= \frac{\Delta k_y}{k_z} = \frac{\lambda}{\pi w_0} \\ \text{or} \qquad \theta_0 &= \frac{2\lambda}{\pi w_0} \end{aligned}\right\} \tag{1.7.1}$$

Thus, a large beam does not spread. Indeed, a uniform plane wave (one with $w_0 = \infty$) has a zero spread, in accordance with every elementary text on electromagnetic theory. (It has no place to go!)

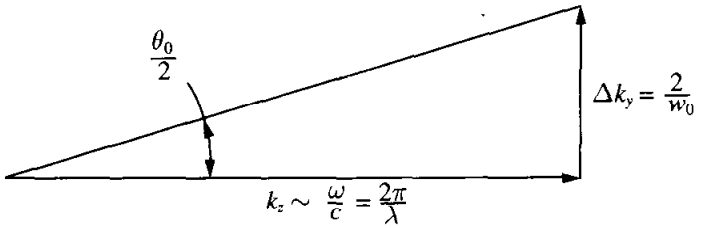


FIGURE 1.4. Vector addition of k_z and Δk_y to estimate the beam spread.

It is instructive to consider some numbers here. Let $\lambda = 694.3$ nm and $2w_0 = 0.1$ cm; then θ_0 is 8.8×10^{-4} rad. To achieve the same beam spread at 10-cm wavelength would require an antenna aperture $2w_0$ of 144 m. Such a small divergence of an optical beam justifies the simple ray-tracing approach of Chapter 2.

1.8 WAVE PROPAGATION IN ANISOTROPIC MEDIA

Materials that are anisotropic to electromagnetic waves have many uses in optical electronics: modulation, sensing, and harmonic generation are just a few examples. Indeed, most crystalline materials are anisotropic and even some of the amorphous ones, such as glass, become so when subjected to an electric field, a magnetic field, or mechanical stress. This section introduces the formalism for handling such cases.

We limit our attention to uniaxial media whose dielectric "constant" depends on the direction of the electrical field, and thus the displacement vector **D** is described by a matrix multiplication of ϵ with the electric field **E**.

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \epsilon_0 \begin{bmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_1 & 0 \\ 0 & 0 & \epsilon_2 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{E}_x \\ \mathbf{E}_y \\ \mathbf{E}_z \end{bmatrix} \tag{1.8.1}$$

Our goal is to predict the value of the wave vector **k** as the wave propagates at an angle θ with respect to the z axis (the optical axis) as shown in Fig. 1.5.

From the algebraic form of Maxwell's equations, we know that the wave vector **k** is perpendicular to **D** in any and all cases—anisotropy or no anisotropy!

$$\begin{aligned} \mathbf{k} \times \mathbf{h} &= -\omega \mathbf{D} \\ \mathbf{k} \cdot (\mathbf{k} \times \mathbf{H}) \equiv 0 &= -\omega \mathbf{k} \cdot \mathbf{D} \end{aligned} \tag{1.8.2}$$

Hence there is one orientation of the electric field where we *know* the answer for the orientation of the fields with respect to **k**. This is shown in Fig. 1.5(b), and since the case is so "ordinary," it is given that name. Note that if **k** is constrained to the yz plane, then **D** is always in the x direction, and thus $\mathbf{E} = E_x \mathbf{a}_x$. The same argument can be applied to the case where the displacement vector is perpendicular to the plane containing **k** and the z axis, the so-called optic axis. For such cases, the propagation constant is given by

$$k^2 = \omega^2 \mu_0 \epsilon_0 \epsilon_1$$

or

$$\frac{k_0^2}{k^2} = \frac{1}{\epsilon_1} = \frac{1}{n_1^2} \qquad \text{(independent of } \theta\text{)} \tag{1.8.3}$$

If, however, **D** is not perpendicular to the plane containing **k** and the optic axis (i.e., $[\mathbf{a}_z \times \mathbf{k}] \cdot \mathbf{D} = 0$) as shown in Fig. 1.5(c), we have a problem. **D** is still perpendicular to **k**, since ($\mathbf{D} \cdot \mathbf{k} = 0$), but **E** is not! Hence we can expect a mixture of ϵ_1 and ϵ_2 in the expression for the propagation constant, and a somewhat "extraordinary" behavior as a function of θ, a task to which we turn.

For this polarization shown in Fig. 1.5(c), **k** and **D** can be expressed as

$$\mathbf{k} = k(\cos\theta\, \mathbf{a}_z + \sin\theta\, \mathbf{a}_y) \tag{1.8.4a}$$

$$\mathbf{D} = D(-\cos\theta\, \mathbf{a}_y + \sin\theta\, \mathbf{a}_z) \tag{1.8.4b}$$

(Note that $\mathbf{k} \cdot \mathbf{D} \equiv 0$.)

We use (1.8.4b) in conjunction with (1.8.1) to find **E**:

$$E_y = \frac{D}{\epsilon_0 \epsilon_1}[-\cos\theta] \tag{1.8.5a}$$

$$E_z = \frac{D}{\epsilon_0 \epsilon_2}[\sin\theta] \tag{1.8.5b}$$

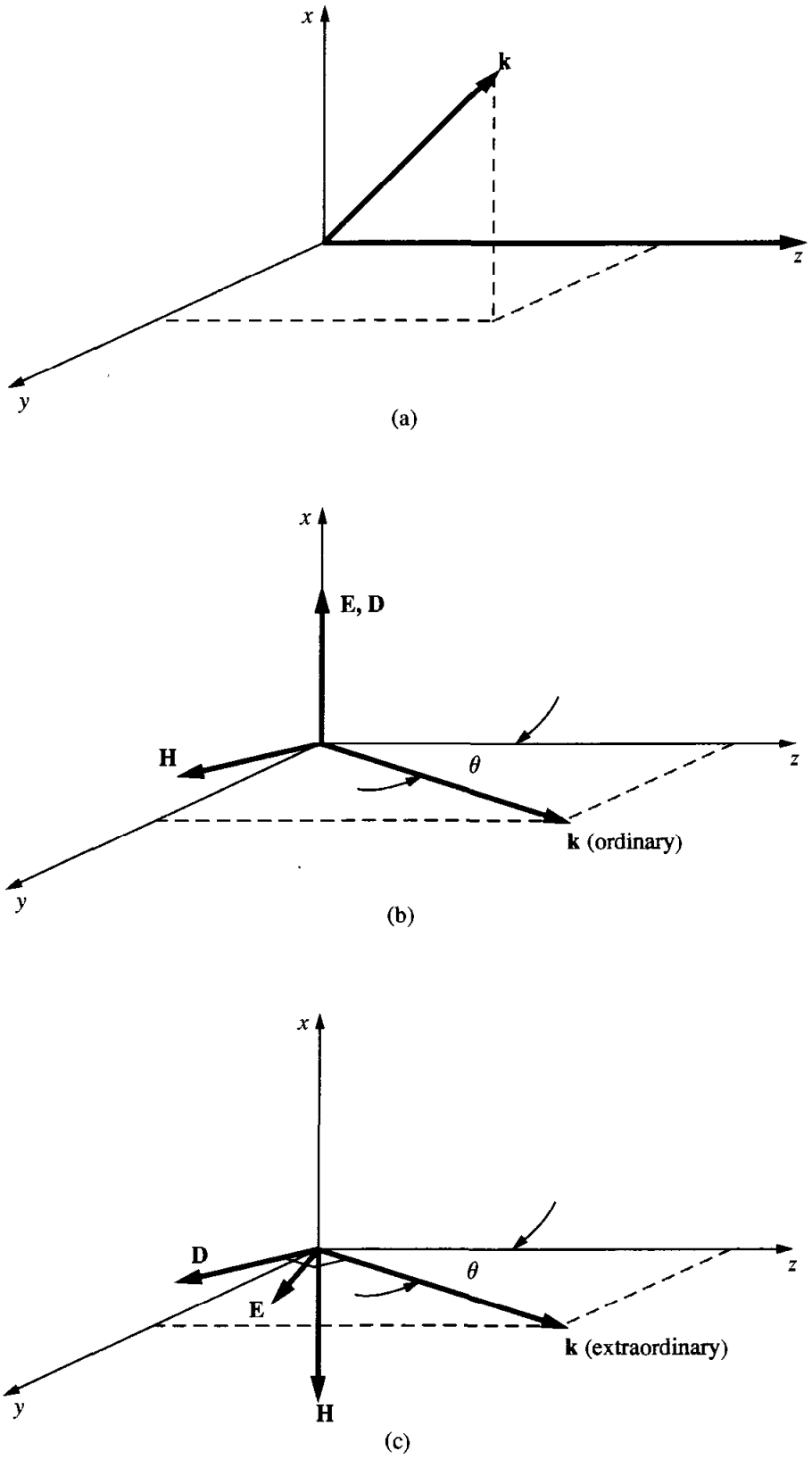


FIGURE 1.5. Orientation of **k**, **E**, and **D** for a uniaxial crystal. (a) The general problem. (b) The ordinary wave. (c) The extraordinary wave.

Now it is a straightforward exercise in vector analysis to show (see Problem 1.3) that

$$k^2 = \omega^2 \mu_0 \frac{\mathbf{D} \cdot \mathbf{D}}{\mathbf{E} \cdot \mathbf{D}} \tag{1.8.6a}$$

or

$$\left(\frac{k_0}{k}\right)^2 = \frac{1}{n_{\text{eff}}^2} = \epsilon_0 \frac{\mathbf{E} \cdot \mathbf{D}}{\mathbf{D} \cdot \mathbf{D}} \tag{1.8.6b}$$

where the effective index is defined by $k/k_0 = n_{\text{eff}}$. Combining (1.8.6b) with (1.8.5) yields

$$\frac{1}{n_{\text{eff}}^2} = \frac{\cos^2\theta}{n_1^2} + \frac{\sin^2\theta}{n_2^2} \tag{1.8.7}$$

The forms of normalized propagation vector (k/k_0) expressed by (1.8.3) and (1.8.7) are conveniently shown on a graph called the index surface (see Fig. 1.6). Equation (1.8.3) states that the effective index for the ordinary wave is independent of the angle θ. Hence it is shown as a circle. The effective index for extraordinary wave does depend on θ in the form of an ellipse.

It is apparent from Fig. 1.6 and from (1.8.3) and (1.8.7) that the phase constants for the ordinary and extraordinary waves are not equal for $\theta \neq 0$. This fact plays a critical role in nonlinear optics where it is crucial that the phase constants of, for example, the fundamental wave and any harmonic or intermodulation terms, *must* be synchronized. Fortunately, the dielectric constants are not constant with frequency (i.e., λ), and thus it is possible to choose a phase matching angle θ_m such that the effective index for the fundamental frequency ω, when propagated as an ordinary (extraordinary) wave, equals the effective index for the second (third, etc.) harmonic when it is propagated as an extraordinary (ordinary) wave.

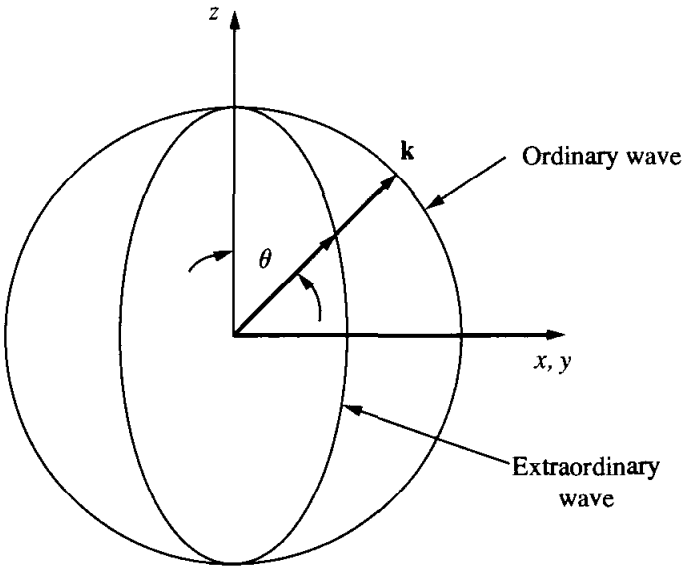


FIGURE 1.6. The index ellipsoid for a uniaxial crystal.

1.9 ELEMENTARY BOUNDARY VALUE PROBLEMS IN OPTICS

The propagation of electromagnetic waves is determined by Maxwell's equations, but these are incomplete without a specification of boundary conditions. After all, they are partial differential equations that presume that all field variables and material properties are continuous functions of the coordinates. However, we will have many occasions to consider abrupt junctions between different materials (windows, mirrors, etc.) where the electrical parameters are different, and, as a consequence, the field variables change discontinuously.

Most elementary texts derive the relationship between the tangential and normal components of the field at each side of an abrupt interface:

$$\mathbf{a}_n \times (\mathbf{E}_1 - \mathbf{E}_2) = 0 \tag{1.9.1a}$$

$$\mathbf{a}_n \cdot (\mathbf{D}_1 - \mathbf{D}_2) = \rho_{s2} \tag{1.9.1b}$$

$$\mathbf{a}_n \times (\mathbf{H}_1 - \mathbf{H}_2) = \mathbf{J}_{s2} \tag{1.9.1c}$$

$$\mathbf{a}_n \cdot (\mathbf{B}_1 - \mathbf{B}_2) = 0 \tag{1.9.1d}$$

where $\mathbf{a}_n$ is a unit vector from 2 to 1 and perpendicular to the interface. The concept of a surface charge, ρ_s, and surface current, $\mathbf{J}_s$, both existing in zero depth in medium 2, are useful approximations at low frequencies, $\nu < 10^{12}$ Hz, but those approximations are almost never utilized in the optical domain. Hence, we will let the right-hand side of (1.9.1b) and (1.9.1c) be zero.

The formal method of handling the interface problem is to first solve Maxwell's equations in the two media and then match the fields at the boundary with (1.9.1a) and (1.9.1c). It is sufficient to match tangential components only, because the normal components will then be matched automatically, provided the fields in the respective media obey Maxwell's equations.

Many times we can sidestep a lot of dull mathematics implied by what we just did by applying some elementary physical reasoning. Some very important examples of this approach are shown below.

1.9.1 Snell's Law

Consider a *u*niform *p*lane *w*ave (upw) impinging on the interface shown in Fig. 1.7 making an angle θ_1 with respect to the normal to the surface. The discontinuity generates a second wave

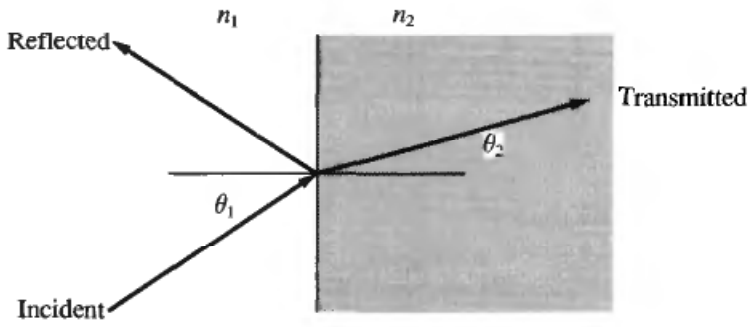


FIGURE 1.7. Geometry for Snell's Law.

at an angle θ_2 and a reflected wave. We could grit our teeth and match field components at the interface and solve the problem completely. This procedure is necessary if the amplitude and phase of the *transmitted* and *reflected* waves are desired. However, if only the *direction* is desired, the procedure can be greatly simplified.

The point to be remembered is that the incident wave is the *source*, and the transmitted and reflected waves are the *responses*. Hence the phases of both responses, whatever they are, must be synchronized with respect to the source along the boundary where the responses are generated.

The relative phase of the source along the interface is

$$\phi = (\omega/c)n_1 \sin\theta_1 \tag{1.9.2}$$

and this must be the phase of both responses as measured along the interface. If medium 1 is isotropic, this fact forces the incident and reflected waves to make the same angle with respect to the normal.

For the transmitted field, we force the phases along the boundary to be the same:

$$(\omega/c)\, n_1 \sin\theta_1 = (\omega/c)\, n_2 \sin\theta_2 \tag{1.9.3a}$$

or

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{1.9.3b}$$

For an anisotropic medium for 2, the incident wave can generate two transmitted waves, but both must remain tied to the phase of incident wave along the interface.

1.9.2 Brewster's Angle

Windows oriented at Brewster's angle are commonly used on gas lasers because, in principle, they transmit waves *without* reflection for one polarization of the electric field. The geometry of the electromagnetic problem is shown in Fig. 1.8 for two possible polarizations of the incident field. In both cases, Snell's law is applicable, and thus the wave vector $\mathbf{k}$ is bent toward the normal in the window material.

There are some artifacts added to Fig. 1.8 to help visualize the physical situation: The orientations of the induced dipoles in the dielectric material are shown, for it is their reradiation that generates the reflected wave.

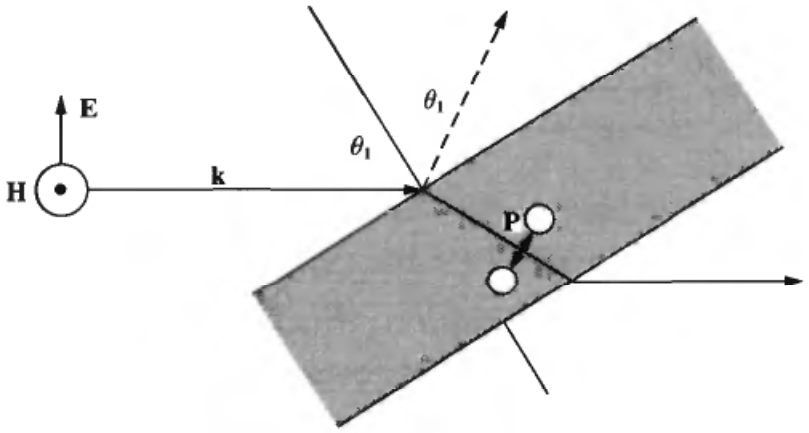
Now every elementary test in electromagnetic theory shows that electric dipoles radiate perpendicular to the axis and *not* along it. Thus for the TE orientation there is no problem in generating a reflected wave. However, for the TM case and a particular angle of the incident wave, the reflected wave would try to come off the ends of the dipole, which is impossible. Hence there is *no* reflected wave when the angles $\theta_1 + \theta_2 = \pi/2$. Combining this fact with Snell's law yields an expression for an angle of zero reflection:

$$\theta_1 + \theta_2 = \frac{\pi}{2}$$

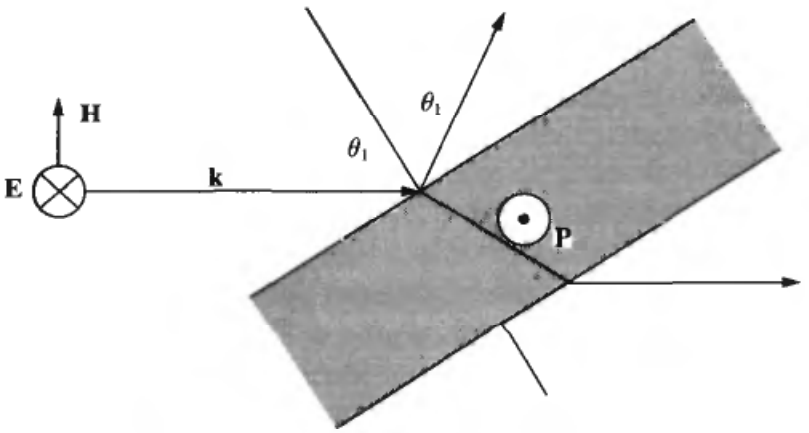
$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \qquad \text{(Snell's law)} \tag{1.9.4}$$

Hence

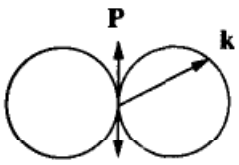
$$n_1 \sin\theta_1 = n_2 \sin(\pi/2 - \theta_1) = n_2 \cos\theta_1$$



(a) TM or "p" polarized



(b) TE or "s" polarized



(c) Dipole radiation

FIGURE 1.8. Brewster's angle windows.

Therefore $$\tan\theta_1 = \frac{n_2}{n_1} \qquad \text{(Brewster's angle)} \tag{1.9.5}$$

It should be emphasized that mathematics involved in matching fields across an interface will lead to the same result, but we should appreciate the physical reasoning just presented also.

1.10 COHERENT ELECTROMAGNETIC RADIATION

Let us reiterate the goals of this book: to understand the physical bases for the *generation*, *transmission*, and *detection* of electromagnetic radiation in the "optical" portion of the spectrum. But we should be more precise and focus our attention on a specific characteristic that distinguishes the laser from a simple lamp.

The distinguishing characteristic is the generation of *coherent* electromagnetic radiation. Now, the topic of coherence is most involved and complex to describe with precision, but it is relatively easy to understand the first-order consequences.

Most who have had electronic experience at low frequencies, say less than 30 GHz, with classical generators never address this subject, because most of our generators had a long coherence time or length. In other words, they are almost perfectly coherent. But what does this mean, and how would we measure either coherence time or length?

In a loose sort of way, coherence time is the net delay that can be inserted in a wave train and still obtain interference. Since electromagnetic waves travel with a velocity of c, the longitudinal coherence length is simply c times the coherence time. Note that the key word is *interference*. Let us illustrate these ideas with a "thought" experiment taken from low-frequency electronics and compare it with a similar experiment at optical frequencies (visible wavelengths).

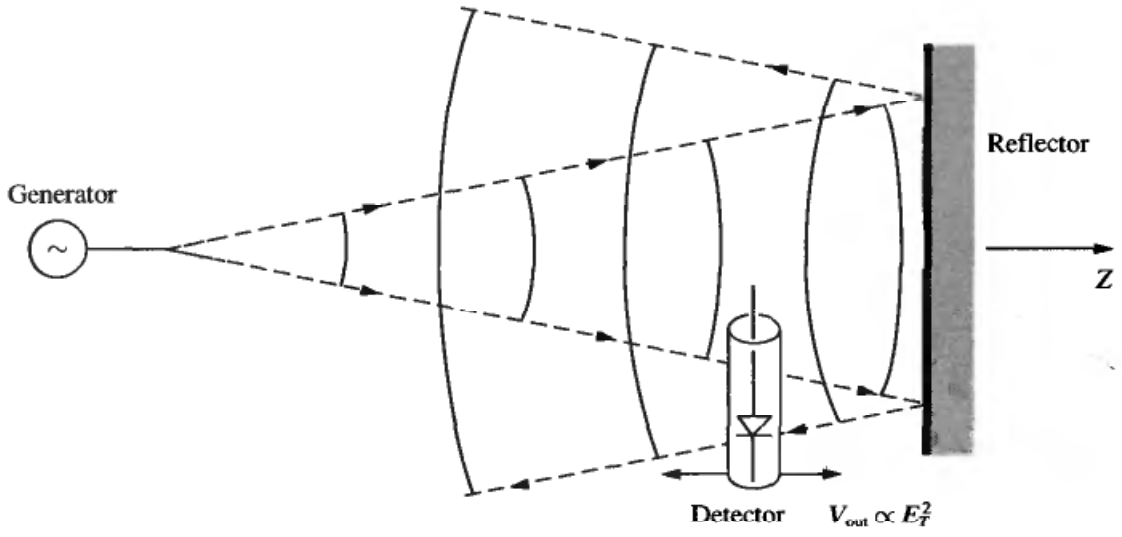


FIGURE 1.9. Simple interference experiment.

Consider a simple transmission-line measurement of the standing-wave ratio on a short-circuited transmission system as shown in Fig. 1.9. To make the conventional "slotted-line" measurement of the "voltage" standing-wave ratio (VSWR), we move a short dipole antenna and a rectifying diode along the z axis. The output of the detector is proportional to the square of the electric field (usually); hence, the relative output of the detector would be as shown in Fig. 1.10. The VSWR, $V_{\max}/V_{\min}$, is very large, and for all practical purposes it is infinite. This is precisely what we observe in a normal laboratory.* Even elementary theory would predict this result, as is demonstrated next.

The electric field traveling to the right is given by

$$E^+ = E_0 \exp(-jkz) \qquad z < 0 \qquad k = \frac{2\pi}{\lambda} \tag{1.10.1}$$

with the time factor, $\exp(j\omega t)$, suppressed. The reflected wave is given by

$$E^- = -E_0 \exp(+jkz) \tag{1.10.2}$$

Hence, the output of the detector is given by

$$V_{\text{out}} \propto E_T E_T^* = 4E_0^2 \sin^2 kz \qquad (E_T = E^+ + E^-) \tag{1.10.3}$$

Although this analysis is quite adequate for normal laboratory experiments at low frequencies, we have made the serious assumption of a perfectly coherent source. Such a device does not exist.

We have assumed that the phase of the incoming wave at a point z is predictable from the phase of the wave that crossed this point at a time $2z/c$ seconds earlier. But, of course, it is not tied perfectly to this earlier waveform; its phase could have "wandered" in the time it took the initial wave to traverse the distance from the observation point to the reflector and back. Thus, we should modify (1.10.1) to read

$$E^+ = E_0 \exp\left\{-j\left[kz + \Delta\phi(t)\right]\right\} \tag{1.10.1a}$$

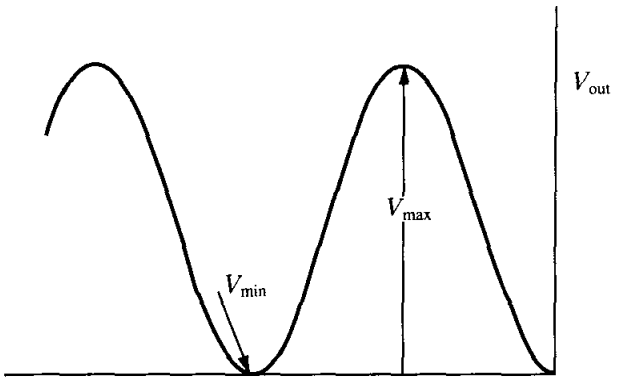


FIGURE 1.10. Measurements of the VSWR. (NOTE: Most detectors produce an output [i.e., voltage] proportional to the power sampled by the antenna. Consequently, the quantity $V_{\max}/V_{\min}$ would correspond to the power standing-wave ratio.)

*In fact, Fig. 1.9 bears a close resemblance to the original experiments of Hertz, who demonstrated the equivalence of light and low-frequency waves as predicted by Maxwell's theory.

where $\Delta\phi(t)$ is a random variable, characteristic of the source.* Thus the output of the detector changes to

$$V_{\text{out}} \propto E_T^2 = 4E_0^2 \sin^2\left[kz + \frac{\Delta\phi(t)}{2}\right] \tag{1.10.4}$$

In this case, the minimum (or maximum) is *not* where we think it should be, and worse yet, it wanders in time according to the whims of $\phi(t)$. It is almost as if the standing-wave pattern is "jittering" back and forth in a random fashion, as indicated in Fig. 1.11. Normally, the time rate of change of ϕ is small when compared with the angular frequency ω, and this fact explains why we never see this effect at low frequencies from any "decent" source.

Example

Suppose that the maximum value of $d\phi/dt$ was 10^{-4} of the angular frequency ω_0 of the source (a rather poor one, but let us use it). Let the nominal frequency of the source be 1 GHz. If the observation point z of our detector were a "room-like" distance away from the reflector, say 3 m, the time interval between the passage of the first wave train and its return is only

$$\Delta t = \frac{2z}{c} = \frac{2 \times 3}{3 \times 10^8} = 20\text{ns}$$

and the phase could, at most change by

$$\Delta\phi = \left.\frac{d\phi}{dt}\right|_{\max} \Delta t = 10^{-4} \times 2\pi \times 10^{+9} \times 20 \times 10^{-9} = 0.004\pi \Rightarrow 0.72^\circ$$

In other words, the position of the minimum is only jittering by $0.72^\circ/360^\circ = 0.2\%$ of a wavelength (30 cm) or $\Delta L = 0.6$ mm (probably smaller than the wire used for the dipole antenna).

However, the numbers and the effects change considerably if we perform the same type of interference experiment at optical frequencies. Since most components and detectors are huge when compared with optical wavelengths, the techniques are slightly different but not in their essential function.

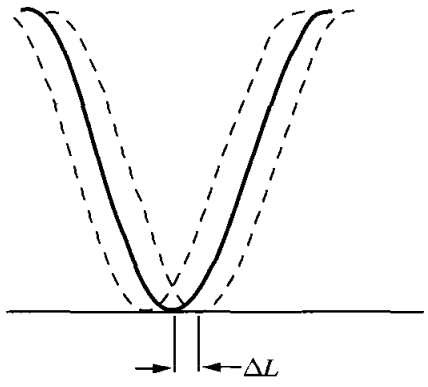


FIGURE 1.11. "Jittering" of the minimum position owing to the random jumps in phase of the later portion of the wave.

*$\Delta\phi$ is the amount by which the phase can change in the round trip delay time $2z/c$.

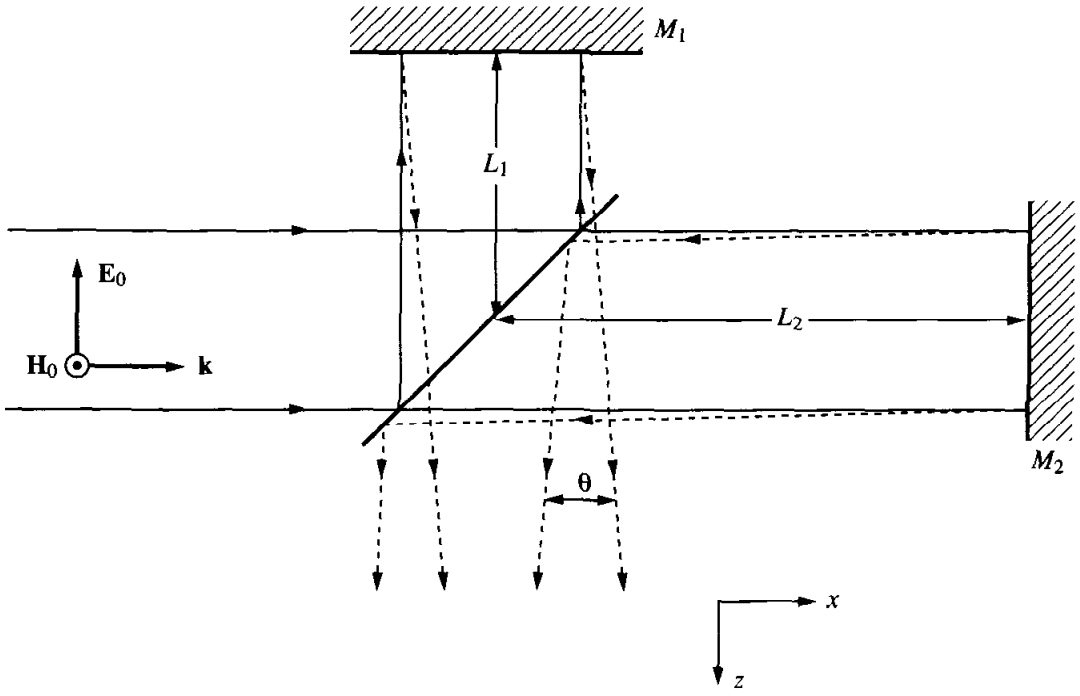


FIGURE 1.12. Michelson interferometer.

Consider the Michelson interferometer shown in Fig. 1.12. Collimated light is divided and passed around the two arms of the interferometer in the manner indicated in Fig. 1.12. Obviously, the radiation that went the M_2 route is retarded in time by $2(L_2 - L_1)/c$ with respect to that returning from M_1. Shown also is the probable situation of the two beams propagating at a slight angle with respect to each other. Thus the respective electric fields at the plane of the detector are given by

$$E_1 = \frac{E_0}{\sqrt{2}} \exp\left[-j\left(k\cos\frac{\theta}{2}z + k\sin\frac{\theta}{2}x\right)\right] \exp(-j2kL_1) \exp(-j\Delta\phi)$$
$$E_2 = \frac{E_0}{\sqrt{2}} \exp\left[-j\left(k\cos\frac{\theta}{2}z - k\sin\frac{\theta}{2}x\right)\right] \exp(-j2kL_2) \tag{1.10.5}$$

In (1.10.5), we have allowed for the phase of one wave to wander with respect to the other. Again, we must remember that $\Delta\phi(t)$ is the change in phase of the reference during the time that the other signal is delayed. If the two arms of the interferometer are exactly equal, the phase term disappears. Optical detectors (eyes, photographic emulsions, and photoelectric devices) respond to the intensity of the radiation; hence, the relative intensity along the x direction for a fixed z is given by

$$I(x, y = \text{constant}) = \frac{(\mathbf{E}_1 + \mathbf{E}_2) \cdot (\mathbf{E}_1^* + \mathbf{E}_2^*)}{2\eta_0}$$
$$= 2\left(\frac{E_0^2}{2\eta_0}\right)\cos^2\left[\frac{k\theta x}{2} + k(L_2 - L_1) + \frac{\Delta\phi}{2}\right] \tag{1.10.6}$$

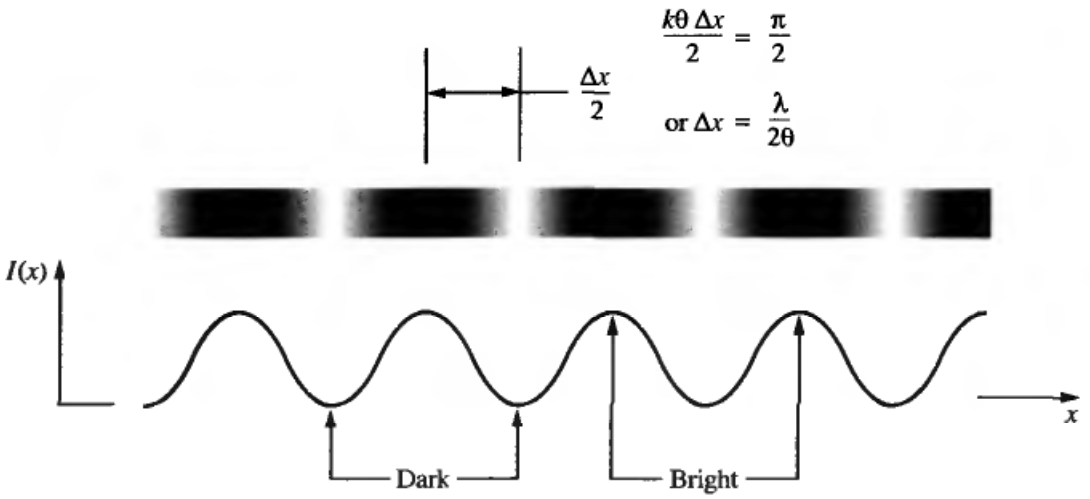


FIGURE 1.13. Interference fringes formed with the Michelson interferometer.

where we have assumed that θ is small. Thus the detector—say our eyes—would see a series of bright and dark bands in the manner indicated in Fig. 1.13.

As indicated in Fig. 1.13, θ must be very small for Δx to be a reasonable size. For instance, if $\lambda = 0.5\ \mu$m (green-blue) and Δx is to be greater than 0.5 cm, then $2\theta \leq 10^{-4}$ rad. Note, too, that the position of the minimum in intensity depends on the difference in optical path length. Hence, if this difference changes a fraction of a wavelength (owing to room vibrations), the fringe position will jitter accordingly.

Even if these significant mechanical stability problems are overcome and the alignment is achieved, the fringe position will still change, owing to the random wandering of $\phi(t)$. Whereas in the microwave "experiment," we could conceive of measuring the jitter in the fringes, our optical detector would average the intensity over its time constant. [For instance, the eye retentivity (or time constant) is on the order of 1/10 sec.] This leads to a degradation of the fringe visibility, defined as

$$V = \frac{\langle I_{\max}\rangle - \langle I_{\min}\rangle}{\langle I_{\max}\rangle + \langle I_{\min}\rangle} \tag{1.10.7}$$

where the brackets $\langle\ \rangle$ indicate time averages.

Again, let us take some typical numbers for a spectral line. Let us assume that $d\phi/dt$ is at most 10^{-5} of $\omega = 2\pi c/\lambda$ and solve for the difference in lengths that keeps the fringe position within $\Delta x/2$ of its predicted position (an interchange between the "bright" and "dark" bands). Thus $\Delta\phi/2 = \pi/2$ radians.

$$\Delta\phi = \left.\frac{d\phi}{dt}\right|_{\max} \Delta t = 10^{-5} \times \frac{2\pi c}{\lambda}\,\frac{2(L_2 - L_1)}{c} = \pi$$

or

$$L_2 - L_1 = 10^5 \times \frac{\lambda}{4}$$

For a wavelength of 5000 Å, this yields a coherence length, $2(L_2 - L_1)$, of 2.5 cm (1 in.).

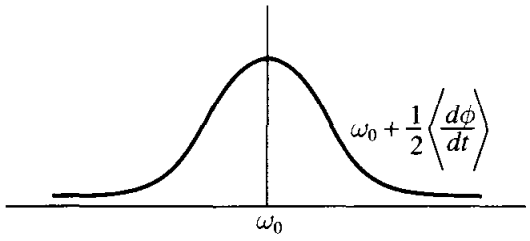


FIGURE 1.14. Profile of an emission line.

Another way of looking at this phenomenon and, at the same time, gaining some insight into the origin of the changes in the phase is to change to the spectral representation of the field. The instantaneous frequency of the source is given by the time rate of change of the phase:

$$\begin{aligned} \omega &= \frac{d}{dt}\left[\omega_0 t + \phi(t)\right] \\ &= \omega_0 + \frac{d\phi}{dt} \end{aligned} \tag{1.10.8}$$

If we consider the source as being a collection of identical oscillators being turned on and off randomly,* each contributing a part of the field, then the distribution of frequencies will be centered about ω_0 but containing a smaller amplitude on either side, as shown in Fig. 1.14. This will be discussed in more detail when the line shape is considered.

1.11 EXAMPLE OF COHERENCE EFFECTS

Let us consider a problem that emphasizes the role of the random jumps in phase owing to the finite spectral width of the source. Assume that we are measuring the interference between the two waves by means of a camera and a photographic film, as shown in Fig. 1.15. We can choose the exposure time by the shutter and thus control the "density" of the blackening of the film. This density is proportional to the time-integrated optical power density incident on a point of the film.

We assume that the two waves have the same amplitude, polarization, and nominal frequency, ω_0. However, we allow the phase of 1 to jump discontinuously by $\Delta\phi$ every T seconds according to the following random prescription.

Use three different denominations of coins, say, a penny, a nickel, and a dime. Flip the coins every T seconds, and assign the jump in phase according to Table 1.2. Now let us compute the fringe visibility by determining the density of the exposure of the film assuming that the shutter is open for $t < T$ sec and $t = NT$ sec. In the short-exposure case, the integrated exposure yields perfect fringe contrast according to (1.10.6). But for a

*As we shall see, the atoms are the oscillators. In a laser, the field stimulates the atoms to give up their energy in a predictable manner, and thus the coherence time is much larger.

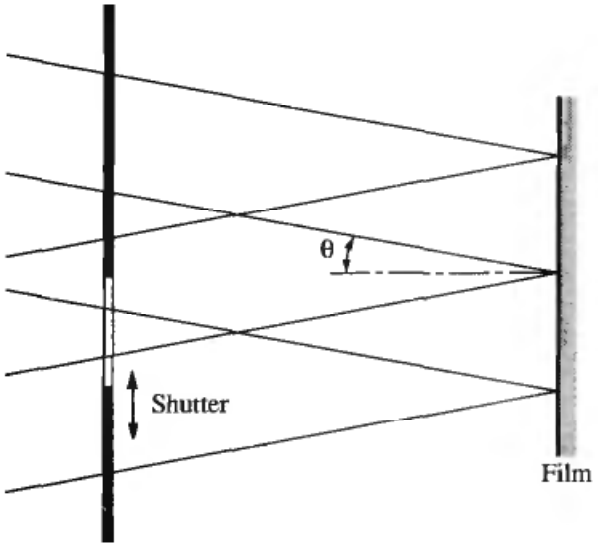


FIGURE 1.15. Paper experiment to demonstrate coherence.

long-exposure time, we have

$$D_N(x) = \frac{1}{NT}\left[I_{\max}T\cos^2\left(\frac{k\theta x + \phi_0}{2}\right)\right.$$
$$+ I_{\max}T\cos^2\left(\frac{k\theta x + \phi_0 + \phi_1}{2}\right)$$
$$+ I_{\max}T\cos^2\left(\frac{k\theta x + \phi_0 + \phi_1 + \phi_2}{2}\right) + \cdots$$
$$\left. + I_{\max}T\cos^2\left(\frac{k\theta x + \phi_0 + \phi_1 + \cdots + \phi_{N-1}}{2}\right)\right]$$

TABLE 1.2

Penny	Nickel	Dime	Jump in phase $\Delta\phi$
H	H	H	0
H	H	T	+45°
H	T	H	+90°
T	H	H	+135°
T	T	T	+180°
T	T	H	−45°
T	H	T	−90°
H	T	T	−135°

Obviously the results depend on the toss of the coin. The author obtained the results given in Table 1.3 on page 30 (you should generate your own sequence). A plot of the relative film density for various values of N is shown in Fig. 1.16. If we are able to photograph the fringes in a time interval short compared to the phase change, very sharp fringes result. However, if the shutter is open for many phase changes, the visibility decreases more or

TABLE 1.3 Results of the Toss of Coins

Coin Toss	1	2	3	4	5	6	7	8
Penny	H	H	T	H	T	T	T	H
Nickel	T	T	H	H	H	T	T	H
Dime	H	H	H	T	T	H	T	H
$\Delta\phi$ =	+90	+90	+135	+45	−90°	−45°	180°	0
$\phi_0 = 0, \phi =$	+90	180	−45	0	−90°	−135°	+45°	45°

less exponentially with exposure time. For very long periods, the film would be exposed more or less uniformly.

With this loose understanding of longitudinal coherence, let us turn to the problem of "transverse" coherence length. Here we inquire whether the phase changes along a transverse coordinate in a smooth and slow fashion, as wave 1 of Fig. 1.17, or rapidly in space and time and in an unpredictable fashion, as wave 2. We could have the same power in the two waves, and as far as visual observation is concerned, the beam size at $z = 0$ could be the same, but there would be a remarkable difference in the divergence of the beam in the two cases. This is easily shown by applying the previous discussion of the uncertainty principle considering the fundamentals of wave propagation.

If the phase is to be changing in the x direction, then there must be a component of the wave vector in that direction, even though the wave is traveling primarily in the z direction. The magnitude of this phase change can be estimated by using the mean change in phase $\overline{\Delta\phi}$ divided by the distance over which this occurs:

$$\Delta k_x \sim \frac{\overline{\Delta\phi}}{L} \tag{1.11.1}$$

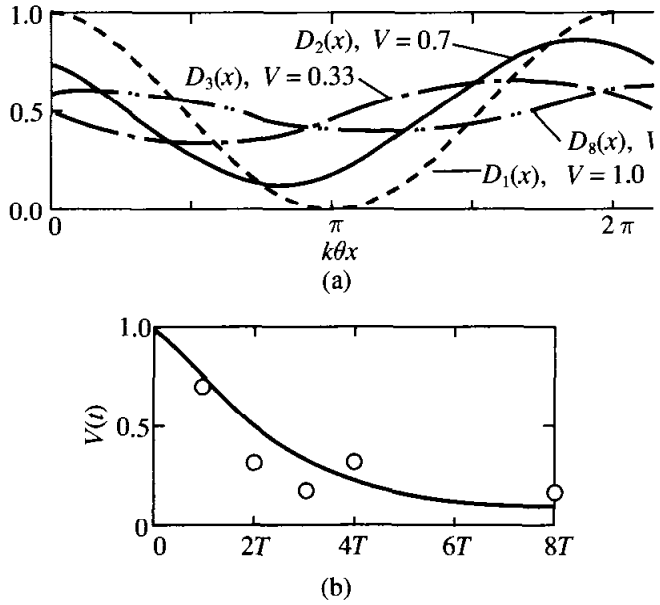


FIGURE 1.16. Relative density (D) or blackening of the film is shown in (a) along with fringe visibility V. The latter is plotted in (b) as a function of exposure time.

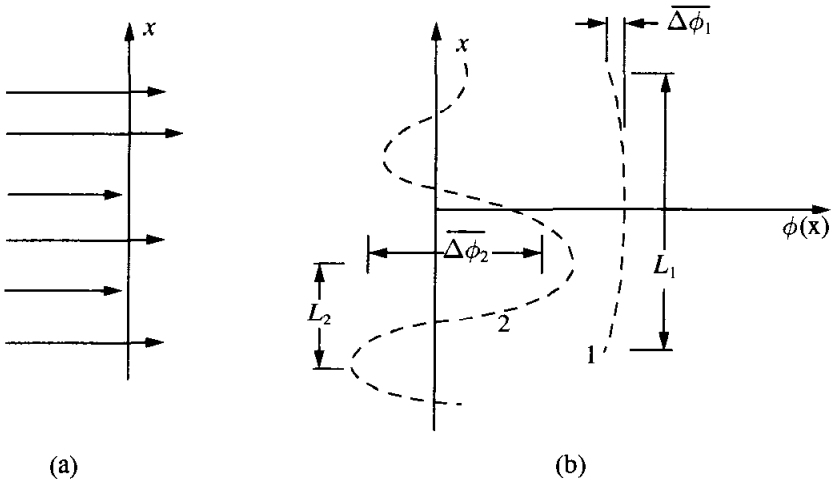


FIGURE 1.17. Two beams of the same size but with radically different variations of phase in the transverse direction.

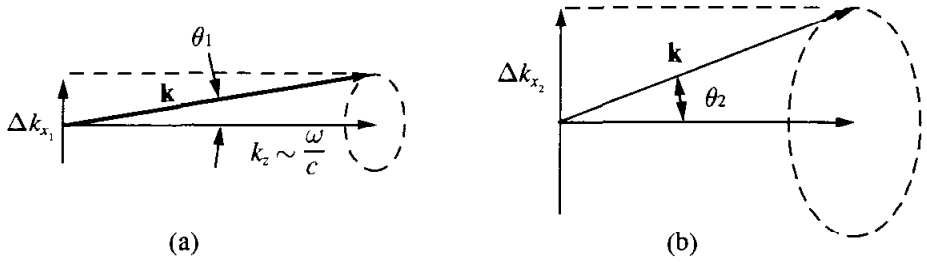


FIGURE 1.18. Beam spreads for the two beams of Fig. 1.17.

Thus for the two waves shown in Fig. 1.17, $\Delta k_{x1} = \overline{\Delta\phi_1}/L_1$ and $\Delta k_{x2} = \overline{\Delta\phi_2}/L_2$. Hence, the diagrams for the transverse and longitudinal wave vectors are as shown in Fig. 1.18. Thus the far-field divergence angle is given by

$$\theta \sim \frac{\Delta k_x}{k_z} \qquad \theta_1 \sim \frac{\overline{\Delta\phi_1}}{L_1}\frac{\lambda}{2\pi} \qquad \theta_2 = \frac{\overline{\Delta\phi_2}}{L_2}\frac{\lambda}{2\pi} \tag{1.11.2}$$

Obviously, $\theta_2 \gg \theta_1$, for the situation shown in Fig. 1.17 and the divergence of the second beam is much greater than the first. Incidentally, (1.11.2) is much more restrictive than one that is usually specified in terms of the amplitude of the wave [i.e., (1.7.1)].

PROBLEMS

1.1. Why is the factor 2 present in the expression for Poynting vector (i.e., $\mathbf{S} = \mathbf{E} \times \mathbf{H}^*/2$)?

1.2. Assume that the electromagnetic fields vary as $\exp(-j\mathbf{k} \cdot \mathbf{r})$ and use the rules for the curl, gradient, and divergence to derive the algebraic form of Maxwell's equations (1.4.2).

1.3. The algebraic forms for Maxwell's equations for a linear homogeneous anisotropic medium are

$$\mathbf{k} \times \mathbf{H} = -\omega\mathbf{D}$$

$$\mathbf{k} \times \mathbf{E} = \omega \mathbf{B}$$

where **B** is related to **H** and **D** to **E** by

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$$

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}$$

For many materials, the polarization vector **P** is not collinear with **E**; hence, **D** is not collinear with **E** either. The same comments apply to **B**, **M**, and **H**. Assume a dielectric medium with $\mathbf{M} = 0$ but with no restrictions placed on **D** and **E**.

(a) Show that $\mathbf{k} \cdot \mathbf{D} \equiv 0$.

(b) Show that the wave vector **k** always points in the direction of $\mathbf{D} \times \mathbf{B}$.

(c) Show that the amplitude of the wave vector **k** is given by

$$\mathbf{k}^2 = \omega^2 \mu_0 \frac{\mathbf{D} \cdot \mathbf{D}}{\mathbf{E} \cdot \mathbf{D}}$$

(d) Show that the Poynting vector, $\mathbf{S} = \mathbf{E} \times \mathbf{H}^*/2$, can point in a direction other than that of the wave vector **k**.

1.4. Suppose that we are using an optical beam of diameter D to monitor the particle content of a column of gas. For many applications we would prefer to sample as small a volume as possible, and consequently we would first choose a very small beam. But if the path length is long, a very small beam would diverge quickly and thus sample a larger cross-sectional area of the gas column. Use the uncertainty relations to derive an expression for the beam diameter to minimize the volume of gas sampled. Assume a helium/neon probing laser ($\lambda = 632.8$ nm) and a simple cone describing the convergence and divergence of the beam envelope so as to evaluate for a gas column 10 m long.

1.5. There are various ways to specify the frequency and photon energy of a laser; some use the energy in eV; some specify the wavelength in angstrom units (10^{-10} m), in nanometers, (10^{-9} m), or in micrometers (10^{-6} m); others use the wave number $\bar{\nu}$, the number of wavelengths that will fit inside a centimeter of vacuum; and still others specify cycles (Hz). Convert the specification of photons from common coherent sources to the other units.

Source	eV	λ (Å)	λ (nm)	ν(Hz)	$\bar{\nu}$ (cm^{-1})
GaAs	1.47				
Ar^+		5145			
He:Ne			632.8		
CO_2					943
ISM band		(in meters)		13.56 MHz	
KrF			249		

1.6. Repeat the coin-flipping routine of Sec. 1.11, and find the fringe visibility as a function of exposure time.

1.7. Show that the Fourier transform of the field given by

$$E(y) = E_0 \exp\left[-\left(\frac{y}{w_0}\right)^2\right]$$

is

$$E(k_y) = \pi^{1/2} w_0 E_0 \exp\left[-\left(\frac{k_y w_0}{2}\right)^2\right]$$

1.8. The $TEM_{0,0}$ Gaussian beam has the smallest value of the product $\Delta x \Delta k_x = 1/2$ allowed by the uncertainty relationship. (The meaning of the terminology $TEM_{0,0}$ will be covered in Chapter 3.) The quantities Δx and Δk_x are to be interpreted as

$$\Delta x^2 = \int x^2 |E(x)|^2 dx \Big/ \int |E(x)|^2 dx$$

$$\Delta k_x^2 = \int k_x^2 |E(k_x)|^2 dk_x \Big/ \int |E(k_x)|^2 dx$$

with $E(x)$ and $E(k)_x$) being related by the Fourier transform.

(a) What are the values for Δx and Δk_x for $E(x) = E_0 \exp[-(x/w_0)^2]$ (i.e., $TEM_{0,0}$)?

(b) What is the uncertainty product for a field given by

$$E_{10} = (\sqrt{2}x/w) \exp[-(x^2 + y^2)/w_0^2]$$

(c) Sketch the intensity $E_{10} \cdot E_{10}^*/2\eta_0$ as a function of x.

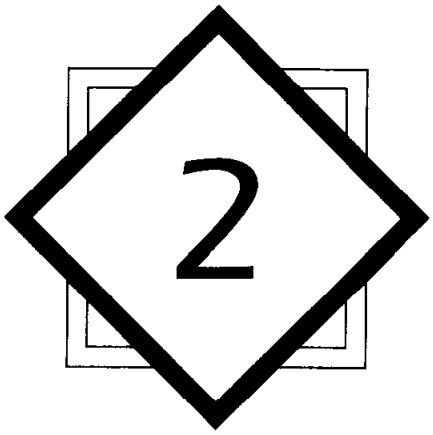
1.9. Show that the factor of 2 belongs in (1.10.4).

1.10. Quartz windows oriented at Brewster's angle are commonly used for He:Ne lasers in the manner indicated in Fig. 0.1. What is the angle measured from the axis of the cavity? (η[quartz] = 1.43).

REFERENCES AND SUGGESTED READINGS

1. E. C. Jordan and K. G. Balmain, *Electromagnetic Waves and Radiating Systems*, 2nd ed. (Englewood Cliffs, N.J.: Prentice Hall, 1968), Chaps. 1–7.
2. S. Ramo, J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics* (New York: John Wiley & Sons, 1965), Chaps. 1–6.
3. A. Nussbaum and R. Phillips, *Contemporary Optics for Scientists and Engineers* (Englewood Cliffs, N.J.: Prentice Hall, 1976), Chap. 6.
4. R. M. Eisberg, *Fundamentals of Modern Physics* (New York: John Wiley &Sons, 1961), Chap. 6.
5. F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 3rd ed. (New York: McGraw-Hill, 1957). An excellent introduction to optics.
6. W. H. Hayt, Jr. *Engineering Electromagnetics* (New York: McGraw-Hill, 1974).

7. D. K. Cheng, *Field and Wave Electromagnetics* (Reading, Mass.: Addison-Wesley, 1983).
8. N. N. Rao, *Elements of Engineering Electromagnetics*, 2nd Ed. (Englewood Cliffs, N.J.: Prentice Hall, 1987).



Ray Tracing in an Optical System

2.1 INTRODUCTION

In Chapter 1 we found that a physically small beam of electromagnetic energy at very high frequencies does not spread to any degree. This allows us to follow our intuition and follow rays of light as they traverse an optical path. Let us first define a ray as the *path* that the center of a very slowly diverging electromagnetic beam would take as it goes through the system. Note that the word *path* was emphasized, for a ray does not have an amplitude. Furthermore, we require that the spatial extent of this beam in the transverse direction be small compared to the size of the optical components.

2.2 RAY MATRIX

In order to describe the "gyrations" of a ray in an optical system consisting of lenses, lengths of free space, interfaces between different dielectric media, mirrors, and so on, we should first ask, What is necessary to specify everything about it? The answer is deceptively simple:

1. Where is it with respect to some arbitrarily chosen axis?
2. In what direction is it heading?

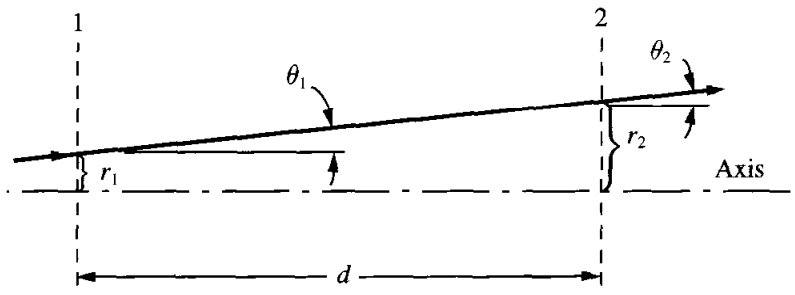


FIGURE 2.1. Ray in a homogeneous dielectric of length d.

We can answer these questions by inspection for the most important building block of an optical system, a length of free space (see Fig. 2.1). Obviously, if we know where the ray is at plane 1 and know its slope with respect to the axis, then we know where the ray emerges and where it is going at the exit plane 2. Indeed, the example is so trivial that it could be boring, but since it is so simple, we use it to introduce the notation.

We assume here and in everything to follow that all rays are paraxial, so that $\tan\theta \approx \sin\theta \sim \theta$ for all angles measured with respect to the optic axis, and thus the angle θ equals the slope of the ray, r'. The output parameters are related to the input parameters by

$$\begin{aligned} r_2 &= 1 \cdot r_1 + d \cdot r_1' \\ r_2 &= 0 \cdot r_1 + 1 \cdot r_1' \end{aligned} \tag{2.2.1}$$

or we could (and will) write this in matrix form as

$$\begin{bmatrix} r_2 \\ r_2' \end{bmatrix} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_1' \end{bmatrix} \tag{2.2.2}$$

In general, the relation between the output and input parameters of a general optical system is given by the $ABCD$ matrix of the form

$$\begin{bmatrix} r_{\text{out}} \\ r_{\text{out}}' \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} r_{\text{in}} \\ r_{\text{in}}' \end{bmatrix} \tag{2.2.3}$$

Thus ray tracing through a sequence of optical components is reduced to simple 2×2 matrix multiplication. For instance, if we consider two lengths of free space (see Fig. 2.2), the

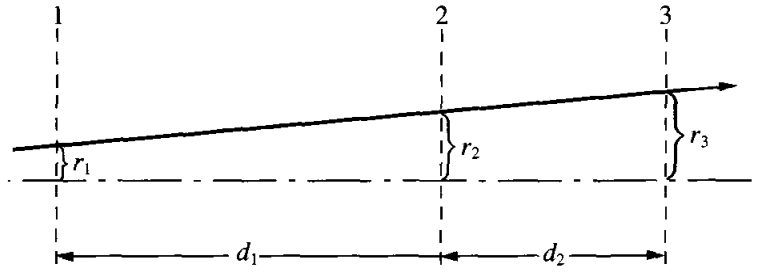


FIGURE 2.2. Example of two lengths of space.

output of one optical component is the input to the other. Thus we have

$$\left.\begin{aligned} &\begin{bmatrix} r_3 \\ r_3' \end{bmatrix} = \begin{bmatrix} 1 & d_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_2 \\ r_2' \end{bmatrix}; \qquad \begin{bmatrix} r_2 \\ r_2' \end{bmatrix} = \begin{bmatrix} 1 & d_1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_1' \end{bmatrix} \\ \text{or} \\ &\begin{bmatrix} r_3 \\ r_3' \end{bmatrix} = \begin{bmatrix} 1 & d_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & d_1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_1' \end{bmatrix} = \begin{bmatrix} 1 & d_1 + d_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_1' \end{bmatrix} \end{aligned}\right\} \tag{2.2.4}$$

Equation (2.2.4) is a very complex way of obtaining an obvious result; if (2.2.2) is correct, then the length of $d_1 + d_2$ is substituted for d for the cascade.

Some may start to recognize the analogy between the formalism used here and that used for the cascade of electrical networks (see Fig. 2.3). The analogy is excellent (and intentional). The notation is the same, $ABCD$, and even some of the special details are the same. For instance, the determinant of coefficients of T is unity for a bilateral electrical network:

$$AD - BC = 1 \tag{2.2.5}$$

Note that this holds for the matrices in (2.2.2) and (2.2.4). For optical rays this is true provided that the index of refraction at the exit plane is the same as the entrance plane.

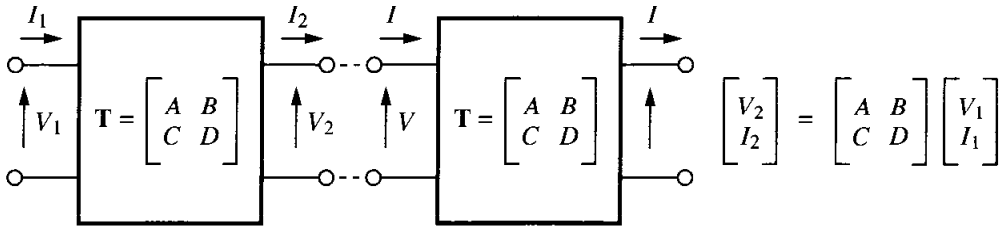


FIGURE 2.3. Correspondence with electrical network theory.

2.3 SOME COMMON RAY MATRICES

The most important ray matrix in optical systems is that of a length of free space. The next most important element is a thin lens of focal length f. Let us use the formalism of the preceding section and some rather obvious facts about lenses to obtain its ray matrix (see Fig. 2.4). Here we assume that the lens is so thin that there is negligible distance between the entrance plane (1) and exit plane (2). Thus no matter what the slope of the incoming ray, the output position is always equal to the input position, or

$$r_2 = r_1$$

Therefore, $A = 1$ and $B = 0$.

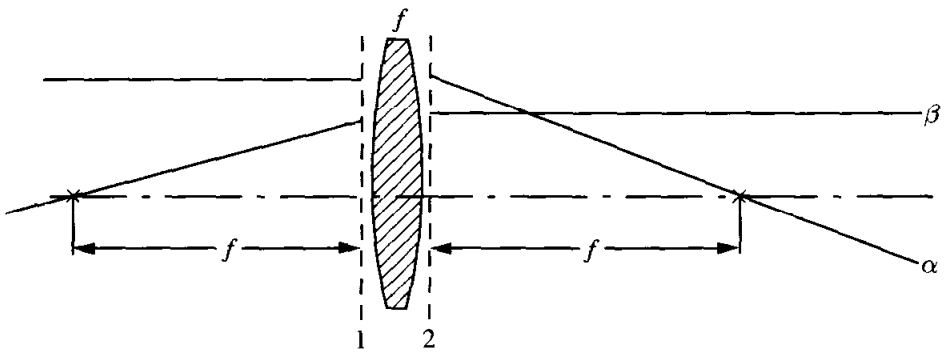


FIGURE 2.4. Paper experiment with a "thin" lens.

Now consider the circumstances provided by ray α in the diagram. For this special case the input slope $r_1' = 0$, yet it is obvious that the output slope is $-r_1/f$.

$$r_{2\alpha}' = Cr_{1\alpha} + Dr_{1\alpha}' = -\frac{r_{1\alpha}}{f} = Cr_{1\alpha} + D \cdot 0$$

or

$$C = -\frac{1}{f}$$

In the other case, ray β comes in with a slope of $+r_1/f$ and obviously exits parallel to the axis. Thus, $r_2' = 0$.

$$r_{2\beta}' = 0 = -\frac{1}{f} r_{1\beta} + Dr_{1\beta}' \qquad r_{1\beta}' = \frac{r_{1\beta}}{f}$$

Therefore $D = 1$

Hence the ray matrix of a thin lens is given by

$$\mathbf{T} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix} \tag{2.3.1}$$

Again note that the determinant of **T** is unity.

Figure 2.5 on page 39 combines the two elements in cascade. The transmission matrix between planes 1 and 3 is given by

$$\mathbf{T} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix} \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & d \\ -\frac{1}{f} & 1 - \frac{d}{f} \end{bmatrix} \tag{2.3.2}$$

Note that the matrices appear in reverse order to that encountered as the ray goes through the system. (This is because the output of the first component is the input to the second.) Again, note that the determinant is unity.

Another important component is a mirror (see Fig. 2.6 on page 39). Here we run into a bit of a problem, because the element causes a major redirection of the ray. In other words,

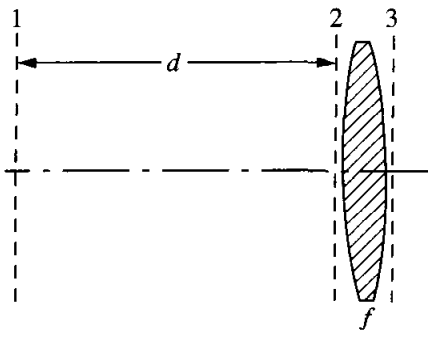


FIGURE 2.5. Combination of a lens plus free space.

the entrance and exit planes are on the same side of the surface of the mirror. If we ride *with* the ray, the effect of the mirror is to slightly redirect its path according to the laws of geometric optics. To an observer riding with the ray, the effect of the spherical mirror of radius R is to direct the ray toward the axis just like a thin lens. It is very easy to show that the focal length of a spherical mirror is just one-half of the radius of curvature. Thus the transmission matrix is given by

$$\mathbf{T} = \begin{bmatrix} 1 & 0 \\ -\dfrac{2}{R} & 1 \end{bmatrix} \tag{2.3.3}$$

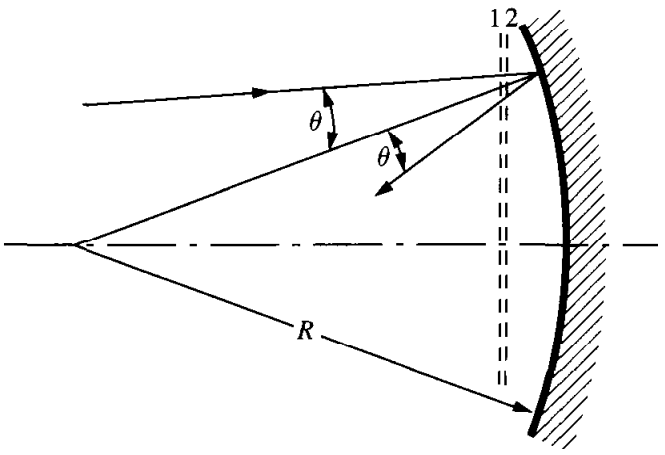


FIGURE 2.6. Mirror. Note that the entrance (1) and exit (2) planes are on the same side of the mirror.

2.4 APPLICATIONS OF RAY TRACING: OPTICAL CAVITIES

One of the most important uses of ray tracing is in the analysis of optical cavities such as the very simple but very important one shown in Fig. 2.7. Obviously, if a ray gets started inside the cavity, it will bounce back and forth between the two mirrors, being redirected and focused each time it hits the surface.

If the ray position stays "close" to the optical axis even after many transits between the mirror, the system is *stable*; if the ray naturally "walks off" one of the mirror surfaces, it is *unstable*; and if the mirrors must be perfectly aligned to keep the ray near the axis, it is *conditionally stable*.

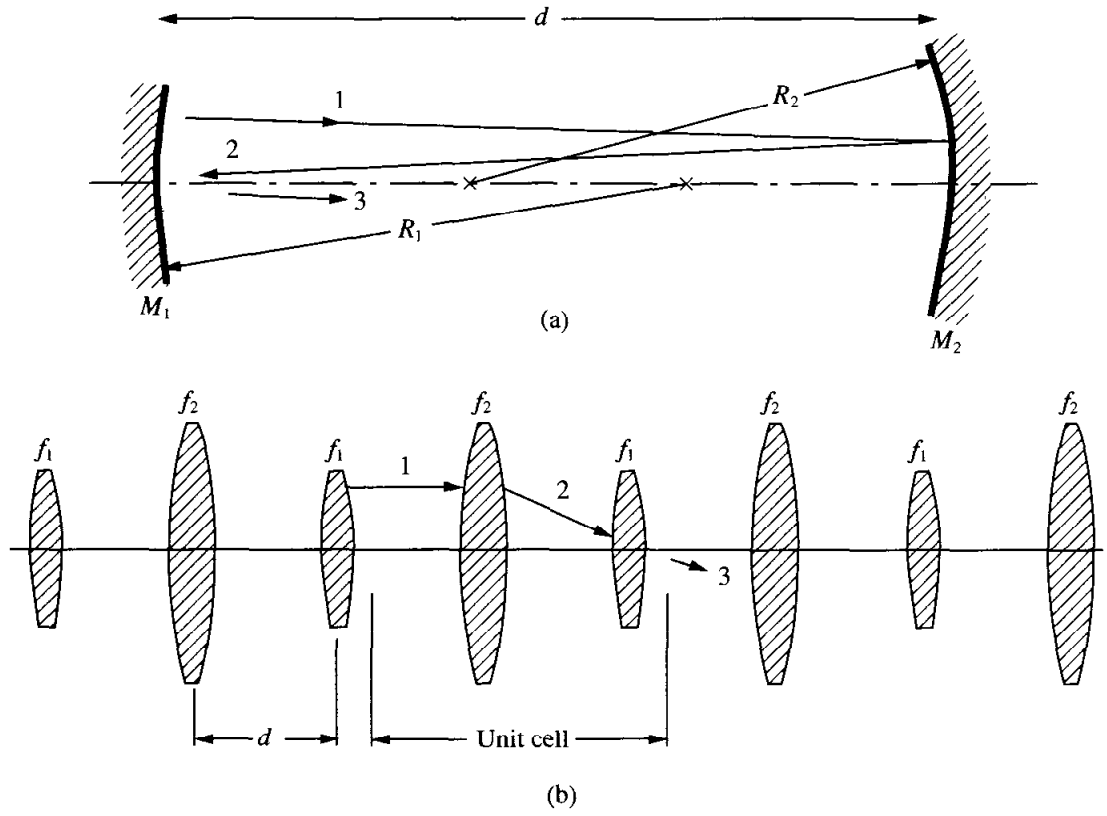


FIGURE 2.7. (a) Optical cavity showing a ray bouncing back and forth between the mirrors; (b) lens-waveguide equivalent to the mirror system shown in (a).

To analyze the stability of such a cavity, we construct an equivalent lens waveguide that would redirect the beam in the same manner as the mirrors. For instance, if we follow the rays shown in Fig. 2.7(a), we first encounter a length d, then a focusing element M_2, then another length d, and finally a focusing element M_1; then we start all over again. The infinite sequence of lenses shown in Fig. 2.7(b) would redirect the ray in the same manner as the cavity and is called the equivalent-lens waveguide.

Given the information of preceding sections, it is easy to compute the transmission matrix of the unit cell shown; it is just the product of two matrices of the form given by (2.3.2).

$$\mathbf{T} = \begin{bmatrix} 1 & d \\ -\dfrac{1}{f_1} & \left(1 - \dfrac{d}{f_1}\right) \end{bmatrix} \begin{bmatrix} 1 & d \\ -\dfrac{1}{f_2} & 1 - \dfrac{d}{f_2} \end{bmatrix}$$

$$\mathbf{T} = \begin{bmatrix} 1 - \dfrac{d}{f_2} & d + d\left(1 - \dfrac{d}{f_2}\right) \\ -\dfrac{1}{f_1} - \dfrac{1}{f_2}\left(1 - \dfrac{d}{f_1}\right) & \left(1 - \dfrac{d}{f_1}\right)\left(1 - \dfrac{d}{f_2}\right) - \dfrac{d}{f_1} \end{bmatrix} \tag{2.4.1}$$

This matrix is sufficiently messy to entice us to use the general letter symbols $ABCD$ for it and thereby cover all unit cells for all cavities at one time. After we finish, we can return to (2.4.1) to examine the details about this particular cavity.

Let us find a second-order difference equation for the ray as it passes the various planes of the succeeding unit cells that correspond to observing a ray as it makes successive round-trips through the cavity. We do this by eliminating the slope from these equations:

$$r_{s+1} = Ar_s + Br'_s \qquad \text{or} \qquad r'_s = \frac{1}{B}\left(r_{s+1} - Ar_s\right)$$

and

$$r'_{s+1} = \frac{1}{B}\left(r_{s+2} - Ar_{s+1}\right) = Cr_s + Dr'_s$$

Substituting r'_s from the first line, we obtain

$$\frac{1}{B}\left(r_{s+2} - Ar_{s+1}\right) = Cr_s + \frac{D}{B}\left(r_{s+1} - Ar_s\right)$$

Combining terms and remembering that $AD - BC = 1$ for a complete roundtrip in *any* cavity leads to

$$r_{s+2} - 2\left(\frac{A+D}{2}\right)r_{s+1} + r_s = 0 \tag{2.4.2}$$

We now ask the question, Does (2.4.2) have solutions in which the magnitude of r is less than some maximum value?

It is important to realize why we should ask the question. If the answer to it is yes, then the position of the ray form the axis undulates as it propagates along the lens waveguide (or between the two mirrors). If the ray's position is not bounded, it will eventually become so big that it will "miss" one of the components and thus walk off the mirror. These possibilities are shown in Fig. 2.8. The solid circles represent a typical position variation of a stable situation, and the open dots represent an unstable resonator. The points are connected by a dotted line to help in visualization, but it must be emphasized that the $ABCD$ matrix relates the two planes, $s + 1$ to s, but tells you nothing about the trajectory of the ray between the two planes. For complicated cavities it could make some wondrous gyrations between the two planes.

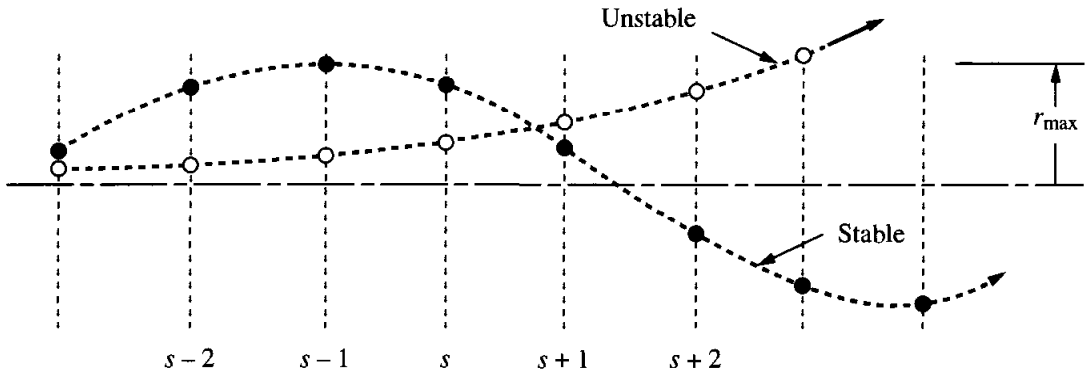


FIGURE 2.8. Example of the ray's position at the various planes of the lens waveguide.

However, the visualization provided by the dotted curve leads us to attempt a trigonometric solution. Let us assume that the solution to (2.4.2) has the following form:

$$r_s = r_0(e^{j\theta})^s = r_0 e^{js\theta} \tag{2.4.3}$$

Now, of course, the position r must be real. Hence we anticipate that we will obtain two solutions (to the second-order difference equation) whose combination yields a real answer. Substituting this guess* into (2.4.2), we obtain

$$r_0 e^{js\theta}\left[e^{j2\theta} - 2\left(\frac{A+D}{2}\right)e^{j\theta} + 1\right] = 0 \tag{2.4.4}$$

Now r_0 is specified by the initial conditions. It cannot be set equal to zero for the general case. The exponential term is not zero; hence, the second factor must be zero. That is a quadratic equation in $\exp(j\theta)$. The two solutions are

$$e^{j\theta} = \frac{A+D}{2} \pm j\left[1 - \left(\frac{A+D}{2}\right)^2\right]^{1/2} \tag{2.4.5}$$

Note that we did obtain two solutions, and, if all the quantities in (2.4.5) are real, then the solutions are complex conjugates. Thus the general solution to (2.4.2) is a linear combination of the form

$$\left.\begin{aligned} r_s &= r_0 e^{js\theta} + r_0^* e^{-js\theta} \\ \text{or} \qquad r_s &= r_{\max} \sin(s\theta + \alpha) \end{aligned}\right\} \tag{2.4.6}$$

This last form emphasizes the fact that the position must be real.

2.5 STABILITY: STABILITY DIAGRAM

Equations (2.4.5) and (2.4.6) only make sense provided that all quantities are real. If they are complex, the physical interpretation is considerably different and (2.4.6) must be changed. It is sufficient for $(A + D)/2$ to be less than ± 1 so that θ be real and thus have a bounded solution implied by (2.4.6). This is the general condition for stability:

$$-1 \le \left(\cos\theta = \frac{A+D}{2}\right) \le 1 \tag{2.5.1}$$

By adding 1 to this equation and then dividing by 2, we obtain

$$0 \le \frac{A+D+2}{4} \le 1 \qquad \text{(stable)} \tag{2.5.2}$$

Before we proceed further, let us return to the specific case of the two-spherical-mirror cavity. Equation (2.4.1) gives the elements of the ray matrix for the unit cell. Thus the

*Guessing is an acceptable method of solving difference *and* differential equations, provided, that is, that it is successful!

stability condition becomes

$$\frac{A+D+2}{4} = \frac{1}{4}\left[1 - \frac{d}{f_2} - \frac{d}{f_1} + \left(1 - \frac{d}{f_2}\right)\left(1 - \frac{d}{f_1}\right) + 2\right]$$
$$= 1 - \frac{d}{2f_1} - \frac{d}{2f_2} + \frac{d^2}{4f_1 f_2} = \left(1 - \frac{d}{2f_1}\right)\left(1 - \frac{d}{2f_2}\right) \tag{2.5.3a}$$

Since $2f_1 = R_1$ and $2f_2 = R_2$, the stability condition can be written in a compact and simple form:

$$0 \leq g_1 g_2 \leq 1 \tag{2.5.3b}$$

where

$$g_{1,2} = 1 - \frac{d}{R_{1,2}} \tag{2.5.4}$$

Even though this is a trivial equation, it is so important that it is graphed in Fig. 2.9. We can

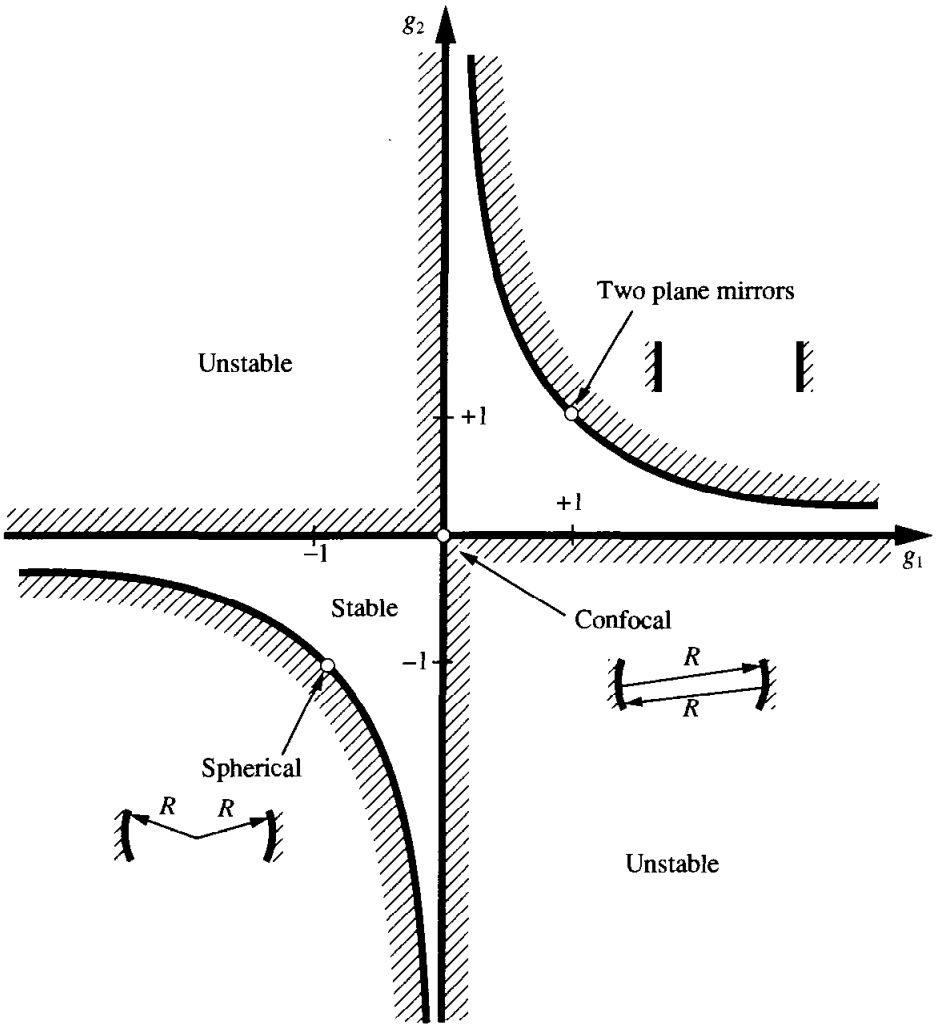


FIGURE 2.9. Stability diagram for the cavity of Fig. 2.7.

tell by a glance at this graph whether the system is stable. If the dimensions of the cavity are such that product of the g parameters is inside the cross-hatched region, the cavity is stable; if outside, it is unstable; and if on the border, it is conditionally stable, requiring perfect alignment.

One of the surprises that this diagram provides is that the confocal geometry, consisting of two identical mirrors and separated by the sum of the two focal lengths, is on the borderline of stability ($g_1 = g_2 = 0$). Some would think that this should be the most stable of all cavities. Indeed, it was one of the first cavities analyzed "exactly" with analytic techniques. (In fact, it is the only one that yields an analytical answer. See Sec. 12.7.3.) Nevertheless, it is on that borderline, and most lasers are made to avoid that situation by increasing or decreasing d slightly.

2.6 THE UNSTABLE REGION

The unstable region is described by the condition

$$\left(\frac{A+D}{2}\right)^2 > 1 \qquad \text{(unstable)} \tag{2.6.1}$$

and is shown by the cross-hatched region in Fig. 2.9. It is a natural human tendency to avoid the "unstable" region, because of its name. However, resonators operating in the unstable region have become very useful for high-gain laser systems. In that case, the rays that walk off the mirrors constitute the output.

A very crude example illustrates why we should not let personal misgivings about these resonators interfere with scientific judgment. Suppose that we had a small pencil-like beam of light (i.e., a ray) starting out in the previous cavity and that after 10 reflections this beam missed one of the mirrors. Assume further that the medium inside the cavity increases the power to the beam by a factor of 5 each time the beam passes through the medium. Since it makes 10 passes, the emerging beam that misses a mirror is amplified 10 times and thus contains much more power than the original one by a factor of $5^{10} = 9.76 \times 10^6$. If the initial beam contained only 1 mW of power, then the emerging beam has nearly 10 kW of power. Although the example is crude, it illustrates that unstable resonators have their place. We have more to say about this in Chapter 12.

2.7 EXAMPLE OF RAY TRACING IN A STABLE CAVITY

There are many applications of ray tracing, and deciding whether a cavity is stable or unstable is just one. To show some of the possibilities afforded by the formalism introduced here, let us consider the cavity shown in Fig. 2.10.

We imagine a situation where the incoming position and slope are specified in the manner illustrated. We take the specific case first and then generalize the result.

We assume that the incoming ray is perpendicular to the flat mirror ($R_1 = \infty$) and displaced from the axis by a specified amount $-r_0$. Thus this ray is directed toward the

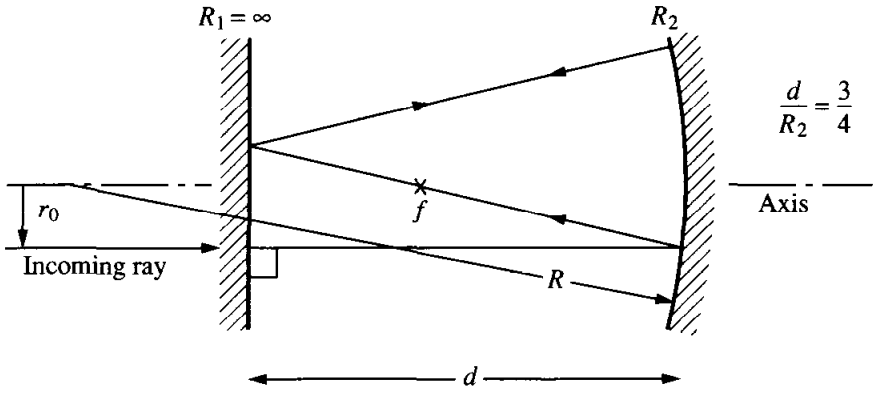


FIGURE 2.10. Ray tracing in a stable cavity.

focal point of M_2, reflects from the flat mirror, and heads toward the spherical mirror along a radius. Once this ray reaches M_2, it starts a retrace of its initial path and forevermore stays along paths indicated. This is an example of a repetitive ray path.

Let us now apply the formalism developed here to this case to establish general conditions for discrete beam paths in a cavity and to apply our previous analysis. First, note that this cavity is stable because $g_1 = 1 - d/R_1 = 1$, and $g_2 = 1 - d/R_2 = \frac{1}{4}$; hence, $g_1 g_2 = \frac{1}{4} < 1$. Note, too, that the beam stays within a maximum displacement of $2r_0$ from the axis, consistent with the ideas of stability developed earlier. Let us now consider the lens-waveguide equivalent to this cavity, shown in Fig. 2.11. The transmission matrix for the first unit cell is given by

$$\mathbf{T}_1 = \begin{bmatrix} 1 - \dfrac{d}{f} & d + d\left(1 - \dfrac{d}{f}\right) \\ -\dfrac{1}{f} & 1 - \dfrac{d}{f} \end{bmatrix} \tag{2.7.1}$$

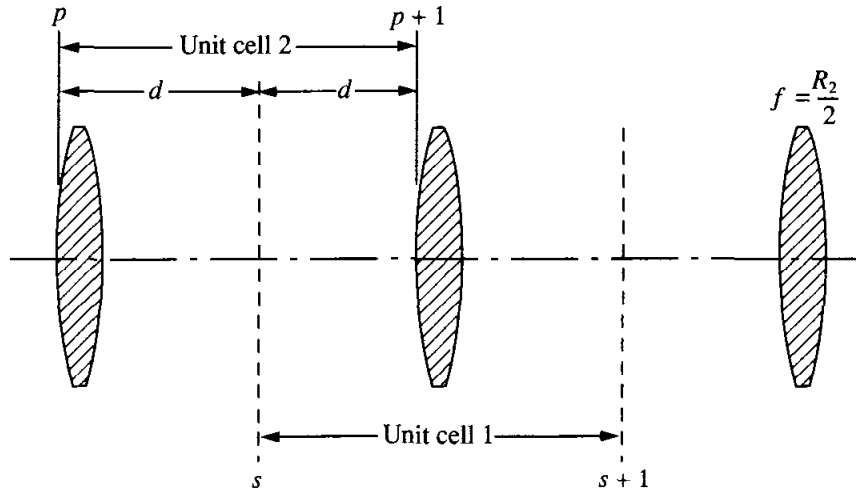


FIGURE 2.11. Lens-waveguide equivalent to the cavity of Fig. 2.10.

From (2.4.5), we have

$$e^{j\theta} = \frac{A+D}{2} + j\left[1 - \left(\frac{A+D}{2}\right)^2\right]^{1/2} = \cos\theta + j\sin\theta$$

or

$$\cos\theta = \frac{A+D}{2} = 1 - \frac{d}{f}$$

for this case. Since $f = R/2$, we find that $\cos\theta = 1 - \frac{3}{2} = -\frac{1}{2}$ or $\theta = 2\pi/3$ (120°).

Let us also try to apply the second form of the solution to the difference equation and follow the real ray position as it leaves (and enters) unit cell 1. For this case, (2.4.6) becomes

$$r = +r_0 \sin\left(s\frac{2\pi}{3} - \frac{\pi}{2}\right) \tag{2.7.2}$$

where the phase angle $-\pi/2$ expresses the fact that the ray is $-r_0$ below the axis at the start of the first round trip (call this plane $s = 0$). At the end of this round trip (and one unit cell later), $s = 1$, and the ray's position is given by

$$r = -r_0 \sin\left(\frac{2\pi}{3} + \frac{\pi}{2}\right) = -r_0 \sin 210° + \frac{r_0}{2}$$

After another unit cell and one more round trip, $s = 2$ and (2.7.2) predicts that

$$r = -r_0 \sin\left(\frac{4\pi}{3} + \frac{\pi}{2}\right) = -r_0 \sin 330° = \frac{r_0}{2}$$

and then finally, after three round trips, $s = 3$, we have

$$r = -r_0 \sin\left(\frac{6\pi}{3} + \frac{\pi}{2}\right) = -r_0 \sin 90° = -r_0$$

and we start all over again.

There appears to be something wrong here. According to (2.7.2), the maximum displacement is r_0, whereas the simple walk through the cavity at the beginning of Sec. 2.6 indicated that the maximum displacement was $2r_0$. Why is there a difference?

There is nothing wrong; indeed, unit cell 1 was chosen to illustrate this particular point. The $ABCD$ matrix for a unit cell relates the position and slope of the ray as it enters and leaves the reference planes. It tells us nothing about what happens in between. If we need to know about the position between the reference planes of the unit cell (say at the spherical mirror), we must apply the individual ray matrix to obtain that information. For the choice of reference plane indicated for unit cell 1, the maximum displacement at those terminals is r_0.

If we had chosen the other unit cell, say cell 2, of Fig. 2.11, the ray matrix is different but the angle θ is the same.

$$\mathbf{T}_2 = \begin{bmatrix} 1 & 2d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix} = \begin{bmatrix} 1 - \frac{2d}{f} & 2d \\ -\frac{1}{f} & 1 \end{bmatrix}$$

$$\cos\theta = \frac{A+D}{2} = \frac{1}{2}\left(2 - \frac{2d}{f}\right) = 1 - \frac{d}{f} = 1 - \frac{2d}{R} = -\frac{1}{2} \tag{2.7.3}$$

$$\text{or: } \theta = \frac{2\pi}{3} \qquad \text{i.e. } (120^\circ)$$

Thus the displacement can be related to the pth reference plane of this unit cell by

$$r = 2r_0 \sin\left(p\frac{2\pi}{3} - \frac{\pi}{6}\right)$$

At the start of this first unit cell, $p = 0$, and the ray is impinging on the spherical mirror at $r = r_0$ below the axis. After one round trip, $p = 1$, and the ray's position is $2r_0 \sin(120^\circ - 30^\circ) = 2r_0$. At the end of the second round trip, $p = 2$, and we find that

$$r = 2r_0 \sin\left(2 \times \frac{2\pi}{3} - \frac{\pi}{6}\right) = 2r_0 \sin 210^\circ = -r_0$$

Note that this last choice of the unit cell tells us nothing about the trials and tribulations of the ray at the flat mirror.

2.8 REPETITIVE RAY PATHS

The preceding example illustrates a case whereby the beam retraces its path after a discrete number of round trips. In the particular case shown, the ray went three complete round trips and then was in the same position, with the same slope, as it was when it started. We can generalize this result for any cavity.

Equation (2.4.6) indicates that the ray position could be found from

$$r_s = r_{\max} \sin(s\theta + \alpha)$$

where (2.4.6)

$$\theta = \cos^{-1}\left(\frac{A+D}{2}\right)$$

If s is increased by m units, corresponding to m round trips, then the ray returns to its original position after these m round trips, when θ satisfies

$$m\theta = 2n\pi \qquad \text{and} \qquad n < \frac{m}{2} \tag{2.8.1}$$

when n and m are integers. For the case analyzed in the preceding section, $n = 1$ and $m = 3$. Note that inequality guarantees that θ is always less than π, in accordance with the principal value of $\cos\theta = (A + D)/2$.

2.9 INITIAL CONDITIONS: STABLE CAVITIES

The example given in Sec. 2.7 was transparent enough to make the evaluation of the constants in (2.4.6) easy and was obtained almost by inspection. Let us now attempt to formalize the procedure for a more complex cavity. Let us suppose that we have unfolded a complex cavity and have found the transmission matrix for a unit cell, and that the ray's position a and the slope m are known at a given reference plane (call it $s = 0$).

If a cavity is stable, we can find the angle θ from the transmission matrix of the unit cell:

$$\theta = \cos^{-1}\left(\frac{A + D}{2}\right) \tag{2.9.1}$$

The initial conditions $r_0 = a$ and $r_0' = m$ yield the first equation involving $r_{\max}$ and α:

$$a = |r_{\max}| \sin(0 \cdot \theta + \alpha) = |r_{\max}| \sin\alpha \tag{2.9.2}$$

The ray matrix also provides us with a second equation for the ray position after traversing one unit cell (or round trip):

$$\begin{aligned} r_1 = Ar_0 + Br_0' &= |r_{\max}| \sin(1 \cdot \theta + \alpha) \\ A\alpha + Bm &= |r_{\max}| \sin(\theta + \alpha) \end{aligned} \tag{2.9.3}$$

Expanding the $\sin(\theta + \alpha)$ and recognizing that $|r_{\max}| \sin\alpha = a$ and $\cos\theta = (A + D)/2$ yields

$$|r_{\max}| \cos\alpha = \frac{1}{\sin\theta}\left[a\left(\frac{A - D}{2}\right) + Bm\right] \tag{2.9.4}$$

Thus the angle α is given by

$$\alpha = \tan^{-1}\left[\frac{a\left(1 - \left(\frac{A + D}{2}\right)^2\right)^{1/2}}{a\left(\frac{A - D}{2}\right) + Bm}\right] \tag{2.9.5}$$

2.10 INITIAL CONDITIONS: UNSTABLE CAVITIES

If the cavity is unstable, then the trigonometric solution (2.4.6) is confusing, misleading, and in any case more trouble than it is worth. It is best to return to the difference equation and reinterpret the conclusions that followed.

We can still assume the same functional form for a solution to (2.4.2), although we give it a slightly different "name." Let

$$r_s = r_0(F)^s \tag{2.10.1}$$

Then (2.4.4) becomes

$$rF^s\left[F^2 - 2\left(\frac{A+D}{2}\right)F + 1\right] = 0 \tag{2.10.2}$$

Thus there are two solutions (again), but now they are both real:

$$F_1 = \frac{A+D}{2} + \left[\left(\frac{A+D}{2}\right)^2 - 1\right]^{1/2} \tag{2.10.3a}$$

$$F_2 = \frac{A+D}{2} - \left[\left(\frac{A+D}{2}\right)^2 - 1\right]^{1/2} \tag{2.10.3b}$$

The general solution becomes

$$r_s = r_a(F_1)^s + r_b(F_2)^s \tag{2.10.4}$$

For unstable cavities $(A + D)/2$ is either greater than 1 or less than -1. In either case, one of the solutions has a magnitude greater than 1. Thus after a few steps through the unit cell, the ray's position is dominated by the larger quantity raised to the sth power:

$$r_s \sim r_0(F_>)^s$$

and is farther and farther away from the axis of the system.

The constants r_a and r_b are evaluated in the same manner as in Sec. 2.9. At the plane where the positions a and the slope m are specified, $s = 0$ and (2.10.4) yields

$$a = r_a + r_b \tag{2.10.5}$$

After one round trip, we have

$$r_1 = aA + Bm = r_aF_1 + r_bF_2 \tag{2.10.6}$$

Solving for r_a and r_b yields

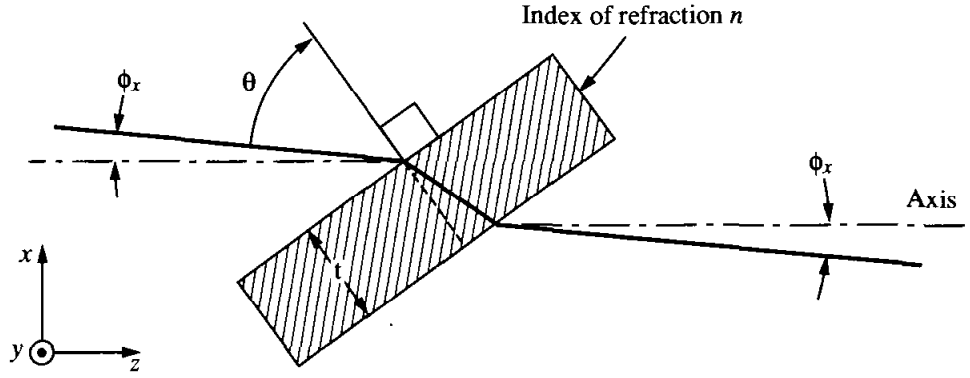
$$r_b = \frac{1}{F_1 - F_2}\left[a(F_1 - A) - Bm\right] \tag{2.10.7}$$

$$r_a = \frac{1}{F_2 - F_1}\left[a(F_2 - A) - Bm\right] \tag{2.10.8}$$

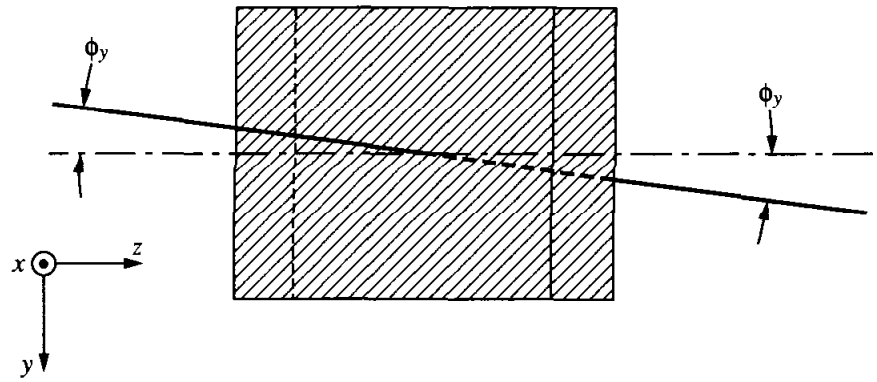
2.11 ASTIGMATISM*

When a material body is placed in the path of a ray and is tilted in the manner shown in Fig. 2.12, we must account for the change in the optical path in two orthogonal directions. For instance, windows placed at the Brewster angle are very common in gas lasers. The angles ϕ_x and ϕ_y can be considered as small and thus paraxial to the optic axis. However, if the dielectric material is a Brewster's angle window on a gas laser tube, then $\theta = \tan^{-1} n$, and for quartz with $n = 1.45845$,† $\theta = 55.56°$. Obviously, the angle $\theta - \phi_x$ is not small, and we must account for bending and resulting displacement of the beam in the xz plane as shown in Fig. 2.12(a). The paraxial approximation is quite adequate for rays in the yz plane. It is left as a problem to show that optical paths traversed by the two rays through a Brewster's angle window are given by

$$d_y = t\frac{\left(n^2+1\right)^{1/2}}{n^2} \tag{2.11.1a}$$



(a)



(b)

FIGURE 2.12. Astigmatism of a window. (a) Side view. (b) Top view.

*Some of the consequences of astigmatism for lasers are discussed in Ref. 8.

†From the *Handbook of Chemistry and Physics* (Chemical Rubber Co., 1983) at $\lambda = 589.29$ nm. p. E-364.

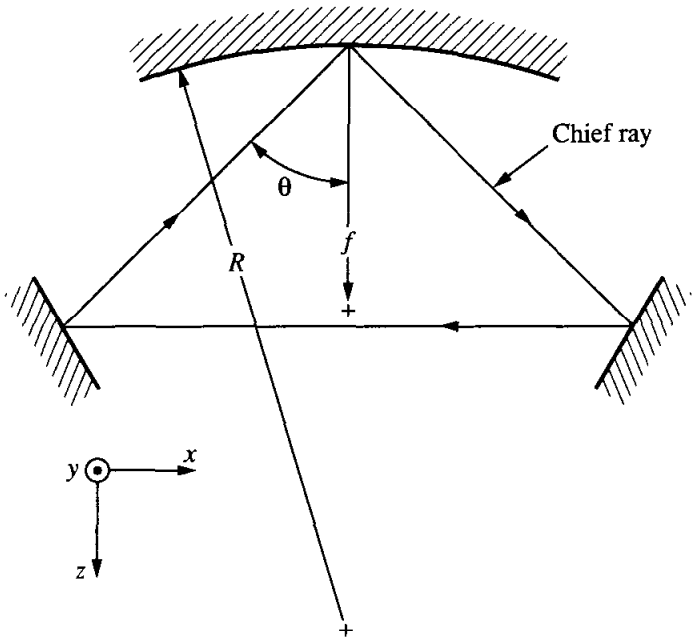


FIGURE 2.13. Astigmatic laser cavity.

$$d_x = t\frac{(n^2+1)^{1/2}}{n^4} \tag{2.11.1b}$$

when $\tan\theta_B = n$. The fact that these distances are not equal gives rise to astigmatism.

A curved mirror is often used in ring laser cavities in the manner shown in Fig. 2.13. It is a tiring exercise in geometry to show that the mirror focuses parallel rays in the two planes at different locations, leading to different effective focal lengths in the xy and xz planes. Thus, for rays that are paraxial to the "chief ray path," we use an effective focal length given by

$$f_x = f\cos\theta \tag{2.11.2a}$$

$$f_y = \frac{f}{\cos\theta} \tag{2.11.2b}$$

Astigmatism leads to elliptical beams in ring lasers and plays a critical role in dye-laser cavities.

2.12 CONTINUOUS LENS-LIKE MEDIA

Consider the case in which a ray is propagating in a medium in which the index of refraction is nonuniform in the transverse direction. We assume that the medium is not a function of the axial (i.e., z) coordinate.

Two contrasting examples of this type of a medium are shown in Figs. 2.14 and 2.15. The first case is that of an optical fiber used as a waveguide for communication purposes. The manufacturing process involves the drawing of different glasses with different dopants and different coatings through a die to attain a radial variation in the index of refraction.

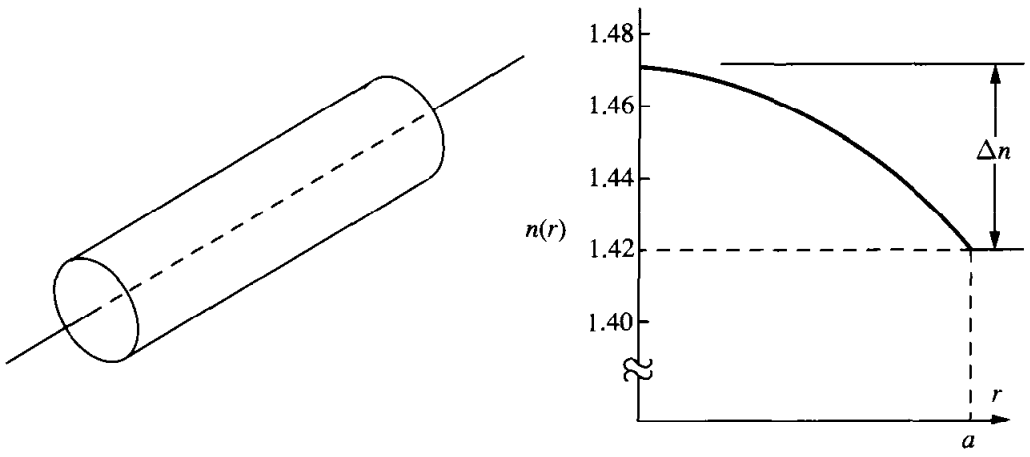


FIGURE 2.14. Optical fiber with an index of refraction depending on r.

The variation of the index of refraction, $n(r) = [\epsilon(r)]^{1/2}$, is shown with typical numbers to emphasize that the change is usually quite small. As we shall see, this small change is sufficient to guide the electromagnetic energy efficiently over long distances.

The second case is that of a gas discharge used to pump some of the common low-power lasers such as helium/neon (632.8 nm), argon ion (488.8 nm, 514.5 nm), or CO_2 (10.6 μm) systems. It will be shown much later that the current is *not* uniform throughout the bore of the tube shown but rather is distributed nonuniformly in the manner sketched. Since the current is maximum at the center, the temperature of the gas is highest there also and decreases to the value of the walls. To a first approximation, the pressure is constant across the bore of the tube, and since the temperature is a function of radius, the density of

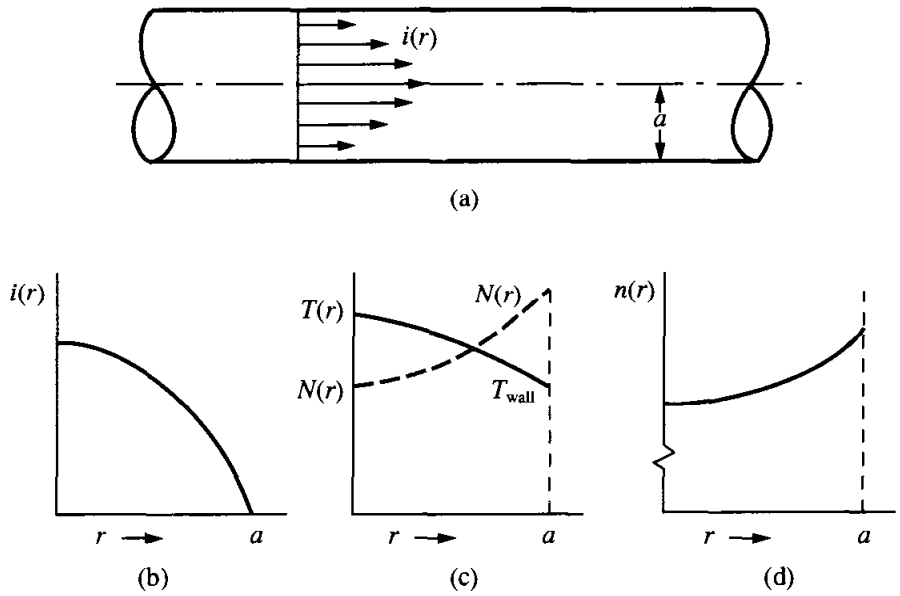


FIGURE 2.15. Variation of the index of refraction of a gas due to local heating by the current density distributed as in (b).

neutral gas atoms must be the inverse function

$$N(r)kT(r) = \text{constant} \tag{2.12.1}$$

Inasmuch as the index of refraction of the gas is directly related to the density by

$$n - 1 = 2\pi\alpha N(r) \tag{2.12.2}$$

where $2\pi\alpha$ is the specific refractivity of the gas, it follows that the index of refraction is also a function of r, as shown in Fig. 2.15(d). In both cases the index of refraction is a function of r.

2.12.1 Propagation of a Ray in an Inhomogeneous Medium

We again assume that the paraxial ray is propagating primarily in the z direction, as shown in Fig. 2.16. We shall apply Snell's law to the interface shown on the diagram:

$$\frac{\omega}{c} n(r) \cos\theta_1 = \frac{\omega}{c} n(r + \Delta r) \cos(\theta_1 + \Delta\theta) \tag{2.12.3}$$

By using the first two terms of a Taylor series expansion for $n(r)$ and the expansion of $\cos(\theta_1 + \Delta\theta)$ yields

$$n(r)\cos\theta_1 = \left[n(r) + \frac{\partial n}{\partial r}\Delta r\right](\cos\theta_1 \cos\Delta\theta - \sin\theta_1 \sin\Delta\theta)$$

or

$$\frac{\partial n}{\partial r} = n(r)\tan\theta_1 \left(\frac{\Delta\theta}{\Delta r}\right) \tag{2.12.4}$$

Within the context of the paraxial ray approximation, $\tan\theta_1 = \Delta r/\Delta z$, and it follows that

$$\tan\theta_1 \cdot \frac{\Delta\theta}{\Delta r} = \frac{1}{n} \cdot \frac{\partial n}{\partial r} = \frac{\Delta r}{\Delta z} \cdot \frac{\Delta\theta}{\Delta r} = \frac{\Delta\theta}{\Delta z} \to \frac{\partial^2 r}{\partial z^2} \tag{2.12.5}$$

The combination of (2.12.4) and (2.12.5) yields the basic equation for the propagation of a ray in an inhomogeneous dielectric medium:

$$\frac{d^2 r}{dz^2} = \frac{1}{n(r)} \frac{dn(r)}{dr} \tag{2.12.6}$$

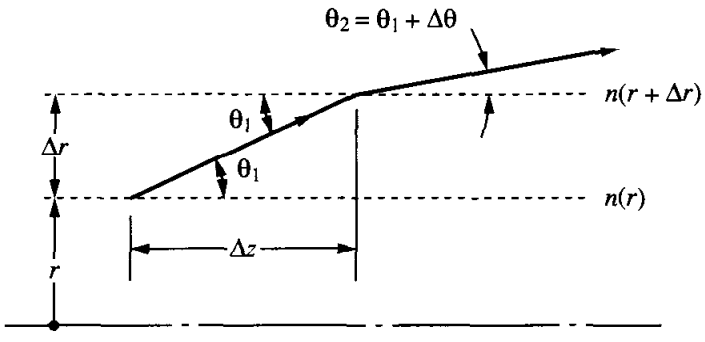


FIGURE 2.16. Propagation of a beam in an inhomogeneous medium.

2.12.2 Ray Matrix for a Continuous Lens

A simple thin positive lens bulges at $r = 0$, indicating more mass and a longer optical path there than at the edges. Similarly, a positive continuous lens has an index of refraction that decreases with r. For a cylindrically symmetric positive lens, the first two terms of Taylor series expansion for $n(r)$ would have the following format:

$$n(r) = n_0\left[1 - \frac{1}{2l^2}r^2 + O(r^4)\cdots\right] \tag{2.12.7}$$

We assume that the second term, $r^2/2l^2$, is much less than 1 for all values of the radius of concern to us. It should also be noted that the parameter l in (2.12.7) is simply a scale factor that indicates how fast n varies with r. For instance, if $a = 50\ \mu$m for the radius of the fiber in Fig. 2.14 and $n(a) = 1.42$ with $n_0 = 1.47$, then $\Delta n = 0.05$. The relationship between Δn and l is given by simple arithmetic:

$$\Delta n = n_0 - n(a) = \frac{n_0 a^2}{2l^2} \text{ or } l = a \cdot \sqrt{\frac{n_0}{2\Delta n}}$$

For the numbers just quoted, $l = 191.7\ \mu$m, which is large when compared with the radius of the fiber. Most fibers have a change in index much less than the value chosen here, and hence l is even larger.

Because of the inequality we can use the first term of (2.12.7) for n in (2.12.6) and the derivative of the second for dn/dr:

$$\frac{d^2r}{dz^2} = \frac{r}{l^2} \tag{2.12.8}$$

The solution is straightforward. Assume that $z = 0$ is the input plane to this optical component where the position r_1 and slope r_1' are known:

$$r(z) = r_1\left[\cos\left(\frac{z}{l}\right)\right] + r_1'\left[l\sin\left(\frac{z}{l}\right)\right] \tag{2.12.9}$$

Then, by differentiation, we obtain the slope at any position z:

$$r'(z) = r_1\left[-\frac{1}{l}\sin\left(\frac{z}{l}\right)\right] + r_1'\left[\cos\left(\frac{z}{l}\right)\right] \tag{2.12.10}$$

Thus the ray matrix for a length $z = d$ is

$$\mathbf{T} = \begin{bmatrix} \cos\left(\frac{d}{l}\right) & l\sin\left(\frac{d}{l}\right) \\ -\frac{1}{l}\sin\left(\frac{d}{l}\right) & \cos\left(\frac{d}{l}\right) \end{bmatrix} \tag{2.12.11}$$

for

$$n(r) = n_0\left(1 - \frac{r^2}{2l^2}\right)$$

Note that if $l \to \infty$, then the medium is uniform and (2.12.11) reproduces the ray matrix given by (2.2.2).

If the index of refraction increases with r (as in the case of a gas discharge), then the geometric fudge factor l must take on imaginary values to predict $n(r)$ from (2.12.7). If we let $l = jL$ and use the fact that $\cos j\theta = \cosh\theta$ and $\sin j\theta = j \sinh\theta$, we obtain the ray matrix for a negative lens directly from (2.12.11):

$$\mathbf{T} = \begin{bmatrix} \cosh\left(\frac{d}{L}\right) & L \sinh\left(\frac{d}{L}\right) \\ \frac{1}{L}\sinh\left(\frac{d}{L}\right) & \cosh\left(\frac{d}{L}\right) \end{bmatrix} \tag{2.12.12}$$

for

$$n(r) = n_0\left(1 + \frac{r^2}{2L^2}\right) \tag{2.12.13}$$

Be cautioned that the matrices (2.12.11) and (2.12.12) describe the ray propagation within the medium. The matrices do *not* describe the changes that might occur rather abruptly at the entrance and exit planes. For instance, if n_0 is significantly different from 1, as indicated by Fig. 2.14 for a glass fiber, then the slope of a ray changes in passing from the air into the fiber (and conversely). Since the variation of the index with the radius is slight, it is ignored for the inclusion of the air $\leftrightarrow$ fiber transmission matrix.

Let us examine (2.12.11) in a very imprecise manner so as to anticipate a result of a later section. The C term in this matrix is the negative of the focal length of this lens:

$$C = -\frac{1}{f} = -\frac{1}{l}\sin\left(\frac{d}{l}\right)$$

Thus the focal length is positive (i.e., converging) provided that the argument of the sine function is less than π radians. If d is long enough—as in a fiber-optic communication system—the focal length is alternately positive and negative, corresponding to a converging and diverging system. Thus we could anticipate the "beam" undulating as shown in Fig. 2.17 as it propagates down the fiber. Going one step further, anticipate a situation where the natural divergence of a small beam is continuously counteracted by the convergence of the medium. Such fibers exist and are used in fiber-optic communication links.

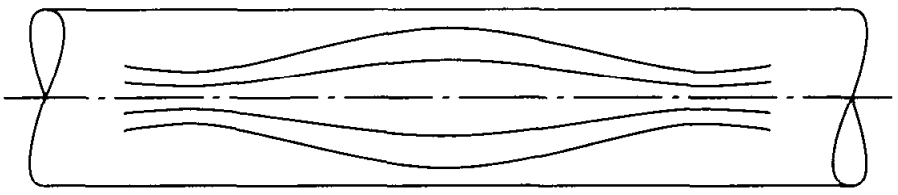


FIGURE 2.17. Propagation of a beam within a fiber.

2.13 WAVE TRANSFORMATION BY A LENS

Let us now broaden our viewpoint and consider the effect of a lens on a limited extent uniform plane wave. We can consider each of the horizontal lines in Fig. 2.18 to be a ray, and the bundle of rays to be a beam. Furthermore, we assume that the phase *surface* of the incident beam is more-or-less planar, as indicated by the vertical dashed lines.

The following question naturally arises: What is the shape of the phase front after passing through the lens? Obviously, the phases of A and B are equal at the left of the lens *and* at the focal point at the right. (This is sometimes referred to as the Fermat principle; that is, the light always takes the path that makes the transit time a minimum.) But if the waves are in phase at f, then A' and B' are not in phase (for equal z coordinates). The ray at B' has farther to travel; hence, its phase must be ahead of that of A'. This extra distance is

$$\left.\begin{aligned} &\Delta d = (r^2 + f^2)^{1/2} - f \doteq f\left(1 + \frac{1}{2}\frac{r^2}{f^2}\right) - f \\ \text{or} \quad &\Delta d \doteq \frac{1}{2}\frac{r^2}{f} \end{aligned}\right\} \tag{2.13.1}$$

Thus we have an approximate expression for the field just to the right of the lens:

$$E = E_0 \exp\left(+j\frac{kr^2}{2f}\right) \tag{2.13.2}$$

A moment's consideration indicates that the wave front is now curved (toward the focal point). The more exact analysis of Chapter 3 shows that this is correct.

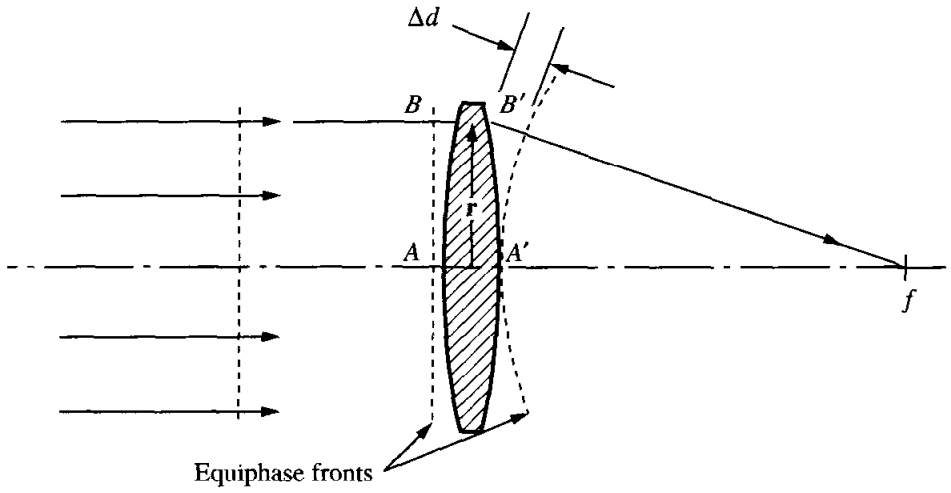
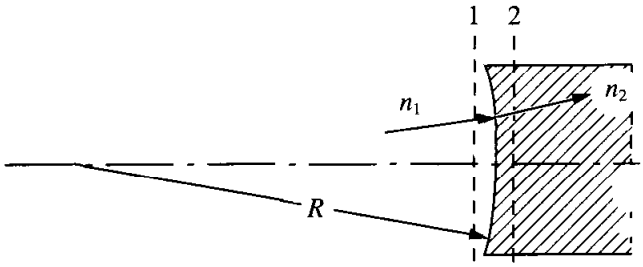


FIGURE 2.18. Beam transformation by a lens.

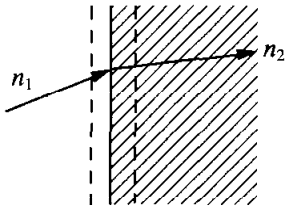
PROBLEMS

2.1. Derive the ray matrix for a ray entering a spherical dielectric interface.



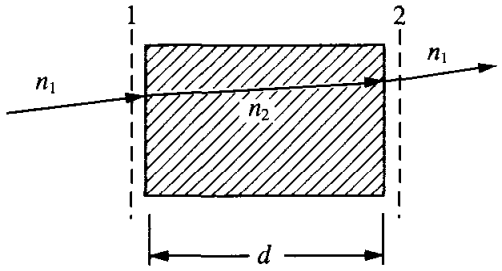
Ans.: $\mathbf{T} = \begin{bmatrix} 1 & 0 \\ \left(1 - \frac{n_1}{n_2}\right)\frac{1}{R} & \frac{n_1}{n_2} \end{bmatrix}$

2.2. Derive the ray matrix for the plane dielectric interface.



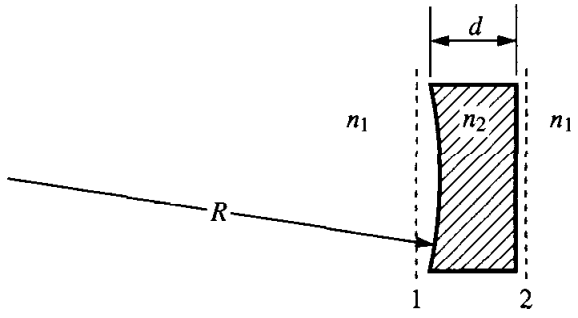
Ans.: $\mathbf{T} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{n_1}{n_2} \end{bmatrix}$

2.3. Derive the ray matrix for the plane dielectric slap of thickness d.



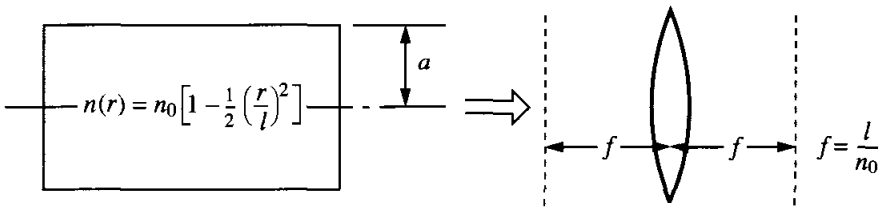
Ans.: $\mathbf{T} = \begin{bmatrix} 1 & \frac{n_1 d}{n_2} \\ 0 & 1 \end{bmatrix}$

2.4. Combine the results of problems 2.1 and 2.2 to derive the ray matrix for the negative lens. (Assume that $R \gg d$.)

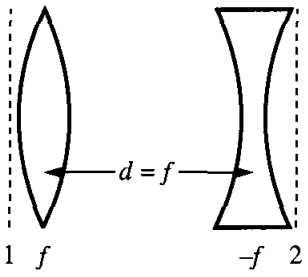


2.5. The GRIN lens shown in the diagram below consists of a graded index fiber with $n(r) = n_0[1 - (r^2/2l^2)]$ of length d.

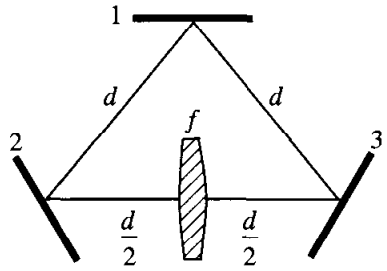
(a) Find the $ABCD$ matrix for this lens. Do not ignore the air-index boundary (use the ray matrix found in Problem 2.2 for that operation). Keep d arbitrary.
(b) Evaluate for $d = \pi l/2$.
(c) Show that this lens is equivalent to the system shown at the right.
(d) If $n_0 = 1.53$, $n(a) = 1.525$, and $a = 2$ mm, find f.



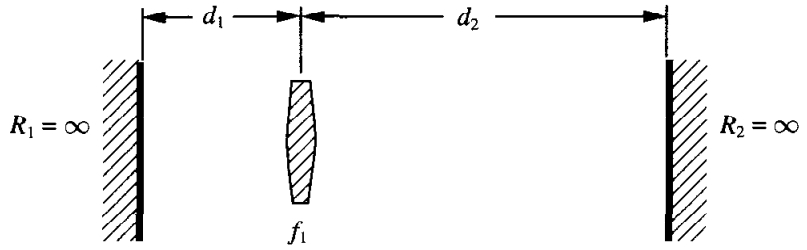
2.6. Find the $ABCD$ matrix for the lens combination shown below.



2.7. Consider the ring laser cavity shown in the accompanying diagram.
(a) Show an equivalent-lens waveguide for this cavity and identify a unit cell starting just after the lens and proceeding counterclockwise around the triangle.
(b) What is the transmission matrix for this unit cell? (Demonstrate that you have the component matrices in proper order.)
(c) What are the values of d/f that make this a stable cavity?



2.8. Consider the cavity shown in the accompanying diagram.
(a) Construct an equivalent-lens waveguide.
(b) Indicate a unit cell starting at a flat mirror, R_1.
(c) Find the ray matrix for the unit cell of (b).
(d) Discuss the stability of this cavity by constructing a diagram similar to Fig. 2.9.



2.9. Sketch the ray pattern shown in Fig. 2.10 if $d/R = 1/2$. Assume the initial ray is a distance r_0 below the axis.

(a) What is the maximum excursion of the path from the axis?

(b) Figure 2.10 shows three round trips to obtain a repetition. How many are required here?

2.10. Consider the obviously unstable cavity shown in the accompanying diagram. Suppose that a ray starts out at the flat mirror with a zero slope and at a position that is one tenth of the radius of the mirror.

(a) Show the equivalent-lens waveguide for the cavity. Identify two unit cells, one starting at the flat mirror and the second starting just before the spherical mirror.

(b) Since the cavity is unstable, we should return to the second-order difference equation to obtain a solution of the form (use unit cell 1)

$$r_s = r_a(F_+)^s + r_b(F_-)^s$$

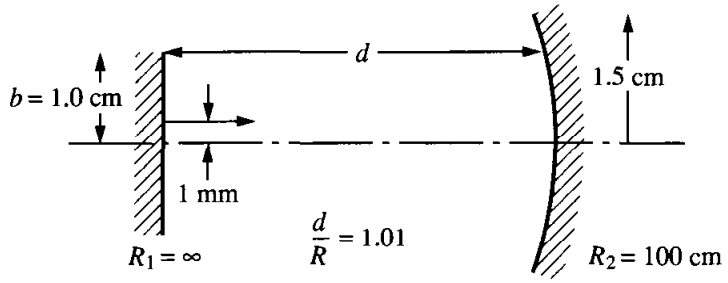
(1) What are the values of F_+ and F_-?

(2) What are the values of r_a and r_b?

(c) How many passes through the cavity does the ray make before it misses the flat mirror?

(d) Where does the ray leave the cavity: at the flat mirror or at the curved one? How many passes does it make? (To answer this, you will have to repeat (b) and (c) for the second unit cell and then decide.)

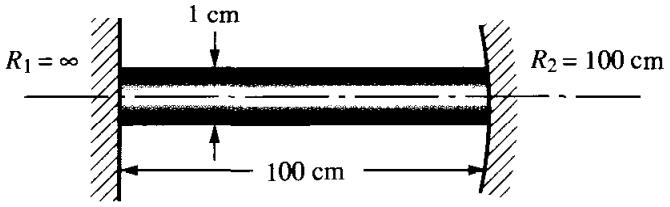
(e) If the beam associated with this ray started with 1 μW of power and the power gain per pass $G = 5$, what is the power leaving the cavity? (Assume that the mirrors are perfectly reflecting.)



2.11. In the analysis of the cavity shown in Fig. 2.13, one must keep track of displacements from the chief ray in the plane of and perpendicular to the paper. If the chief ray forms

an equilateral triangle with sides d and the spherical mirror has a radius of curvature R, what is the limit of d/R to ensure stability?

2.12. Consider the accompanying diagram of an optical cavity that is to be used for a gas discharge laser. The empty cavity is on the borderline of stability. As discussed in Sec. 2.12, the nonuniform gas heating by the discharge current creates a *negative* gas lens. Does this fact create an unstable situation? Assume that the parameter $d/L \ll 1$ and estimate the amount by which the stability margin is improved or degraded by finding the first *nonvanishing* term of a Taylor series expansion of $(A + D + 2)/4$.



2.13. **(a)** Suppose a cylindrical gas discharge tube of Problem 2.12 was filled with helium to a pressure of 1 torr at 23°C (i.e., 1/760 of an atmosphere). It was then sealed and detached from the vacuum system. Suppose that the current through the tube heated the gas such that the temperature on axis were 300°C decreasing in a parabolic fashion to 50°C at the wall. The pressure increases because of the higher temperature but must be constant independent of radius; that is, $p = N(r)kT(r) \neq f(r)$. Estimate the value of the parameter d/L. At standard temperature and pressure (STP), n (helium) = 1.000036. (Remember that the total number of atoms cannot change in a sealed-off tube.)

(b) Repeat (a) if the gas were CO_2 and the initial fill was to 100 torr. At STP, $n(CO_2) = 1.000449$.

2.14. Typical parameters of a glass fiber are radius $= 20\mu$m, $n_0 = 1.5$, and $\Delta n = 8 \times 10^{-3}$. How many times would a ray cross the axis of a fiber 1 km long?

2.15. In the derivation of (2.4.2), use was made of $AD - BC = 1$. However, here is a restriction on this relationship. Is it possible for $AD - BC \neq 1$ for a unit cell of an optical cavity?

2.16. Read the paper by H. Kogelnik and T. Li, "Laser Beams and Resonators," *Appl. Opt.* 9, 1550–1566, Oct. 1966. Verify ray matrices 4, 5, and 6 listed in Table I of that paper. (Caution: System 5 uses a different notation from that used in Sec. 2.11.)

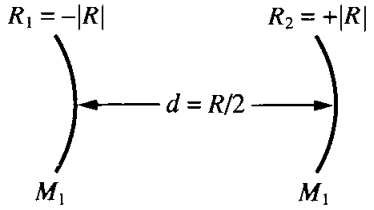
2.17. Examine the stability of the cavity shown on the diagram below by the following steps.

(a) Draw an equivalent lens waveguide.

(b) Identify a unit cell that starts on the right side of M_1.

(c) Find the transmission matrix for the unit cell.

(d) Apply the stability criterion.



2.18. Consider a general equivalent lens waveguide that is symmetric about a plane midway between the terminals 1 and 2. Show that $A = D$ for such a case.

2.19. Consider two lenses with equal focal lengths separated by a distance $\delta d \ll f$. Find the equivalent focal length of the combination.

2.20. The ray matrix for a flat mirror cocked at an angle θ with respect to the axis of the incident beam can be expressed as $A = D = \pm 1$, $B = C = 0$ depending upon the choice for the definition of positive r_2: counterclockwise (up) or clockwise (down) with respect to the chief reflected ray. Present an argument to show why each alternative is logical but that the stability criterion is not affected. (The text will always use the $+1$ choice.)

2.21. An isotropic radiator (i.e., a point source) of light is 10 cm below the surface of a large body of water ($n = 1.33$).

(a) What is the radius of the largest circle through which the light can emerge into air?

(b) The radiation from the source is independent of (θ, ϕ) but that emerging into air is not. Why not?

2.22. Show that the transmission matrix for the reverse path is

$$\mathbf{T}_{21} = \begin{bmatrix} D & B \\ C & A \end{bmatrix} \text{ if } \mathbf{T}_{12} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

is the ray matrix going from plane 1 to plane 2.

2.23. Suppose a unit cell for a cavity can be represented by the following:

$$\underset{\text{Unit cell}}{\longleftarrow\!\!\!\!\!\!\longrightarrow} \qquad \mathbf{T} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \qquad \mathbf{T}' = \begin{bmatrix} D & B \\ C & A \end{bmatrix}$$

where $\mathbf{T}$ is the transmission matrix for a half-way passage through the unit cell.

(a) What is the round trip transmission matrix for this unit cell?

(b) What is the stability criterion for such a cavity?

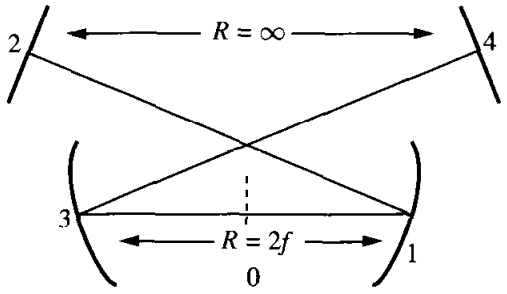
2.24. Now suppose the unit cell can be represented by:

$$\underset{\text{Unit cell}}{\longleftarrow\!\!\!\!\!\!\longrightarrow} \qquad \mathbf{T} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \quad \mathbf{T}' = \begin{bmatrix} D_1 & B_1 \\ C_1 & A_1 \end{bmatrix} \quad \mathbf{T} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \quad \mathbf{T}' = \begin{bmatrix} D_1 & B_1 \\ C_1 & A_1 \end{bmatrix}$$

(a) What is the transmission matrix for this unit cell?

(b) What is the stability criterion in terms of $A_1 \cdots D_1$.

(c) Show that the above is appropriate for the folded cavity shown below by identifying the sections for **T** and **T'**.



(d) Apply this analysis to the folded cavity.

REFERENCES AND SUGGESTED READINGS

There are many excellent references to matrix methods in optical design. However, there is a sequence of classic papers which bear directly on the subject matter of this and the following four chapters. It is recommended that the student read the papers [1] to [6].

1. J. R. Pierce, "Modes in Sequences of Lenses," *Proc. Natl. Acad. Sci. 47*, 1808–1813, 1961.
2. A. G. Fox and T. Li, "Resonant Modes in a Maser Interferometer," *Bell Syst. Tech. J. 40*, 453–488, Mar. 1961.
3. G. D. Boyd and J. P. Gordon, "Confocal Multimode Resonator for Millimeter through Optical Wavelength Masers," *Bell. Syst. Tech. J. 40*, 489–508, Mar. 1961.
4. G. D. Boyd and H. Kogelnik, "Generalized Confocal Resonator Theory," *Bell Syst. Tech. J. 41*, 1347–1369, July 1962.
5. H. Kogelnik, "Imaging of Optical Modes—Resonators with Internal Lenses," *Bell Syst. Tech. J. 44*, 455–494, March 1963.
6. H. Kogelnik and T. Li, "Laser Beams and Resonators," *Appl. Opt. 5*, 1550–1556, Oct. 1966. This article contains an additional 44 references.
7. F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 3rd ed. (New York: McGraw-Hill, 1957).
8. H. W. Kogelnik, E. P. Ippen, A. Dienes, and C. V. Shank, "Astigmatically Compensated Cavities for CW Dye Lasers," *IEEE J. Quant. Electron. QE-8*, 373, Mar. 1972.
9. A. Yariv, *Introduction to Optical Electronics* (New York: Holt, Rinehart & Winston, 1971) Chap. 2.
10. A. Yariv, *Quantum Electronics* (New York: John Wiley & Sons, 1975), Chap. 6.
11. A. Siegman, *Lasers* (Mill Valley, Calif.: University Science Books, 1986), Chap. 15.



Gaussian Beams

3.1 INTRODUCTION

Although the concept of ray tracing is intuitively satisfying (and simple), it is important to realize its limitations. A ray is a path and is not a field and does not have an amplitude, a phase, or a spatial extent. Yet this information is vital to all phases of electronics.

Consequently, it is necessary for us to obtain a more complete wave description of the beams produced by the laser. This will be accomplished by an approximate analytic solution to the wave equation yielding fields that are completely specified at all points in space. After a bit of mathematics, it will be shown that there exists a complete set of "modes" that are naturally generated by stable laser cavities.

It is most important to continually question the significance of the mathematical results so that we can appreciate *why* these modes are generated by lasers. The reason, in the simplest terms, is that these modes *match* the mirror surfaces. They fit, as will be demonstrated in Chapter 5.

3.2 PRELIMINARY IDEAS: TEM WAVES

We should first realize that most optical beams propagating in free space are almost pure TEM (transverse electric and magnetic). In other words, the field components lie in the plane

perpendicular to the direction of propagation. This is easy to show by a simple application of the divergence equation

$$\nabla \cdot \mathbf{E} = 0 \qquad \text{(for free space)} \tag{3.2.1a}$$

Obviously, this same comment applies to $\mathbf{H}$ since $\nabla \cdot \mathbf{H} = 0$. The divergence operation can be broken up into transverse divergence (∇_t) of the transverse field (E_t) and the longitudinal derivative of the z component of the field:

$$\nabla \cdot \mathbf{E} = \nabla_t \cdot \mathbf{E}_t + \frac{\partial}{\partial z} E_z = 0 \tag{3.2.1b}$$

Here some physical insight can save hours of arithmetic. The wave is obviously propagating with a velocity of approximately c. Hence, the *major* variation of the field with z is a term of the approximate form $\exp(-jkz)$ with $k \sim \omega n/c = 2\pi n/\lambda_0$. At optical frequencies of $> 10^{13}$ Hz, λ_0 is quite small, and hence k is a large number. Thus

$$\frac{\partial E_z}{\partial z} \sim -j \frac{2\pi n}{\lambda_0} E_z \tag{3.2.2}$$

Furthermore, a beam has a finite diameter D, which is typically 1 cm or so. Thus we can *approximate* the transverse divergence by

$$\nabla_t \cdot \mathbf{E}_t \sim \frac{|E_t|}{D} \tag{3.2.3}$$

The use of (3.2.2) and (3.2.3) in (3.2.1) yields a comparison of the magnitude of the z component of the field in terms of the transverse components:

$$|E_z| \sim \frac{\lambda_0}{2\pi n D} |E_t| \tag{3.2.4}$$

This ratio is exceedingly small for visible wavelengths (0.5 μm) and reasonable beam diameters ($\sim$ 1 cm). But a word of caution is in order here. If one is dealing with the focal spot of a lens, D can be quite small, making some of these conclusions questionable. Nevertheless, the assumption of a TEM wave for optical beams is usually quite good.

Let us now return to the point made earlier: that $k = 2\pi n/\lambda_0$ is a very large number. This means that the complete description of the fields must involve functions that change very rapidly with z. If we were to attempt to solve the wave equation on a digital computer, a numerical nightmare would result because of the rapid variation with z.

Thus it behooves us to eliminate this rapid variation with z from our equation. After all, we know that the fields propagate with a velocity of roughly c/n. Thus we look for solutions of the form

$$E(x, y, z) = E_0 \psi(x, y, z) e^{-jkz} \tag{3.2.5}$$

It is most important to realize the physical significance of (3.2.5). The factor E_0 is the customary amplitude factor expressing the intensity of the wave, the factor $\exp(-jkz)$ expresses our feelings that the wave should propagate more or less as a uniform plane wave, and the factor ψ measures how the beam deviates from a uniform plane wave.

We substitute (3.2.5) into the time-independent wave equation to derive another equation involving ψ alone. In other words, we say to ourselves, "We know about uniform plane waves. What about the deviation?"

$$\nabla^2 E + \frac{\omega^2}{c^2} n^2 E = 0$$

or

$$\nabla_t^2 E + \frac{\partial^2 E}{\partial z^2} + \frac{\omega^2}{c^2} n^2 E = 0 \tag{3.2.6}$$

Let

$$E = E_0 \psi(x, y, z) \exp(-jkz)$$

where

$$k = \frac{\omega}{c} n$$

In (3.2.6) we presume a uniform index of refraction: if n is a function of r, then the derivation must be repeated.

The following derivatives are necessary:

$$\nabla_t^2 E = E_0 (\nabla_t^2 \psi) \exp(-jkz)$$

$$\frac{\partial E}{\partial z} = E_0 \left(-jk\psi + \frac{\partial \psi}{\partial z} \right) \exp(-jkz)$$

$$\frac{\partial^2 E}{\partial z^2} = E_0 \left(-k^2 \psi - j2k \frac{\partial \psi}{\partial z} + \frac{\partial^2 \psi}{\partial z^2} \right) \exp(-jkz)$$

When these derivatives are substituted into (3.2.6), the term with k^2 cancels that with $(\omega n/c)^2$, and, of course, the common factor $\exp(-jkz)$ cancels out of all terms:

$$\nabla_t^2 \psi - j2k \frac{\partial \psi}{\partial z} + \frac{\partial^2 \psi}{\partial z^2} = 0 \tag{3.2.7}$$

Equation (3.2.7) is exact and is just a different representation of the wave equation. But now the first approximation is used; the second derivative term is neglected, with the following justification.

We can anticipate that the "beam" parameters contained in ψ do change with z leading to nonzero values for both $\partial \psi / \partial z$ and $\partial^2 \psi / \partial z^2$. But the first derivative is multiplied by this very large number k, whereas unity is the coefficient for the last term. Thus it is neglected, to yield

$$\nabla_t^2 \psi - j2k \frac{\partial \psi}{\partial z} = 0 \tag{3.2.8}$$

Equation (3.2.8) is the central equation for Gaussian beams. (Incidentally, it has the same form as the time-dependent Schrödinger equation.) It is called the paraxial wave equation.

3.3 LOWEST-ORDER $TEM_{0,0}$ MODE

To keep the arithmetic to a minimum, we look for a solution that is cylindrically symmetric. Equation (3.2.8) becomes

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \psi}{\partial r}\right) - j2k\frac{\partial \psi}{\partial z} = 0 \tag{3.3.1}$$

As is typical with the solution to differential equations, we "guess" at the functional form of the solution and then force the unknown coefficients or functions to fit the equation. We choose

$$\psi_0 = \exp\left\{-j\left[P(z) + \frac{kr^2}{2q(z)}\right]\right\} \tag{3.3.2}$$

where the subscript 0 indicates the fundamental lowest-order $TEM_{0,0}$ mode (more about the meaning of the subscripts in Sec. 3.5). Our goal is to find ψ_0 by reducing the partial differential equation (3.3.1) to ordinary differential equations for the unknown functions $P(z)$ and $q(z)$. Thus the following derivatives are necessary:

$$-j2k\frac{\partial \psi_0}{\partial z} = \left[-2kP'(z) + \frac{k^2r^2q'(z)}{q^2(z)}\right]\psi_0$$

$$\frac{\partial \psi_0}{\partial r} = -j\frac{kr}{q(z)}\psi_0 \qquad \therefore r\frac{\partial \psi_0}{\partial r} = -j\frac{kr^2}{q(z)}\psi_0$$

$$\frac{\partial}{\partial r}\left(r\frac{\partial \psi_0}{\partial r}\right) = -j\frac{kr^2}{q(z)}\left[-j\frac{kr}{q(z)}\right]\psi_0 - j2\frac{kr}{q(z)}\psi_0$$

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \psi_0}{\partial r}\right) = \left[-\frac{k^2r^2}{q^2(z)} - j\frac{2k}{q(z)}\right]\psi_0$$

These functions are substituted into (3.3.1), and the terms with equal powers of r are grouped together:

$$\left\{\left[\frac{k^2}{q^2(z)}(q'(z) - 1)\right]r^2 - 2k\left[P'(z) + \frac{j}{q(z)}\right]r^0\right\}\psi_0 = 0 \tag{3.3.3}$$

For the assumed form (3.3.2) to be a solution, every factor of a power of r must be equal to zero. This yields two simple ordinary differential equations:

$$q'(z) = 1 \tag{3.3.4a}$$

$$P'(z) = -\frac{j}{q(z)} \tag{3.3.4b}$$

Since (3.3.4a) is decoupled from the other, its solution and implications will be discussed first. The solution is trivial

$$q(z) = q_0 + z \tag{3.3.5}$$

where q_0 is the value of q at $z = 0$. (Where is $z = 0$? We must face this question later.) Now, it is obvious the dimensions of q_0 must be the same as z—length—so we are tempted to use the letter symbol z_0 for it. But is q_0 real?

To answer this, we must refer back to the initial form involving $q(z)$:

$$\psi_0 = \exp\left[-j\frac{kr^2}{2q(z)}\right]\exp[-jP(z)] \tag{3.3.2}$$

If $q(z)$ were always real, then $|\exp[-jkr^2/2q(z)]| = 1$ for all values of r. This would mean that the phase is changing faster and faster with r, with the amplitude remaining a constant. That does not describe a beam. It is an absurdity. A beam has most of its energy concentrated in a certain location in the transverse plane, and thus the possibility of q being pure real is not interesting. If $q(z)$ has an imaginary part, then the mathematical exercise becomes worthwhile.

Assume that $q(z)$ is complex. Now z is obviously real in (3.3.5), and any real part of q_0 just corresponds to a shift in the spatial coordinate. Consequently, we might as well start with the proper $z = 0$ axis to absorb this real part and let q_0 be imaginary (i.e., $q_0 = jz_0$):

$$q(z) = z + jz_0 \tag{3.3.6}$$

where z_0 is a constant to be determined (and interpreted). If (3.3.6) is substituted into (3.3.2)—say at $z = 0$—we obtain a very satisfying physical picture for part of ψ_0. At $z = 0$,

$$q(z) = jz_0$$

and

$$\psi_0(z = 0) = \exp\left(-\frac{kr^2}{2z_0}\right)\exp[-jP(z = 0)]$$

Note that the exponential term is real and thus the amplitude drops off quite rapidly with r, being down from its peak value of 1 at $r = 0$ to 0.368 at $r = (2z_0/k)^{1/2}$. Obviously, this last quantity is a *scale length* for this beam.

$$w_0^2 = \frac{2z_0}{k} = \frac{\lambda_0 z_0}{n\pi} \qquad \text{or} \qquad z_0 = \frac{\pi n w_0^2}{\lambda_0} \tag{3.3.7}$$

Thus the field varies as $\exp(-r^2/w_0^2)$ at the plane $z = 0$ (Fig. 3.1). Since our eyes respond to the square of the field, we would observe a single "dot" with a major fraction of the power contained in a beam of radius $\sim w_0$. Because of this, w_0 is called the "spot size." (Actually, it is the minimum spot size, as will be shown.)

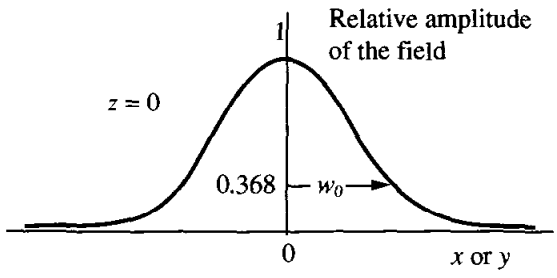


FIGURE 3.1. Variation of the field in the transverse plane.

At any point z, the value of q changes according to (3.3.6). Since we are interested in the inverse of q, let us compute that factor and examine the imaginary part.

$$\frac{1}{q(z)} = \frac{1}{z + jz_0} = \frac{z}{z^2 + z_0^2} - j\frac{z_0}{z^2 + z_0^2} = \frac{1}{R(z)} - j\frac{\lambda_0}{\pi n w^2(z)} \tag{3.3.8}$$

where $R(z)$ and $w(z)$ are characteristic beam parameters to be discussed later.

Again, we return to (3.3.2) for a physical interpretation of this bit of manipulation.

$$\psi_0 = \left\{\exp\left[-\frac{kz_0r^2}{2(z^2 + z_0^2)}\right]\right\}\left\{\exp\left[-\frac{jkzr^2}{2(z^2 + z_0^2)}\right]\right\}\{\exp[-jP(z)]\} \tag{3.3.9}$$

Now, as we move away from the axis, the phase changes faster and faster with r, but the amplitude becomes insignificant at large r. Thus the previous absurdity is avoided.

We again realize that the term multiplying r^2 in the first exponential factor is a scale length, and we name it the spot size of the beam, which is now a function of z:

$$w^2(z) = \frac{2}{kz_0}(z_0^2 + z^2) = \frac{2z_0}{k}\left[1 + \left(\frac{z}{z_0}\right)^2\right]$$

or by using the previous definition of w_0, (3.3.7), we obtain

$$w^2(z) = w_0^2\left[1 + \left(\frac{\lambda_0 z}{\pi n w_0^2}\right)^2\right] \tag{3.3.10}$$

Let us also abbreviate the terms in the second exponential factor by $R(z)$:

$$\begin{aligned} R(z) &= \frac{1}{z}(z^2 + z_0^2) = z\left[1 + \left(\frac{z_0}{z}\right)^2\right] \\ &= z\left[1 + \left(\frac{\pi n w_0^2}{\lambda_0 z}\right)^2\right] \end{aligned} \tag{3.3.11}$$

Both (3.3.10) and (3.3.11) require considerable discussion to appreciate the physical interpretation. This will be done after we have disposed of the remaining bit of mathematics; namely, to find the function $P(z)$. This is related to $q(z)$ by

$$P'(z) = \frac{-j}{q(z)} = \frac{-j}{z + jz_0} \tag{3.3.4b}$$

or

$$jP(z) = \int_0^z \frac{dz'}{z' + jz_0} = ln(z' + jz_0)\Big|_0^z$$

$$= ln(z + jz_0) - ln(jz_0)$$

$$jP(z) = ln\left[1 - j\left(\frac{z}{z_0}\right)\right]$$

Now we use the fact that

$$1 - j\left(\frac{z}{z_0}\right) = \left[1 + \left(\frac{z}{z_0}\right)^2\right]^{1/2} \exp\left[-j \tan^{-1}\left(\frac{z}{z_0}\right)\right]$$

to find the real and imaginary parts of P(z):

$$jP(z) = ln\left[1 + \left(\frac{z}{z_0}\right)^2\right]^{1/2} - j \tan^{-1}\left(\frac{z}{z_0}\right) \tag{3.3.12}$$

We need $\exp[-jP(z)]$:

$$e^{-jP(z)} = \frac{1}{[1 + (z/z_0)^2]^{1/2}} e^{+j \tan^{-1}(z/z_0)} \tag{3.3.13}$$

Thus the complete expression for the fundamental or lowest-order $TEM_{0,0}$ mode is found from the various definitions made previously:

$$\frac{E(x, y, z)}{E_0} = \left\{\frac{w_0}{w(z)} \exp\left[-\frac{r^2}{w^2(z)}\right]\right\} \quad \text{amplitude factor}$$

$$\times \exp\left\{-j\left[kz - \tan^{-1}\left(\frac{z}{z_0}\right)\right]\right\} \quad \text{longitudinal phase} \tag{3.3.14}$$

$$\times \exp\left[-j\frac{kr^2}{2R(z)}\right] \quad \text{radial phase}$$

where

$$w^2(z) = w_0^2\left[1 + \left(\frac{\lambda_0 z}{\pi n w_0^2}\right)^2\right] = w_0^2\left[1 + \left(\frac{z}{z_0}\right)^2\right] \tag{3.3.10}$$

$$R(z) = z\left[1 + \left(\frac{\pi n w_0^2}{\lambda_0 z}\right)^2\right] = z\left[1 + \left(\frac{z_0}{z}\right)^2\right] \tag{3.3.11}$$

$$z_0 = \frac{\pi n w_0^2}{\lambda_0} \tag{3.3.7}$$

3.4 PHYSICAL DESCRIPTION OF $TEM_{0,0}$ MODE

The physical interpretation of (3.3.14), together with various definitions, is the subject of this section. It is hoped that we will be able to penetrate the mathematical maze of the preceding section so as to appreciate the physical simplicity of the results of Sec. 3.3.

3.4.1 Amplitude of the Field

The first term in (3.3.14) describes the amplitude of the field as a function of the radial coordinate and how this changes as the beam propagates along z. It was noted in Sec. 3.3 that at $r = w$ the field is down by e^{-1} of its peak value at $r = 0$.

$$\left|\frac{E(x, y, z)}{E_0}\right| = \frac{w_0}{w(z)} \exp\left[-\left(\frac{r}{w}\right)^2\right] \tag{3.4.1}$$

In Fig. 3.2, the $1/e$ point of the field is sketched as a function of the z coordinate. As the beam propagates along z, the spot size $w(z)$ becomes larger; hence, the $1/e$ points become farther from the axis. It is obvious from this figure that the beam expands from its minimum value of w_0 by a factor of $\sqrt{2}$ when $z = z_0$.*

$$w^2(z) = w_0^2\left[1 + \left(\frac{z}{z_0}\right)^2\right] \tag{3.3.10}$$

Furthermore, the beam has its minimum spot size at a certain point along the z axis (the definition of $z = 0$).

For large z the spot size is asymptotic to the dashed lines described by

$$w(z \gg z_0) = \frac{w_0 z}{z_0} = \frac{\lambda_0 z}{\pi n w_0}$$

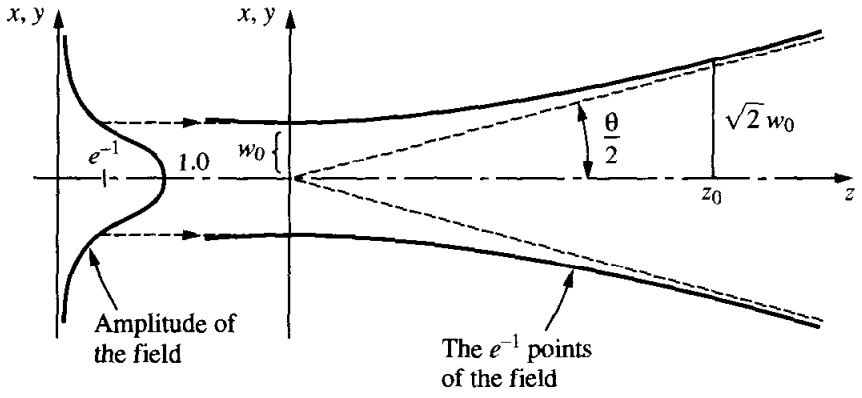


FIGURE 3.2. Spreading of a $TEM_{0,0}$ mode.

*It will become obvious in Chapter 5 that $2 \times z_0$ is the spacing for a confocal mirror arrangement. Hence, $2z_0$ is also called the *confocal parameter*.

Thus the expansion angle of the beam is

$$\frac{\theta}{2} = \frac{dw}{dz} = \frac{\lambda_0}{\pi n w_0}$$
$$\theta = \frac{2\lambda_0}{\pi n w_0} \tag{3.4.2}$$

This is the minimum angular spread that a beam of diameter $2w_0$ can have. (See the discussion on the uncertainty principle in Chapter 1.)

There is a perfectly logical and simple reason why the factor $w_0/w(z)$ appears in the amplitude of the field. To appreciate this, we compute the total power crossing an arbitrary plane z that must be equal to a constant.

$$P = \frac{1}{2}\iint \frac{EE^*}{\eta}\, dA = \frac{1}{2}\frac{E_0^2}{\eta}\frac{w_0^2}{w^2(z)}\int_0^{2\pi}\int_0^{\infty} \exp\left[-\frac{2r^2}{w^2(z)}\right] r\, dr\, d\phi$$
$$= \frac{1}{2}\frac{E_0^2}{\eta}\left(\frac{\pi w_0^2}{2}\right) \tag{3.4.3}$$

where η is the wave impedance $(\mu_0/\epsilon_0 n)^{1/2}$. Note that the power carried by the beam is constant independent of z, a most logical result. It would be embarrassing if it were not true, for we have assumed a lossless medium. Thus the factor $w_0/w(z)$ reduces the peak amplitude of the field in response to the spreading of the beam.

3.4.2 Longitudinal Phase Factor

The second factor in (3.3.14) expresses the change in phase of the wave in the direction of propagation,

$$\phi = kz - \tan^{-1}\left(\frac{z}{z_0}\right) \tag{3.4.4}$$

where k is the customary wave number of a uniform plane wave, $\omega n/c$. Thus the phase velocity of this Gaussian beam is close to, but slightly greater than, the velocity of light (in the uniform medium):

$$v_p \triangleq \left(\frac{\phi}{\omega z}\right)^{-1} = \frac{c/n}{1 - (\lambda_0/2\pi n z)\tan^{-1}(z/z_0)} \tag{3.4.5}$$

Even though the propagation velocity is close to c/n, the difference can be measured and does play a role in resonator theory.

The fact that the wave number is so close to k of a uniform plane wave means that we obtain the magnetic field intensity by the simple relationship

$$\mathbf{H} = \frac{\mathbf{k} \times \mathbf{E}}{\omega\mu_0} = \mathbf{a}_z \times \frac{\mathbf{E}}{\eta} \tag{3.4.6}$$

3.4.3 Radial Phase Factor

The final factor in (3.3.14),

$$\exp\left[-j\frac{kr^2}{2R(z)}\right]$$

indicates that the plane z = constant is *not* an equiphase surface. If we assume $R(z)$ is positive, then the field at $r > 0$ lags the axial value (at $r = 0$). Obviously, if the phase front is not planar, it is curved. Because the letter symbol is used, $R(z)$, we can anticipate that the equiphase surfaces are spherical with a radius of curvature given by $R(z)$.

This is easily shown by considering a limited extent spherical wave as shown in Fig. 3.3 and finding the phase of the wave close to the axis. The field for such a wave would be

$$E \sim \frac{1}{R}\exp(-jkR)$$

where $R = (r^2 + z^2)^{1/2}$. We place ourselves at a large distance from the origin where $R \sim z \gg r$. Thus the binomial theorem is used in a judicious fashion:

$$R = z\left(1 + \frac{r^2}{z^2}\right)^{1/2} \sim z + \frac{1}{2}\frac{r^2}{z} \sim z + \frac{1}{2}\frac{r^2}{R}$$

since $R \sim z$. Hence the phase of the field close to the z axis varies in the following manner:

$$E \sim \frac{1}{R}\exp(-jkz)\exp\left(-j\frac{kr^2}{2R}\right)$$

The last term has the same functional form of the last factor of (3.3.14) and hence its name.

But in the case of a Gaussian beam, the apparent center for the curved wave front changes. Recall the relation for $R(z)$:

$$R(z) = z\left[1 + \left(\frac{z_0}{z}\right)^2\right] \tag{3.3.11}$$

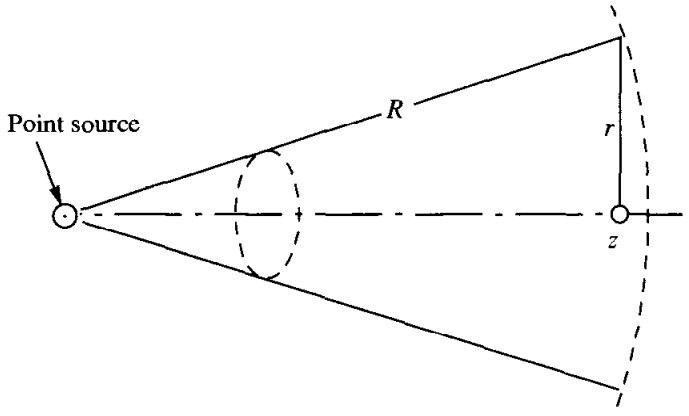


FIGURE 3.3. Origin of the phase front curvature.

Only when z is much larger than z_0 does the beam appear to originate at $z = 0$. As we move closer to the point $z = 0$, the center recedes until at $z = 0$ the "center" of curvature is at infinity and now the wave front is planar. Indeed, it is easy to show that the equiphase surfaces are orthogonal to the beam expansion curves shown in Fig. 3.2.

Thus there are two alternative but equivalent definitions for the plane $z = 0$:

1. Where the spot size is a minimum
2. Where the wave front is planar

Let us now repeat (3.3.14) and assign a brief physical interpretation to each term.

$E(x, y, z)$	The electric field
$\parallel$	
E_0	Amplitude at $z = 0$
$\times$	
$\frac{w_0}{w(z)} \exp\left[-\frac{r^2}{w^2(z)}\right]$	Variation of the amplitude with r
$\times$	
$\exp\left\{-j\left[kz - \tan^{-1}\left(\frac{z}{z_0}\right)\right]\right\}$	Longitudinal phase factor
$\times$	
$\exp\left[-j\frac{kr^2}{2R(z)}\right]$	Radial phase factor

3.5 HIGHER-ORDER MODES

In the previous work, the simplifying assumption of cylindrical symmetry was made. Although this may make the mathematics simple, the laser does not particularly fear (or know about) the complexity. It has no trouble in solving (3.2.8).

Indeed, if we consider the simple gas laser shown in Fig. 3.4, there are trivial reasons why it would not oscillate in the lowest-order mode. For instance, suppose that the window had a streak of "dirt" or "lint" right at the center of the tube, or suppose that the center of the mirror was absent.* If the electric field is as described by (3.3.14), there would be considerable scattering losses owing to the lint and a major coupling loss through the hole. Later, it will be seen that a laser will oscillate in that mode with the highest gain-to-loss ratio. Hence, we must consider possibilities that are not cylindrically symmetric.

*This is called *hole coupling*.

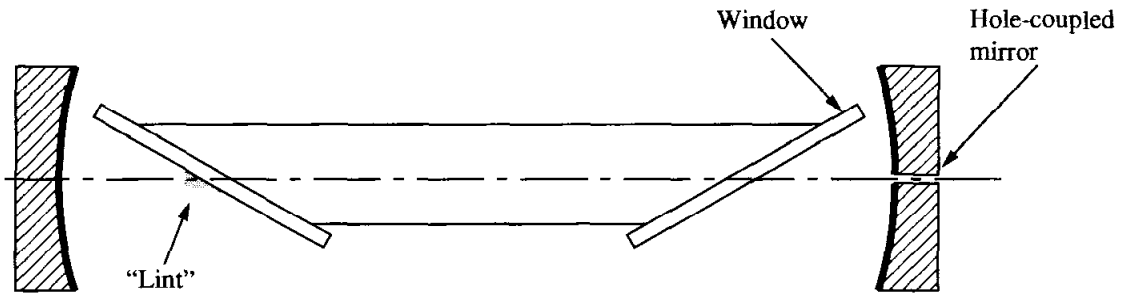


FIGURE 3.4. Simple laser.

We can choose to work in cylindrical (r, ϕ, z) or Cartesian coordinates (x, y, z), allowing for variations in ϕ in the former and different variations in the x and y coordinates in the latter. Different mode descriptions apply for each coordinate system. For the simple system shown in Fig. 3.4, and for most lasers, the Cartesian coordinate system is most appropriate. The reason is that the windows provide a "bias" that discriminates against the purely cylindrical modes.

It can be verified by direct substitution* that the following functions satisfy (3.2.8):

$$\begin{aligned}\frac{E(x, y, z)}{E_{m,p}} &= H_m\left[\frac{2^{1/2}x}{w(z)}\right] H_p\left[\frac{2^{1/2}y}{w(z)}\right] \\ &\quad \times \frac{w_0}{w(z)} \exp\left[-\frac{x^2 + y^2}{w^2(z)}\right] \\ &\quad \times \exp\left\{-j\left[kz - (1 + m + p)\tan^{-1}\left(\frac{z}{z_0}\right)\right]\right\} \\ &\quad \times \exp\left[-j\frac{kr^2}{2R(z)}\right] \end{aligned} \tag{3.5.1}$$

where all symbols, $w(z)$, w_0, z_0, and $R(z)$ are as defined and interpreted previously. The symbol $H_m(u)$, stands for the Hermite polynomial of order m and argument u and is defined by

$$H_m(u) = (-1)^m e^{u^2} \frac{d^m e^{-u^2}}{du^m} \tag{3.5.2}$$

There is a great deal of similarity between (3.5.1) and (3.3.14): the radial-phase factor is the same, the exponential variation with $r^2 = x^2 + y^2$ is the same, and the multiplying factor $w_0/w(z)$ is the same, but there are differences.

First note that the phase shift in the z direction depends on the mode numbers m and p. This will play a role in the oscillation frequency of the laser.

*A very long and painful exercise in arithmetic!

The major change in visible appearance is due to the Hermite polynomials. It is instructive to consider a few of the lower-order ones:

$$H_0(u) = 1 \Rightarrow 1$$

$$H_1(u) = 2(u) \Rightarrow u$$

$$H_2(u) = (2u^2 - 1)2 \Rightarrow 2u^2 - 1$$

where the arrow indicates that we can absorb common numerical factors into an amplitude factor $E_{m,p}$. Thus the field has a more spectacular variation in the transverse plane, as is shown in Fig. 3.5. Note that for large x (or y), the exponential behavior still dominates, and thus the "beam" is tightly bound to the z axis. But at small x, the field is modified considerably by the polynominals being forced to zero at a finite number of points. The number of times that the field goes to zero, other than at ∞, is the mode number m. If the laser is visible, then there will be $m + 1$ dots encountered going across the beam in

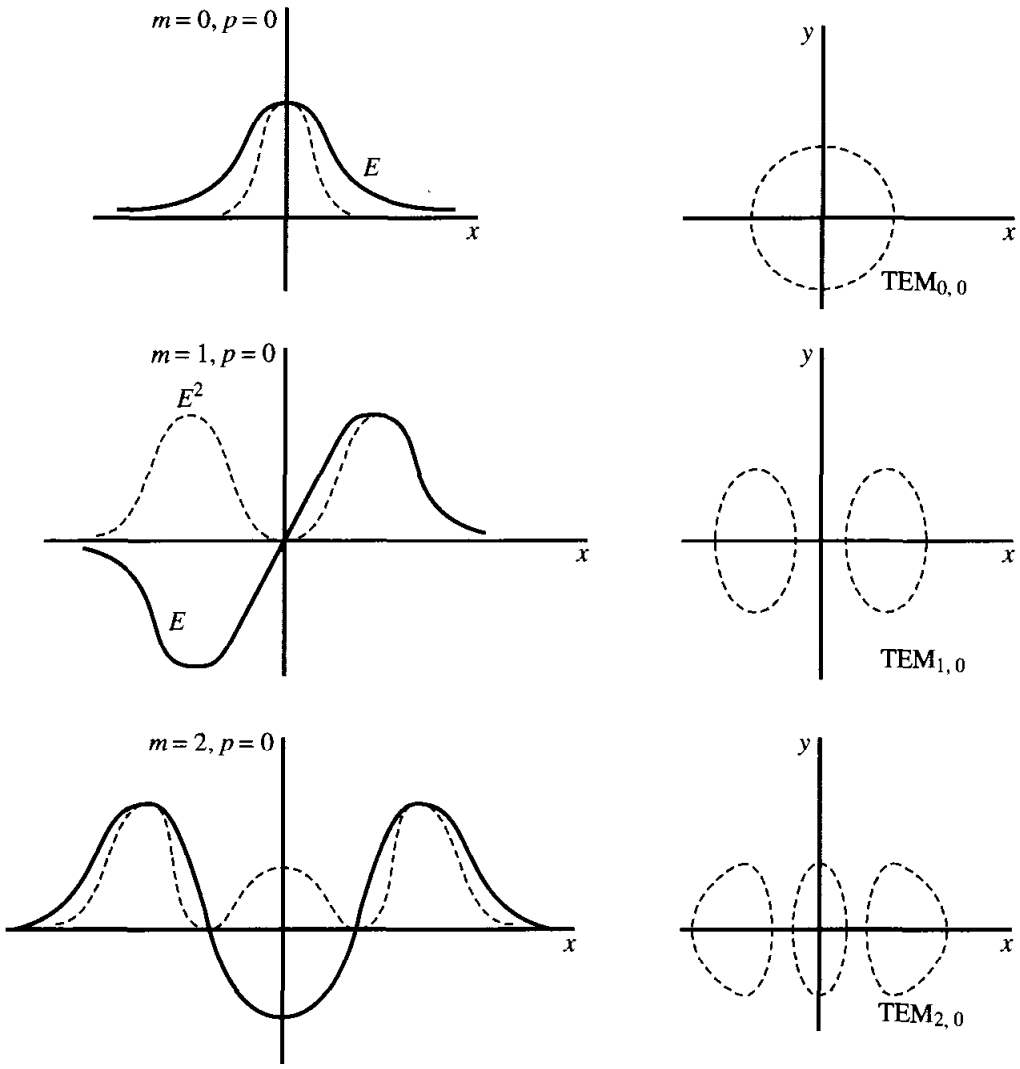


FIGURE 3.5. The field E, intensity E^2, and "dot" pattern of various modes.

the x direction. The same considerations apply to the y direction. Hence there will be $(m + 1)(p + 1)$ "dots" in a $\text{TEM}_{m,p}$ mode.

At this point we should be cautioned that there is a built-in difficulty with the name "spot size" $w(z)$. The spot size $w(z)$ is the ***same*** for all three of the modes illustrated, but the field occupies a bigger area on the paper as the mode number gets larger.

It is a natural tendency to associate the words *spot size* with the radial extent of the beam, but it is wrong to do so if you also associate the same words with $w(z)$. This quantity $w(z)$ is a scale length for measuring the variation of field in the transverse direction. *All* $\text{TEM}_{m,p}$ modes have this same scale length $w(z)$, but the higher-order modes use a larger transverse area.

3.6 *ABCD* LAW FOR GAUSSIAN BEAMS

The $ABCD$ law relates the complex beam parameters, q_2, of a Gaussian beam at plane 2 to the value q_1 at plane 1 by using the $ABCD$ ray matrix.

$$q_z = \frac{Aq_1 + B}{Cq_1 + D} \tag{3.6.1}$$

This is truly an amazing relationship for there is no simple logic sequence leading from rays to the complex beam parameter q, nor is there a simple logic leading to the bilinear transformation format of (3.6.1). No general proof is known to this author and there is no known way of stating that it obviously follows from another equation, but its validity for every known optical component is easily established—one component at a time—and thus is established for any combination.

For instance, the differential equation (3.3.4a) leads to a simple solution relating the output q_2 at a distance z from the input plane

$$q'(z) = 1 \text{ [Eq.(3.3.4a)]} \rightarrow q_2 = q_1 + z \tag{3.6.2}$$

For free space of length z, $A = 1$, $B = z$, $C = 0$, $D = 1$, and (3.6.1) yields precisely the same answer. The complex beam parameter q is most easily interpreted in terms of its reciprocal:

$$\frac{1}{q} = \frac{1}{R} - j\frac{\lambda_0}{\pi n w^2}$$

We can manipulate (3.6.2) to accept and present the information in that format:

$$\frac{1}{q_2} = \frac{C + D(1/q_1)}{A + B(1/q_1)} \tag{3.6.3}$$

If we assume a beam with a minimum spot size w_0 and a planar wave front at $z = 0$ and utilize the $ABCD$ parameters for free space, we recover the expansion law for a Gaussian beam:

$$\frac{1}{q_2(z)} = \frac{0 + 1 \cdot (-j\lambda/\pi w_0^2)}{1 + z \cdot (-j\lambda/\pi w_0^2)} \triangleq \frac{1}{R(z)} - j\frac{\lambda}{\pi w^2(z)} \tag{3.6.4}$$

If we separate (3.6.4) into its real and imaginary parts, we recover the expansion laws (3.3.10) and (3.3.11) directly.

As another example of the veracity and utility of the $ABCD$ law, let us reconsider the beam transformation by a thin lens (as discussed in Sec. 2.13). Assume that a large-diameter Gaussian beam with a *planar wave front* impinges on a thin lens in the manner shown in Fig. 3.6. Equation (3.6.1) would indicate that beam parameter q' to the right of the lens is given by (3.6.3) with the appropriate values for $ABCD$:

$$\frac{1}{q'} = \frac{-1/f + 1 \cdot (1/q_1)}{1 + 0 \cdot (1/q_1)} = -\frac{1}{f} + \frac{1}{q_1} \tag{3.6.5}$$

For air, $n = 1$, and

$$\frac{1}{q_1} = \frac{1}{\infty} - j\frac{\lambda_0}{\pi w_{01}^2} \tag{3.6.6}$$

Hence,

$$\left(\frac{1}{q'} \triangleq \frac{1}{R'} - j\frac{\lambda}{\pi w'^2}\right) = -\frac{1}{f} - j\frac{\lambda_0}{\pi w_{01}^2} \tag{3.6.7}$$

Equation (3.6.7) states that the spot size just to the right of the lens equals that at the left, and the beam appears to be converging toward the focal point f. This latter point is precisely the conclusion of Sec. 2.13, and the equality of spot sizes indicates that power is conserved, a most logical result.

As an example of the utility of the law, we can use it to predict the minimum focal spot size achievable with a lens. The transmission matrix for a lens plus a length of free space z is

$$\mathbf{T} = \begin{bmatrix} 1 - \dfrac{z}{f} & z \\ -\dfrac{1}{f} & 1 \end{bmatrix} \tag{3.6.8}$$

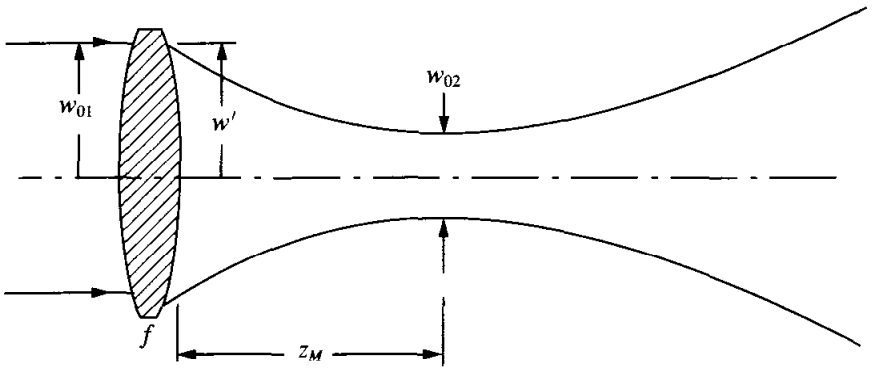


FIGURE 3.6. Focusing of a Gaussian beam by a lens.

Thus the beam parameters at any point z away from the lens are given by

$$\frac{1}{R(z)} = \frac{-1/f + z(1/f^2 + 1/z_{01}^2)}{(1 - z/f)^2 + (z/z_{01})^2} \tag{3.6.9}$$

$$\frac{\lambda_0}{\pi w^2(z)} = \frac{1/z_{01}}{(1 - z/f)^2 + (z/z_{01})^2} \tag{3.6.10}$$

where $z_{01} = \pi w_{01}^2/\lambda_0$ (as usual).

From (3.6.10), we find, much to our surprise, that the spot size does not minimize at $z = f$, but at z_M, given by

$$z_M = \frac{f}{1 + (f/z_{01})^2} \tag{3.6.11}$$

For reasonable sized beams, $z_{01} \gg f$ and thus the minimum occurs at $z \sim f$ in accordance with our intuition. We can use (3.6.11) in (3.6.10) to predict the spot size at the focus

$$\frac{\lambda_0}{\pi w_{02}^2} = \frac{1/z_{01}}{(1 - z_m/f)^2 + (z_m/z_{01})^2} \approx \frac{z_{01}}{f^2} \qquad \text{for} \quad z_{\min} \approx f$$

or

$$w_{02} \approx \frac{\lambda_0 f}{\pi w_{01}} = 3\left[\frac{\lambda_0}{\pi}\frac{f}{(2 \cdot 1.5 w_{01})}\right] \tag{3.6.12}$$

provided $$z_{01} = \frac{\pi w_{01}^2}{\lambda_0} \gg f.$$

Thus, to obtain the smallest spot size at the focal plane, we require that the incoming beam have a large spot size, which is also consistent with the assumption of $z_{01} \gg f$. The clear aperture of the lens (i.e., the clear diameter) must be at least twice $1.5 \times w_{01}$ so as to intercept 99% of the incoming intensity, the reason for the factor of $3 \div 3$ in (3.6.12).

Hence, (3.6.12) can be re-expressed in terms of the lens f number ($f^{\#}$ equals focal length divided by diameter).

$$w_{02} = 3\frac{\lambda_0}{\pi} f^{\#} \sim \lambda_0 \cdot f^{\#} \tag{3.6.13}$$

There are a couple of problems with (3.6.13): We have assumed an infinitesimally thin lens without aberrations,* one that does not exist.

Let us remove the assumption that the incoming beam had a planar wave front (i.e., $R_1 = \infty$), and examine the beam just as it emerges from the lens. According to the $ABCD$ law, we have (3.6.3)

$$\frac{1}{q_2} = \frac{C + D(1/q_1)}{A + B(1/q_1)}$$

But

$$\frac{1}{q_1} = \frac{1}{R_1} - j\frac{\lambda_0}{\pi w_1^2} \qquad C = -1/f \qquad A = 1 \qquad B = 0 \qquad D = 1$$

*The Melles-Griot optics catalog (p. 342) recommends multiplying (3.6.12) by 4/3.

Thus

$$\frac{1}{q_2} = \frac{-1}{f} + \frac{1}{R_1} - j\frac{\lambda_0}{\pi w_1^2} \tag{3.6.14}$$

Thus the thin lens keeps the spot size the same and therefore conserves power (thank heavens) and changes the radius of curvature of the incoming beam to

$$\frac{1}{R_2} = \frac{1}{R_1} - \frac{1}{f} \tag{3.6.15}$$

Note that if $1/R_1 > 1/f$, the lens does *not* focus the beam!

Let us leave the thin lens and turn to the last type of special case for the $ABCD$ law, that of a continuous lens with a parabolic index of refraction $n(r)$. This is reserved for a problem at the end of the chapter. Only the procedure is indicated here.

We use the square of (2.12.7) in the wave equation,

$$\nabla_t^2 E + \frac{\partial^2 E}{\partial z^2} + \frac{\omega^2}{c^2} n^2(r) E = 0 \tag{3.6.16a}$$

or

$$\nabla_t^2 E + \frac{\partial^2 E}{\partial z^2} + \frac{\omega^2}{c^2} n_0^2 \left(1 - \frac{r^2}{l^2}\right) E = 0 \tag{3.6.16b}$$

and then proceed to rederive every equation of Chapter 3 from (3.2.6) to (3.3.14). Unless you lose your mind in the mathematical maze of this operation, you will have verified the $ABCD$ law for a continuous lens.

Thus (3.6.1) is a simple and compact way of describing the evolution of a Gaussian beam in an optical system. We will see other examples of its usefulness in the next two chapters.

3.7 DIVERGENCE OF THE HIGHER-ORDER MODES: SPATIAL COHERENCE

An example will illustrate why we should be very careful about the use of the term *spot size*. Let us ask, what is the divergence of a beam consisting of one or more $\text{TEM}_{m,p}$ modes? For instance, we can obtain a rather large physical "spot" by a linear combination of these modes.

The answer is that *all* Hermite-Gaussian beams have *exactly* the same divergence (or far-field angle), given by

$$\theta = \frac{2\lambda_0}{\pi n w_0} \tag{3.7.1}$$

where w_0 is the minimum spot size for the $\text{TEM}_{0,0}$ mode.

Thus the controlling factor on the beam spread is the characteristic dimension w_0 and not the physical spot seen on the wall. If the beam spread is a factor in the application of

the laser,* we attempt to ensure that oscillation takes place in the $TEM_{0,0}$ mode. Thus this mode has the greatest intensity (power/area) for the minimum beam spread as compared to all other modes or field distributions.

Note that the $TEM_{0,0}$ mode has a uniphase surface, albeit curved but still the field is in phase on this spherical surface. The term *spatial coherence* is used to describe this fact; that is, the field has one common phase on this spherical surface. For contrast, note that the field of the $TEM_{1,0}$ mode reverses direction for negative x; and for the higher-order modes, the field reverses direction many times. Within each dot, the equiphase surface has the same spherical curvature as the $TEM_{0,0}$ mode.

This also explains why a flashlight beam spreads so much faster than a laser beam, even with the parabolic reflector and large aperture on the former. The atoms in the heated filament of the tungsten wire radiate an incoherent wave; that is, the phase from one group of atoms bears little, if any, relationship to another group. Consequently, we cannot, by any stretch of imagination, identify the spot from a flashlight as being a uniphase surface. The characteristic dimension corresponding to w_0 is much, much smaller than the physical size of the parabolic reflector; hence, its divergence is quite large compared to a laser.

PROBLEMS

3.1. The following questions are intended as a review and to test your understanding and appreciation of the Hermite-Gaussian beam modes. Answer these questions with a sketch, some simple mathematics, or a few sentences:

(a) What is the physical significance of the *distance* z_0?

(b) If $z = z_0$ and $r^2 = w^2(z_0)$, by how much does the phase of the field lead or lag that at $r = 0$?

(c) Which factor expresses the idea that the beams are not plane waves and the phase velocity is greater than c?

3.2. **(a)** A certain commercial helium/neon laser is advertised to have a farfield divergence angle of 1 milliradian at $\lambda_0 = 632.8$ nm. What is the spot size w_0?

(b) The power emitted by this laser is 5 mW. What is the peak electric field in volts per centimeter at $r = z = 0$?

(c) How many photons per second are emitted by this laser beam?

(d) Electromagnetic energy can only come in packages of $h\nu$. If one more photon per second were emitted by this laser, what is the new power specification? (The point of this part of the problem is to recognize that there is a time and a place for making the distinction between a classical field and a photon: Should we start here?)

3.3. Given a 1-W $TEM_{0,0}$ beam of $\lambda_0 = 514.5$ nm from an argon ion laser with a minimum spot size of $w_0 = 2$ mm located at $z = 0$

*The beam spread is always a consideration. For a laser transit, laser radar, and laser communications with free-space transmission, we desire a *minimum* beam spread. But these same considerations all apply to focusing. The smallest spot size achievable by a lens is also controlled by w_0. Thus, beam spread is a factor in raw power applications.

(a) How far will this beam propagate before the spot size is 1 cm?
(b) What is the radius of curvature of the phase front at this distance?
(c) What is the amplitude of the electric field at $r = 0$ and $z = 0$?

3.4. A 10-W argon ion laser oscillating at 4880 Å has a minimum spot size of 2 mm.
(a) How far will this beam travel before the spot size is 4 mm?
(b) What fraction of the 10 W is contained in a hole of *diameter* $2w(z)$?
(c) Express the frequency/wavelength of this laser in eV, nm, μm, ν(Hz), and $\bar{\nu}(\text{cm}^{-1})$.
(d) What is the amplitude of the electric field when $w = 1$ cm?

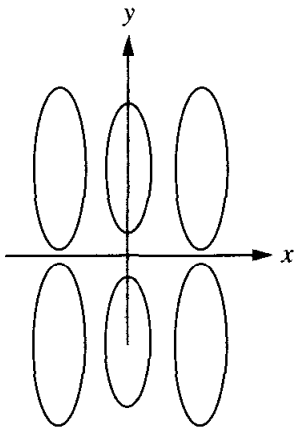
3.5. Sketch the variation of the intensity with x ($y = 0$) of a beam containing 1 W of power in a $\text{TEM}_{0,0}$ mode and 1/2W in the $\text{TEM}_{1,0}$ mode (i.e., total power=1.5W). There are two possibilities: (1) The frequencies and the phases of the two modes are the same, in which case we should add the fields and then square to obtain the intensity, or (2) the phase changes with respect to time. If this change is fast with respect to the observation time, then we should add the intensities.

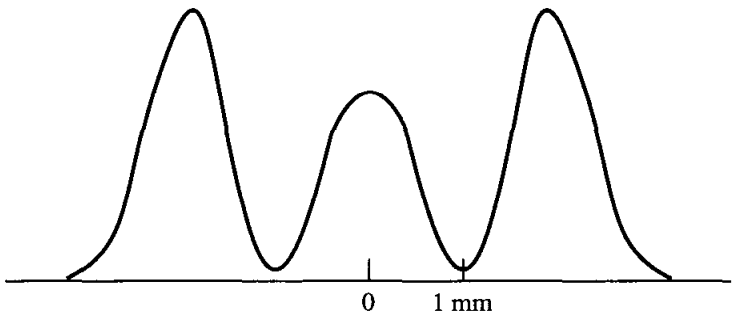
3.6. Consider a linear combination of two equal amplitude $\text{TEM}_{m,p}$ modes given by:

$$\mathbf{E} = E_0 \left\{(\text{TEM}_{1,0})\mathbf{a}_y \pm j(\text{TEM}_{0,1})\mathbf{a}_x\right\}$$

(a) Sketch the "dot" pattern or equal intensity contours for each component (i.e., $\mathbf{a}_x$ or $\mathbf{a}_y$). Indicate the direction of the electric field.
(b) Sketch the pattern for the linear combination.
(c) Label the positions where the intensity is a maximum and a minimum. (This is sometimes referred to as the "donut mode" or $\text{TEM}^*_{0,1}$.

3.7. The intensity of a laser has the following visual appearance when projected on a surface.
(a) Name the mode (i.e., $\text{TEM}_{m,p}$; $m = ?$; $p = ?$).
(b) A plot of the relative *intensity of another* mode as a function of x (for $y = 0$) is shown below at the right. The variation with respect to y is a simple bell-shape curve. What is the spot size w?





3.8. Suppose that a $\text{TEM}_{m,p}$ mode impinged on a perfectly absorbing plate with a hole of radius a centered on the axis of the beam. Plot the transmission coefficient of this

hole as a function of the ratio of a/w for the (0,0), (0,1), and (1,1) modes assuming that the fields are not affected by the plate.

3.9. Show that the Hermite-Gaussian beam modes are orthogonal in the following sense:

$$Re\left[\int (\mathbf{E}_{m,n} \times \mathbf{H}^*_{p,q}) \cdot d\mathbf{S}\right] = 0$$

3.10. Repeat the analysis from (3.2.6) to (3.3.14) for the case where the index of refraction is nonuniform and is given by

$$n(r) = n_0(1 - r^2/2l^2)$$

3.11. The same arguments advanced for the derivation of (3.2.4) can be used for the magnetic field intensity H. Check the accuracy of this equation by considering a dominant $TE_{1,0}$ mode in a rectangular waveguide of width a and height b and computing the ratio of H_z/H_x.

3.12. The news media has shown the astronauts placing laser retroreflectors on the moon. Use the expansion law for Gaussian beams to predict the diameter of a laser beam when it hits the moon. Use $\lambda_0 = 6943$ Å. Consider two cases:

(a) A laser rod of 2 cm diameter.

(b) This same laser sent through a telescope backward so that the beam starts with a diameter of 2 m.

(c) Eye damage intensities are in the range of 10 μW/cm^2. If the laser on earth produced a pulse power of 10 MW, was there danger to the astronauts from the optical radiation?

3.13. Verify the $ABCD$ law for a continuous lens by starting with (3.6.16b) and following the analysis of Sec. 3.1 through 3.3.

3.14. A convenient, if oversimplified, definition of a focal length of a lens is that it converges a parallel beam of light to a point. But if the spot size at the focus were zero, as implied by a point, the expansion of the beam would be infinitely fast and by symmetry would also correspond to its convergence, both statements being obvious contradictions. Use a simple geometric argument based on the convergence (and expansion) to estimate the minimum spot size in the focal region of a lens. Compare with the exact answer.

3.15. Suppose that a Gaussian beam with w =2 cm and a planar wave front impinges on a lens of focal length $f = 4$ cm ($\lambda_0 = 1.0\mu$m).

(a) If $z = 0$ is the location of the lens, where does the output beam reach its minimum spot size?

(b) What is the far-field expansion angle?

3.16. Repeat the analysis of Sec. 3.1 and 3.3 for a medium in which the dielectric constant is complex and depends on r in the following manner:

$$\epsilon(r) = \epsilon_0\left[\epsilon' - j\epsilon''\left(1 - \frac{r^2}{l_G^2}\right)\right]$$

The term ϵ'' can be positive or negative corresponding to gain or loss, but in any case, it is much less than ϵ' and the scale length l_G is much larger than r.

(a) If $\epsilon'' > 0$, does the medium have gain or loss?

(b) With gain or loss, the amplitude of the field does not remain constant with z. Hence, we assume a solution of the form

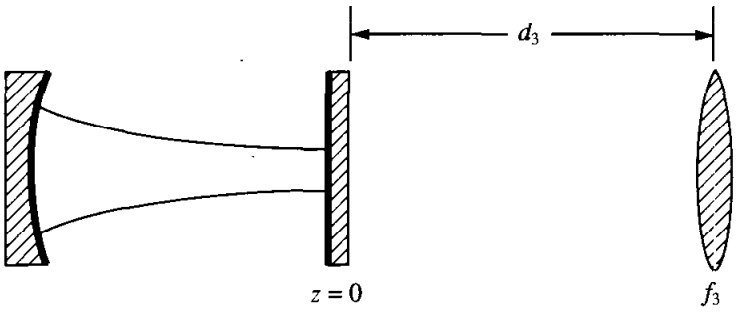
$$E = E_0\psi(x, y, z)\exp\left[\frac{\gamma}{2} - j\frac{\omega}{c}(\epsilon')^{1/2}\right]z$$

(1) What is a logical choice for the relationship between the gain coefficient γ and $\epsilon''(r)$?

(2) What is the differential equation for the wave function ψ?

(c) It is possible to obtain an amplified beam profile in such a medium. Find that beam and relate it to the parameters of the media.

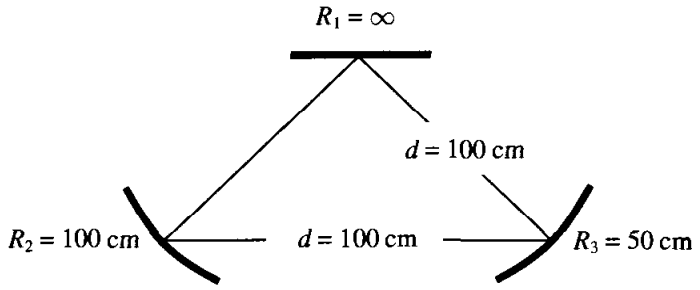
3.17. The laser cavity shown below produces a $TEM_{0,0}$ mode with $z = 0$ located at the flat mirror and its output impinges on a lens of focal length f_3. Assume w_0 is known (0.5 mm), $\lambda_0 = 6328$ Å, $d_3 = 1$ m, and $f_3 = 0.25$m.



(a) What are the spot size and radius of the curvature of the wave impinging on f_3?

(b) What is the radius of curvature after passage through f_3?

3.18. Is the cavity shown below stable? Demonstrate the logic of your answer by (a) constructing a unit cell starting at the flat mirror, (b) finding the $ABCD$ matrix for that cell, and (c) applying the stability criteria. (d) What are the circumstances under which the quantity $[AD - BC]$ can be different from 1? Why is $AD - BC$ always equal to 1 for a cavity?



3.19. The $ABCD$ matrix for a flat mirror with uniform reflectivity is trivial with $A = D = 1$ and $B = C = 0$. This problem concerns a nonuniform flat mirror with a reflectivity that is "tapered" with radius

$$\Gamma^2(r) = \exp[-(tr)^2]$$

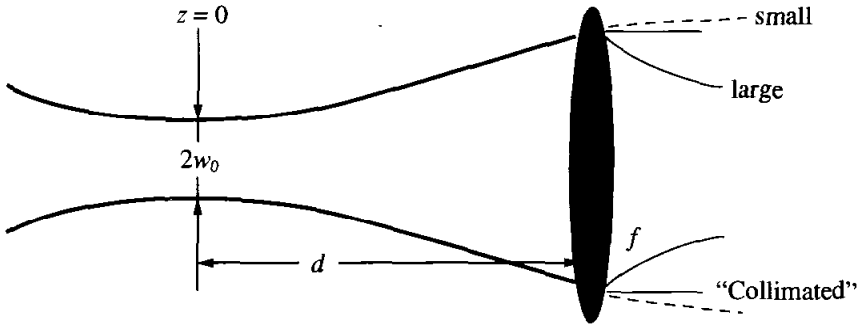
Assume that a $TEM_{0,0}$ Gaussian beam impinges on such a mirror (with the axis of the beam corresponding to the axis of the mirror). Find a new $ABCD$ matrix for this tapered mirror such that the $ABCD$ law still applies for this element.

$$q_2 = \frac{Aq_1 + B}{Cq_1 + D}$$

where $q_{1,2}$ is the complex beam parameter of the incident and the reflected wave, respectively. (NOTE: Your answer must reduce to the usual one when $t = 0$. Do not waste your time tracing rays.)

3.20. Sketch the dot pattern (i.e., contours of the intensity as a function of x/w and y/w) that would be observed from a laser oscillating in the $TEM_{3,2}$ mode. Label the relative coordinates where the electric field goes to zero.

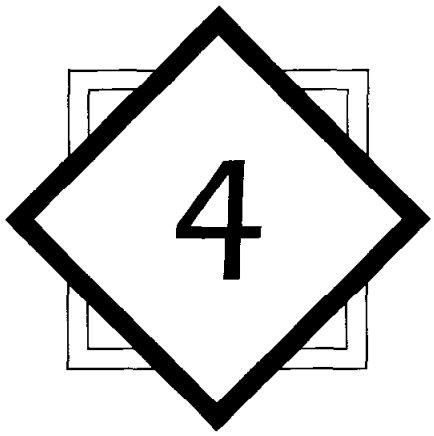
3.21. A focused Gaussian beam reaches its minimum spot size w_0 at $z = 0$ where $R = \infty$ and then propagates to a thin lens of focal length f located a distance d from $z = 0$. If w_0 is large, then the beam exiting the lens will be focused. If it is too small, then the lens merely reduces the far field spreading angle. Find the critical value of w_0 such that the output beam is "collimated"; that is, $R(z = d^+) = \infty$ also.



REFERENCES AND SUGGESTED READINGS

1. G.D. Boyd and J.P. Gordon, "Confocal Multimode Resonator for Millimeter through Optical Wavelength Masers," *Bell Syst. Tech. J.* 40, 489–508, Mar. 1961.
2. G.D. Boyd and H. Kogelnik, "Generalized Confocal Resonator Theory," *Bell Syst. Tech. J.* 41, 1347–1369, July 1962.
3. H. Kogelnik, "Imaging of Optical Modes—Resonators with Internal Lenses," *Bell Syst. Tech. J.* 44, 455–494, Mar. 1963.
4. H. Kogelnik and T. Li, "Laser Beams and Resonators," *Appl. Opt. 5*, 1550–1556, Oct. 1966.
5. A. Yariv, *Introduction to Optical Electronics, 2nd ed. (New York: Holt, Rinehart and Winston, 1976), Chap. 3.*

6. D.R. Herriott, "Applications of Laser Light," *Sci. Am. 219*, 141–156, Sept. 1965.
7. A. Maitland and M.H. Dunn, *Laser Physics* (Amsterdam: North-Holland, 1969), Chaps. 4–7.
8. H.A. Haus, *Waves and Fields in Optoelectronics* (Englewood Cliffs, N.J.: Prentice Hall, 1984).
9. A.E. Siegman, *Lasers* (Mill Valley, Calif.: University Science Books, 1986), Chaps. 16–21.



Guided Optical Beams

4.1 INTRODUCTION

While free space propagation of optical beams has the advantage of being "free," there are some obvious limitations not the least of which is the weather. Communications channels using guided beams are impervious to such limitations and, for the common silica fiber links, have extremely low loss, are immune to electromagnetic interference, are private with virtually no cross talk between adjacent fibers, and are small and flexible.*

The determination of the electromagnetic modes of a fiber requires a head-on attack by using Maxwell's equations and the boundary conditions. We will handle only two very simple and tractable cases here, but the results are similar to all cases; that is, there is a "beam-like" distribution of the fields guided by the index variation of the fiber yielding a phase constant β which depends on ω (or λ), the index of refraction (which is also a function of ω), and the geometry of the fiber.

A major data rate limitation associated with long distance communications using fiber waveguide is identified along with a possible solution; that is, solitons. This last issue is comparatively new and might be considered a research topic. The pay off appears to be very important, however.

*A silica fiber is extremely strong when compared with an equal-sized piece of metal such as steel. But the cross-sectional area is small. Hence, the fiber can be broken.

4.2 OPTICAL FIBERS AND HETEROSTRUCTURES: A SLAB WAVEGUIDE MODEL

Most common fibers are small ($\leq 100\mu$m diameter) and round (some are elliptical) with an index of refraction decreasing with radius in discrete steps from the center core to the cladding and to the outer protective sheath. The technical objective is to transmit information in the form of optical energy over a long path (km $\rightarrow$ 1000 km) with minimum loss at the maximum rate. The exact analysis of such an optical transmission line requires considerable dexterity with Bessel functions that arise naturally in systems whose boundaries are described by cylindrical coordinates.

However, most of the physics of wave guidance is contained in the much simpler slab waveguide model shown in Fig. 4.1, which has the same variation of the index with x as along the diameter of a round fiber. (In a round fiber, there are angular modes that can be represented by rays twisting around the z axis owing to reflections at the core-cladding interface, but not passing through $r = 0$. These are not represented by the one-dimensional slab model.) Furthermore, the slab geometry is a very good representation of the active region of a double heterostructure semiconductor laser, which appears to be well on its way to becoming the dominant source for low power applications.

4.2.1 Zig-Zag Analysis

The analysis starts by considering the slab model of the fiber shown in Fig. 4.1, with the central core having a slightly larger index of refraction. This is usually accomplished by adding a dopant (say GeO_2) to the SiO_2. A major problem is to excite the electromagnetic wave inside the fiber, and a simple suggestion for one way of doing so is shown here. A lens is used to collect and focus the beam into the core. This would excite various combinations of plane waves (at different angles) that undergo total internal reflection at the core-cladding interface and thus follow a "zig-zag" path while advancing along the z direction. Our goal

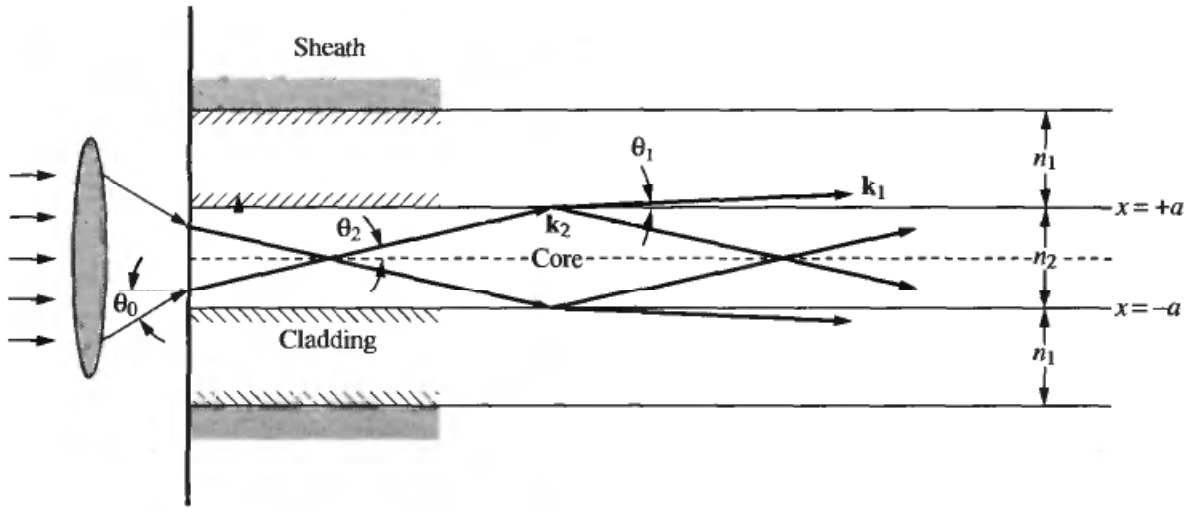


FIGURE 4.1. The slab model of a fiber showing one way of coupling energy into it.

is to find the field configuration or mode that is confined mostly in the core with negligible amplitude extending to the cladding–sheath interface.

To show that such modes exist, it is sufficient to demonstrate that a wave experiences total internal reflection if θ_2 is small enough. This is easily shown by matching the tangential phase constants of the waves in the core and the cladding (a rederivation of Snell's law):

$$\left.\begin{aligned} \frac{\omega}{c} n_2 \cos\theta_2 &= \frac{\omega}{c} n_1 \cos\theta_1 \\ \text{or} \qquad \cos\theta_1 &= \frac{n_2}{n_1} \cos\theta_2 \end{aligned}\right\} \tag{4.2.1}$$

Obviously, if θ_2 is small enough, $\cos\theta_2 \sim 1$, and if $n_2 > n_1$, then $\cos\theta_1 > 1$. What does such an apparent mathematical absurdity mean?

It means that wave is entirely confined to the core and that there is an attenuation of the field amplitude in the x direction but no power loss along z. To demonstrate this last statement, consider a wave propagating in the θ_1 direction in the cladding material:

$$\left.\begin{aligned} \mathbf{E}_1 &= \mathbf{E}_0 \exp(-j\mathbf{k}_1 \cdot \mathbf{r}) \\ \text{where} \qquad \mathbf{k}_1 &= \frac{\omega}{c} n_1(\cos\theta_1 \mathbf{a}_z + \sin\theta_1 \mathbf{a}_x) \end{aligned}\right\} \tag{4.2.2}$$

We use (4.2.1) to express k_1 in terms of the angle θ_2. Now, since $\sin^2\theta + \cos^2\theta = 1$ for all angles—even complex ones—we find that $\sin\theta_1$ must be an imaginary number.

$$\sin\theta_1 = (1 - \cos^2\theta_1)^{1/2} = \pm j\left[\left(\frac{n_2}{n_1}\cos\theta_2\right)^2 - 1\right]^{1/2} \tag{4.2.3}$$

Are we compounding the absurdity by insisting on an angle that yields an imaginary value for the sine? No, the angle, θ_1, is useful only insofar as it describes the field, and the field is perfectly well behaved and not at all mysterious.

$$\begin{aligned} \mathbf{E}_1 = \mathbf{E}_0 \exp &\left\{-\frac{\omega}{c} n_1 \left[\left(\frac{n_2}{n_1}\cos\theta_2\right)^2 - 1\right]^{1/2} (x - a)\right\} \\ &\times \exp\left[-j\frac{\omega}{c} n_1 \left(\frac{n_2}{n_1}\cos\theta_2\right) z\right] \end{aligned} \tag{4.2.4}$$

Thus the field in region $x > a$ decays exponentially away from the interface. However, there is no power flow in the x direction, and hence the energy must be guided by the central core, with a minimal amount contained in the cladding and almost nothing in the sheath. Thus we have achieved the goal of guiding the beam by the slab.

A sketch of the variation of the index of refraction and the field for this slab waveguide is shown by the dashed curves in Fig. 4.2. Note that the field "looks" like that of the Gaussian beam mode of Chapter 3. The fact that the mathematical expression is trigonometric for $|x| < a$, exponentially decaying for $|x| > a$, and not Gaussian [$\sim \exp(-x^2)$] is irrelevant; that is, the field looks like that of a beam.

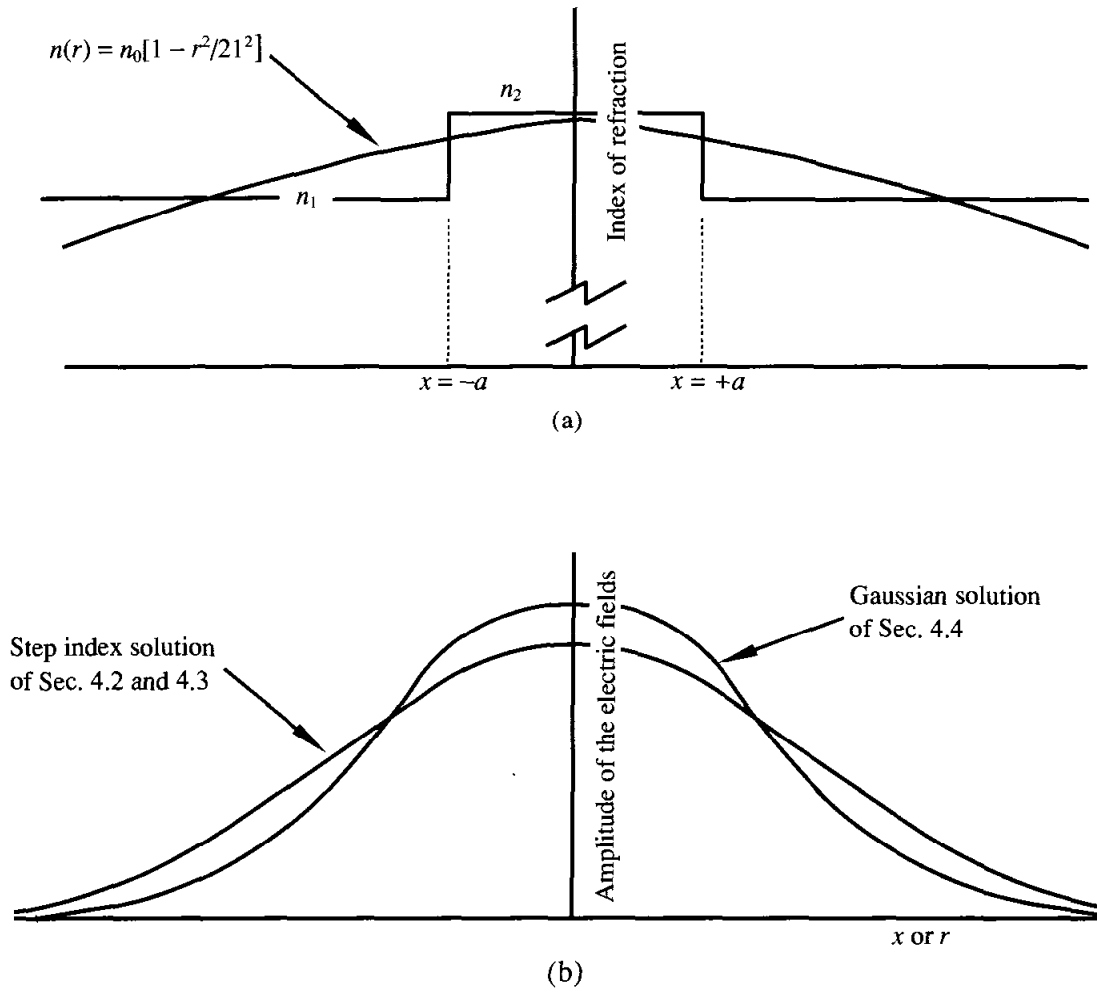


FIGURE 4.2. Variation of (a) the index and (b) the field within a fiber. The shaded curves correspond to each other, as do the solid curves.

It should be remembered, however, that this guidance of the beam was caused by the dielectric discontinuity at $x = \pm a$, in particular the decrease of $n(r)$ there. Thus if we go one step farther and make the index a continuously decreasing function of r [as shown by the shaded curve in Fig. 4.2(a)], we can anticipate guidance there also. In fact, we should not be surprised if the field configuration turned out to be a Hermite-Gaussian beam mode. Section 4.4 will show that this anticipated result is correct.

4.2.2 NUMERICAL APERTURE

Let us return to the slab fiber of Fig. 4.1 and examine the input conditions. If, for instance, the angle θ_2 is too big, then the angle θ_1 is real, and the wave merely propagates outward through the cladding to be absorbed in the sheath or radiated (to $x = \pm\infty$). Such waves are not guided, and the whole purpose of this fiber construction is defeated.

The angle θ_2 is, of course, determined by the air-core interface problem, which is shown in greatly exaggerated form in Fig. 4.1. To have a *guided* wave, we need the sine of the angle θ_1 to be imaginary and $\cos\theta_1 > 1$ in accordance with (4.2.3). However, this is only true provided θ_2 is small enough.

$$\cos\theta_2 < \frac{n_1}{n_2} \tag{4.2.5}$$

On the other hand, θ_2 is controlled by the angle θ_0 and applying Snell's law (again) for the air–n_2 interface yields

$$\frac{\omega}{c}\sin\theta_0 = \frac{\omega n_2}{c}\sin\theta_2 \tag{4.2.6}$$

Combining (4.2.5) and (4.2.6) yields the maximum angle over which the fiber will collect and *guide* the electromagnetic radiation:

$$\sin\theta_0 < n_2[\sin\theta_2 = (1 - \cos\theta_2)^{1/2}]$$

or

$$\sin\theta_0 < (n_2^2 - n_1^2)^{1/2} = NA \tag{4.2.7}$$

The angle θ_0 is the acceptance angle of the fiber and, because $n_2 - n_1$ is quite small for most fibers, the angle is also quite small. The quantity $[n_2^2 - n_1^2]^{1/2}$ is usually referred to as the numerical aperture (NA) in analogy to the corresponding $f\#$ associated with a lens.

While Fig. 4.1 implies that we could use a large-diameter lens to collect a lot of light and focus it onto the core of the fiber, it is quite futile (useless) to do so. That fraction of the radiation contained in angles beyond the limit specified by (4.2.7) is not guided by the fiber.

4.3 MODES IN A STEP-INDEX FIBER (OR A HETEROJUNCTION LASER): WAVE EQUATION APPROACH

Even though most fibers are circular in cross section, we restrict our attention to the symmetric slab geometry shown in Fig. 4.1. This ploy enables us to obtain a formal solution to Maxwell's equations without endless haranguing about the marvels of Bessel functions that arise naturally in cylindrical coordinates. All of the mathematical steps and many of the conclusions of the slab are directly applicable to the round fiber. Furthermore, such a model is a good representation of the active region of a heterojunction* laser. It is a fortuitous fact of nature that the substitution of aluminum for gallium in a GaAs crystal, indicated by $Al_xGa_{1-x}As$, leads to a material with an *increased* bandgap, a *decreased* index of refraction, and, most importantly, almost *identical* lattice constants. Thus the central region of Fig. 4.1 can be assigned to GaAs, and the outer regions to $Al_xGa_{1-x}As$ of a $p - n$ junction laser. There will be much more on the physics of such lasers in Chapter 11, but, for now, our focus is on the electromagnetic problem of guiding waves along this slab.

*A heterojunction uses different materials with different bandgaps in contrast to the junctions in elementary semiconductors silicon or germanium. The latter are called homojunctions.

If we solve the wave equation for the z components, E_z or H_z, then the transverse components for any guide can be found from

$$\mathbf{E}_t = \frac{1}{\beta^2 - (\omega n/c)^2}\left\{j\beta\nabla_t E_z - j\omega\mu_0 \mathbf{a}_z \times \nabla_t H_z\right\} \tag{4.3.1a}$$

$$\mathbf{H}_t = \frac{1}{\beta^2 - (\omega n/c)^2}\left\{j\omega\epsilon_0 n^2 \mathbf{a}_z \times \nabla_t E_z + j\beta\nabla_t H_z\right\} \tag{4.3.1b}$$

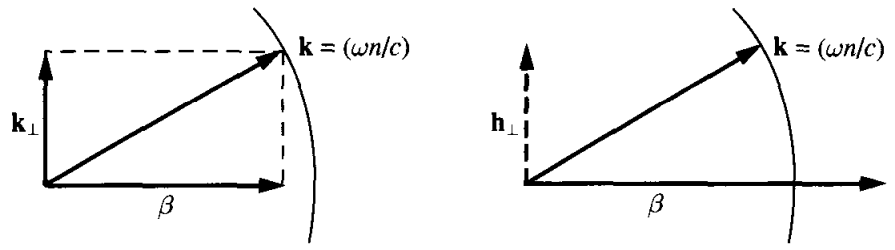
where the fields are assumed to be propagating along z as $\exp(-j\beta z)$. From (4.3.1), we see that there is a natural classification of the types of modes depending on whether H_z or E_z is zero for TE or TM, respectively.* Both may not be zero, or a TEM mode, unless the denominator of (4.3.1) vanishes. Sometimes the boundary conditions require the presence of both E_z and H_z, and these describe the hybrid EH or HE mode depending on which component is dominant. In any case these longitudinal components obey the scalar wave equation:

$$\nabla_t^2 E_z + \left[\left(\frac{\omega n}{c}\right)^2 - \beta^2\right] E_z = 0 \tag{4.3.2a}$$

$$\nabla_t^2 H_z + \left[\left(\frac{\omega n}{c}\right)^2 - \beta^2\right] H_z = 0 \tag{4.3.2b}$$

Note that the character of the solution to (4.3.2) changes depending on whether $[(\omega n/c)^2 - \beta^2]$ is greater than or less than zero. If the former is true, then the equation "looks" like that of a simple harmonic oscillator and we would anticipate a trigonometric or "standing" wave type of solution in the transverse plane. If $[(\omega n/c)^2 - \beta^2] < 0$, then an exponential solution decaying away from the central slab is appropriate. This is, of course, exactly what was found in Sec. 4.2 for fields outside the core.

This idea can be made more formal by considering the Pythagorean relation for components of the wave vector as shown in Fig. 4.3. The wave vector is always related to



(a) $\beta < \omega n/c$ (b) $\beta > \omega n/c$

FIGURE 4.3. The Pythagorean relation for the components of the wave vector.

*Some authors name the mode after the nonvanishing component of the axial field. Thus a TE mode is also referred to as an H mode and a TM as an E mode.

the material properties of the medium by $\mathbf{k} \cdot \mathbf{k} = (\omega n/c)^2$ which holds even if the transverse projection is imaginary.

For the case illustrated in Fig. 4.3(a), $\beta < \omega n/c$ and thus $k_\perp$ can be real but constrained by

$$\mathbf{k} \cdot \mathbf{k} = (\omega n/c)^2 = [\beta \mathbf{a}_z + k_\perp \mathbf{a}_t] \cdot [\beta \mathbf{a}_z + k_\perp \mathbf{a}_t] = \beta^2 + k_\perp^2$$

$$\mathbf{k}_\perp^2 = (\omega n/c)^2 - \beta^2$$

If β is larger than $\omega n/c$ as indicated in Fig. 4.3(b), then the component of $\mathbf{k}$ in the transverse direction must be imaginary (exponentially growing or decaying) in order for the Pythagorean relation to hold:

$$\mathbf{k} \cdot \mathbf{k} = (\omega n/c)^2 = [\beta \mathbf{a}_z \pm j h_\perp \mathbf{a}_t] \cdot [\beta \mathbf{a}_z \pm j h_\perp \mathbf{a}_t] = \beta^2 - h_\perp^2$$

$$h_\perp^2 = \beta^2 - (\omega n/c)^2$$

The analysis of Sec. 4.2 incorporated the above ideas, and it is useful to keep it in mind for the material below.

If the purpose of the structure such as that shown in Fig. 4.4 is to *guide* an electromagnetic mode by region 2, then a trigonometric solution [Fig. 4.3(a)] is appropriate for region 2, and exponentially decaying away from the central region [Fig. 4.3(b)] applies to regions 1 and 3. In any case, *the same β must describe the phase change along z for all three regions*. Hence, it is immediately found that β is constrained by the Pythagorean relationship to obey the following inequality:

$$\frac{\omega}{c} n_{1,3} < \beta < \frac{\omega}{c} n_2$$

or

$$n_{1,3} < \frac{\beta}{(\omega/c)} < n_2 \qquad \text{(for a guided and confined mode)} \tag{4.3.3}$$

The ratio β to $(k_0 = \omega/c)$ is often called the *effective index*, because it describes the propagation of the mode by a familiar relation; that is, $\beta = \omega n_{\text{eff}}/c$.

4.3.1 TE Mode ($E_z = 0$)

Consider the symmetric slab index fiber whose index of refraction variation with respect to x is sketched in Fig. 4.4. To keep the mathematics to a minimum, we assume a symmetrical

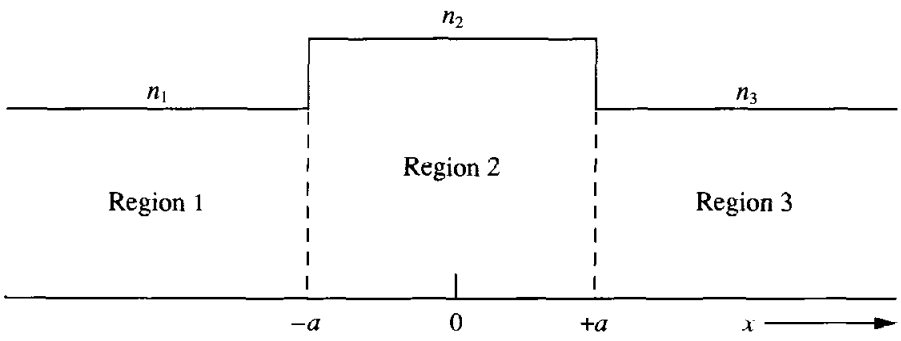


FIGURE 4.4. Index of refraction in a heterojunction laser or a slab fiber.

slab with $n_1 = n_3$ and $n_1 < n_2$. (Thus the sketch would correspond to a radial cut through a round fiber). We search for fields in the three regions that are "guided" by the dielectric step discontinuity at $x = \pm a$.

Let us first agree on what is meant by *guided*. It merely means that most of the field is in the central core, $-a < x < a$, and becomes very small for $|x - a|$ sufficiently large. Furthermore, all would agree that such solutions must obey the wave equation (4.3.2) and the appropriate boundary conditions. For the TE modes for the slab fiber, $\partial/\partial y = 0$ and (4.3.2b) becomes

$$\frac{\partial^2 H_z^{(1,2,3)}}{\partial x^2} + \left[\left(\frac{\omega n_{1,2,3}}{c}\right)^2 - \beta^2\right] H_z^{(1,2,3)} = 0 \tag{4.3.4}$$

There is a very important point to be noted about (4.3.4): There are three different fields corresponding to the three different regions, but there is only one phase constant β common to all three regions, as was mentioned previously. This is what is meant by a guided *mode*: a field configuration that retains its proportionality (i.e., shape) along x but which travels with a constant phase velocity independent of x.

If the mode amplitudes are to vanish at sufficient distances away from the central slab, we must hope for an exponential solution in regions 1 and 3 as was discussed previously:

$$H_z^{(1)} = B_1 \exp[h(x + a)] + A_1 \exp[-h(x + a)] \tag{4.3.5a}$$

$$H_z^{(3)} = A_3 \exp[h(x - a)] + B_3 \exp[-h(x - a)] \tag{4.3.5b}$$

with

$$h^2 = \beta^2 - \left(\frac{\omega n_{1,3}}{c}\right)^2 \tag{4.3.5c}$$

To keep the field finite at $x = -\infty$, $A_1 = 0$; and $A_3 = 0$ to eliminate a similar embarrassment at $x = +\infty$.

The analysis of Sec. 4.2 suggests using a trigonometric form of the solution of (4.3.4) for the central core:

$$H_z^{(2)} = A_2 \cos k_\perp x + B_2 \sin k_\perp x \tag{4.3.6a}$$

with

$$k_\perp^2 = (\omega n_2/c)^2 - \beta^2 \tag{4.3.6b}$$

Equation (4.3.6a) suggests that there is a natural classification of the modes into ones that are symmetric and ones that are antisymmetric about the plane $x = 0$. Our interest lies primarily with the transverse field components E_t and H_t, since they carry the optical power; and (4.3.1) indicates that their symmetry is opposite of H_z. Thus, for example, the symmetric transverse field components of the TE modes involve only the odd sine functions for H_z in (4.3.6) with the cosine functions describing antisymmetric ones. Let us focus on the symmetric case ($A_2 = 0$) and collect our equations together:

$$H_z^{(1)} = B_1 \exp[h(x + a)] \tag{4.3.7a}$$

$$H_z^{(3)} = B_3 \exp[-h(x-a)] \tag{4.3.7b}$$

$$H_z^{(2)} = B_2 \sin k_\perp x \tag{4.3.7c}$$

These fields must be continuous at $x = \pm a$; hence,

$$B_1 = -B_2 \sin k_\perp a \tag{4.3.8a}$$

$$B_3 = +B_2 \sin k_\perp a \tag{4.3.8b}$$

$$\therefore B_1 = -B_3 \tag{4.3.8c}$$

The transverse fields that are parallel to the slab are found by applying (4.3.1) to (4.3.7) and using (4.3.8a-c) for the amplitude of the fields in the three regions.

$$E_y^{(1)} = \frac{j\omega\mu_0}{h} B_3 \exp[h(x+a)] \tag{4.3.9a}$$

$$E_y^{(2)} = \frac{j\omega\mu_0}{k_\perp} B_3 \cos k_\perp x \tag{4.3.9b}$$

$$E_y^{(3)} = \frac{j\omega\mu_0}{h} B_3 \exp[-h(x-a)] \tag{4.3.9c}$$

where the symmetry equation (4.3.8c) was used. Again, these fields must be continuous at $x = \pm a$, and, because of symmetry, it is only necessary to work with one boundary condition.

$$\frac{j\omega\mu_0}{h} B_3 = \frac{j\omega\mu_0}{k_\perp} B_2 \cos k_\perp a \tag{4.3.10}$$

Combining (4.3.10) with (4.3.8b) yields a single implicit equation for the propagation constant β, which is contained in the quantities $k_\perp$ and h.

$$\frac{ha}{k_\perp a} = \tan k_\perp a \tag{4.3.11a}$$

If we follow the same procedure outlined above for the antisymmetric TE modes, we obtain

$$\frac{ha}{k_\perp a} = -\cot k_\perp a \tag{4.3.11b}$$

We will return to (4.3.11) after addressing similar issues for the TM modes.

4.3.2 TM Modes ($H_z = 0$)

The procedure for the TM case parallels the above except that now it is E_z, which is the parent field. Its variation in the various regions is precisely that given by (4.3.7a–c) (after changing H_z to E_z on the left-hand side), the symmetry arguments are the same, and the relationships between the coefficients expressed by (4.3.8a–c) also are the same. However,

the transverse fields analogous to (4.3.9a–c) are magnetic ones and are found by applying (4.3.1) to E_z in the various regions. For the symmetric TM case, we have

$$H_y^{(1)} = \frac{j\omega\epsilon_0 n_1^2}{h} B_1 \exp[h(x+a)] \tag{4.3.12a}$$

$$H_y^{(2)} = \frac{j\omega\epsilon_0 n_2^2}{k_\perp} B_2 \cos k_\perp x \tag{4.3.12b}$$

$$H_y^{(3)} = \frac{j\omega\epsilon_0 n_1^2}{h} B_3 \exp[-h(x-a)] \tag{4.3.12c}$$

For the symmetric mode, $B_1 = B_3$, as before, and, after matching the fields at the boundary, we obtain another implicit equation for the propagation constant, β, for the TM case:

$$\frac{ha}{k_\perp a} = \frac{n_1^2}{n_2^2} \tan k_\perp a \qquad \text{(Symmetric TM)} \tag{4.3.13a}$$

For the antisymmetric modes

$$\frac{ha}{k_\perp a} = -\frac{n_1^2}{n_2^2} \cot k_\perp a \qquad \text{(Antisymmetric TM)} \tag{4.3.13b}$$

4.3.3 Graphic Solution for the Propagation Constant: "R" and "V" Parameters

There is a very convenient and transparent graphic solution procedure for equations (4.3.11) and (4.3.13). Let us define dimensionless variables $X = k_\perp a$ and $Y = ha$ and recall the definitions of $k_\perp$ and h from (4.3.5c) and (4.3.6b). With those definitions, the quantity $R^2 = X^2 + Y^2$ is given by

$$R^2 = \left[(\beta a)^2 - \left(\frac{\omega n_1 a}{c}\right)^2\right] + \left[\left(\frac{\omega n_2 a}{c}\right)^2 - (\beta a)^2\right] = X^2 + Y^2$$

or
$$R = \left(\frac{\omega a}{c}\right)[n_2^2 - n_1^2]^{1/2} \tag{4.3.14}$$

Thus
$$R = (2\pi a/\lambda_0)NA = V$$

Thus the R number of the slab fiber is $2\pi/\lambda_0$ times the dimension a times the numerical aperture of the fiber. (In round fibers, a is the core radius, and the customary name for R is the "V" number.)

Thus (4.3.11) and (4.3.13) can be rewritten in terms of X and Y and coupled to R:

$$\begin{aligned} Y &= X \tan X && \text{Symmetric TE} \\ Y &= -X \cot X && \text{Antisymmetric TE} \\ Y &= (n_1/n_2)^2 X \tan X && \text{Symmetric TM} \\ Y &= -(n_1/n_2)^2 X \cot X && \text{Antisymmetric TM} \end{aligned} \tag{4.3.15}$$

with
$$R^2 = X^2 + Y^2$$

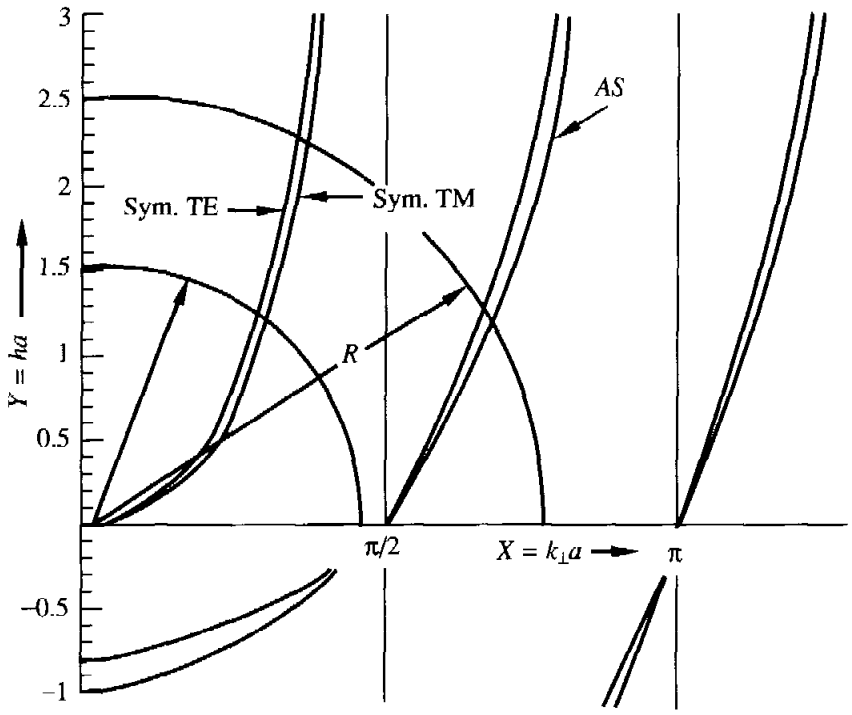


FIGURE 4.5. Graphic solution for the propagation constant in a slab fiber. The quantity (n_1/n_2) is assumed to be < 1 (as usual) for this graph.

The graphic solution is shown in Fig. 4.5 in the form suggested by the choice of variables X, Y, and R. Note that for $R < \pi/2$ only the lowest-order symmetric TE or TM modes can propagate.

Note too that ha represents the exponential decay rate of the fields away from the core of the fiber. A larger value of ha means that the field is more tightly confined to the central region. In this sense, then, the TE mode experiences greater confinement to the central region and thus can be considered the dominant mode. This plays a significant role in semiconductor lasers where the region $|x| > a$ is not excited and thus is lossy. The graphic solution used shows that the parameter ha is smaller for the TM mode compared to the TE for the same value of R and thus the field extends further into region 1 or 3 for the TM case. Such regions are lossy for heterojunction lasers, and, hence, they tend to oscillate in the TE orientation.

4.4 GAUSSIAN BEAMS IN GRADED INDEX (GRIN) FIBERS AND LENSES

A fiber that has a dielectric constant decreasing in parabolic fashion can be analyzed rather easily and precisely and, most importantly, can be made. The relative dielectric constant,

which is the square of the index of refraction is chosen* to be

$$\epsilon(r) \triangleq n_0^2[1 - (r/l_G)^2] \tag{4.4.1}$$

The fiber radial scale length l_G is a measure of the grading of the fiber. Usually l_G is many times the diameter of the fiber and thus there is no danger of the dielectric constant being negative. We can also consider (4.4.1) as the first two terms of a Taylor series expansion of the actual dielectric constant or a continuous approximation to the index of refraction of the step-index fiber in the manner shown as the shaded curve in Fig. 4.2

For such a medium, the wave equation becomes

$$\nabla^2 E + \frac{\omega^2}{c^2} n_0^2[1 - (r/l_G)^2]E = 0 = \nabla^2 E + k^2[1 - (r/l_G)^2]E \tag{4.4.2a}$$

where $$k = \frac{\omega}{c} n_0 \tag{4.4.2b}$$

As we have seen in the previous sections, a mode in a fiber may be composed of many plane waves, but the key issue is that the sum must be added to give a field that propagates along z while keeping its "shape" along the transverse coordinates. The only way this can be accomplished is for the field to factor into a product function $E(r, \phi, z) = E(x, y)e^{-j\beta z}$ where β is the phase constant to be determined. (Even though these fibers are round, we can use Cartesian coordinates, since the only boundary condition to be satisfied is that $|E|^2$ must vanish faster than $(1/r)^2$ for $r \to \infty$.) Substituting this form into (4.4.2a) yields

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \left(k^2 - \beta^2 - \frac{k^2 r^2}{l_G^2}\right) E = 0 \tag{4.4.3}$$

with $$r^2 = x^2 + y^2$$

Now we proceed with the standard method of solving second-order partial differential equations. Assume that $E(x, y)$ is a product function $X(x)Y(y)$. Substitute and differentiate, divide by the product, and rearrange the debris such that all functions of x are on one side of the equation and all functions of y are on the other.

$$\frac{Y''}{Y} + k^2 - \beta^2 - (k/l_G)^2 y^2 = -\frac{X''}{X} + (k/l_G)^2 x^2 = T \tag{4.4.4}$$

The only way the equality demanded by (4.4.4) can be satisfied for all values of x and y is for the right-hand and left-hand sides to be equal to a common constant. Call that separation constant T. Thus we have two ordinary separated differential equations. (The equations have the same appearance as the Schrödinger's equation for a harmonic oscillator.)

$$X'' + [T - (k/l_G)^2 x^2]X = 0 \tag{4.4.5}$$

$$Y'' + [k^2 - \beta^2 - T - (k/l_G)^2 y^2]Y = 0 \tag{4.4.6}$$

*Much of the literature specifies the index by $n(r) = n_0 - n_2 r^2$ with $n_2 r^2 \ll n_0$. To relate that notation to one used here, for the square, $n_0^2 - 2n_2 r^2 + n_2^2 r^4$, neglect the last term and set the remainder equal to (4.4.1) to find $(n_0/l_G)^2 = 2n_2$.

After a frantic search of mathematical tables, we find that the Hermite-Gaussian functions of the form $H_m(u)\exp[-u^2/2]$ solve these differential equations provided the variable u and the constant terms T and $k^2 - \beta^2 - T$, are chosen correctly. If we substitute u for $(k/l_G)^{1/2}x$, (4.4.5) becomes

$$\frac{d^2X}{du^2} + \left(\frac{Tl_G}{k} - u^2\right)X = 0 \tag{4.4.7a}$$

where

$$\left(\frac{k}{l_G}\right)^{1/2} x \triangleq u \tag{4.4.7b}$$

The Hermite polynomial of degree m times a Gaussian is a solution provided that the constant in (4.4.7a) is an odd integer, $2m + 1$,

$$X(x) = H_m[(k/l_G)^{1/2}x]\exp\left[-\frac{kx^2}{2l_G}\right] \tag{4.4.8}$$

if

$$\frac{Tl_G}{k} = 2m + 1 \Rightarrow T = (2m + 1)\frac{k}{l_G} \tag{4.4.9}$$

We repeat the same manipulations on (4.4.6) to find that

$$Y(y) = H_p[(k/l_G)^{1/2}y]\exp\left[-\frac{ky^2}{2l_G}\right] \tag{4.4.10}$$

provided

$$\frac{l_G}{k}(k^2 - \beta^2 - T) = 2p + 1$$

or

$$\beta^2 = k^2 - T - (2p + 1)\frac{k}{l_G} \tag{4.4.11}$$

where $2p + 1$ is a different odd integer.

Now (4.4.9) and (4.4.11) are combined to solve for the "unknown" propagation constant β:

$$\beta^2 = k^2 - \frac{2k}{l_G}(1 + m + p) \tag{4.4.12}$$

or

$$\beta = k\left[1 - \frac{2}{kl_G}(1 + m + p)\right]^{1/2} \tag{4.4.13}$$

If we define a characteristic spot size w by

$$w^2 = 2l_G/k = (\lambda_0 l_G/\pi n_0) \tag{4.4.14}$$

then we obtain a familiar-appearing expression,

$$E(x, y, z) = E_0 H_m\left(\frac{\sqrt{2}x}{w}\right) H_p\left(\frac{\sqrt{2}y}{w}\right) \exp\left[-\frac{r^2}{w^2}\right] \times \exp\left\{-jk\left[1 - \frac{2}{kl_G}(1 + m + p)\right]^{1/2} z\right\} \tag{4.4.15}$$

The only serious approximation made is ignoring the boundary condition at $r = a$, the radius of the fiber. However, a simple example will show that the field is negligible there and thus can be ignored.

Example

Consider a fiber with the following specifications:

1. Diameter of fiber: 50 μm = 0.05 mm; $\therefore a = 25$ μm
2. Index of refraction at center of core: $n_0 = 1.52$
3. Index of refraction at $r = 25$ μm; $n = 1.52 - 0.008 = 1.512$
4. Wavelength region of interest: $\lambda_0 = 1.06$ μm

The index of refraction is given by the square root of $\epsilon(r)$ or

$$n(r) = \epsilon^{1/2}(r) = n_0\left(1 - \frac{1}{2}\left(\frac{r}{l_G}\right)^2\right)$$

From specification 3,

$$\Delta n = \frac{1}{2} n_0 \left(\frac{r}{l_G}\right)^2 n$$

or

$$l_G = \left(\frac{n_0 a^2}{2\Delta n}\right)^{1/2} = 0.0244 \text{ cm} = 244 \ \mu\text{m}$$

Now $k = 2\pi n_0/\lambda_0 = 9.04 \times 10^{+4}$ cm^{-1} $= 9.04$ μm^{-1}
Hence $w = (2l_G/k)^{1/2} = 7.36$ μm.
Thus the exponential term at the boundary of the fiber is

$$\exp\left[-\left(\frac{a}{w}\right)^2\right] = \exp\left[-\left(\frac{25 \ \mu\text{m}}{7.36 \ \mu\text{m}}\right)^2\right] = 9.75 \times 10^{-6} \approx 10^{-5}$$

While the Hermite polynomials peak at $|x, y| > 0$, it takes a very large mode number to make up for the factor of 10^{-5}.

Notice that there was only $\sim$ 0.5% change in index going from the central axis to the outer radius, but that the mode was still tightly confined. Indeed, the index changes by only 0.00045 over the distance $r = w$. Thus it does not take much change in $n(r)$ to promote guidance nor, unfortunately, to promote anti-guiding.

Note also that the phase constant depends upon the mode number and the characteristic length l_G. Thus, if the information is not confined to a single electromagnetic mode, the various parts will travel with different velocities. This is called *modal dispersion* and the general topic of dispersion is the most serious limitation in fiber optic communications.* While the explicit formula for β just derived is pertinent only to the parabolic graded index case considered here, it is representative of many fibers, and thus the dispersion problem will permeate much of the system considerations.

There is another approximate approach to obtaining the field configuration and propagation constant for a graded index fiber or lens that avoids the high power mathematics just used. We can postulate that the Hermite-Gaussian beam modes are the normal modes of the fiber, use the definition of a mode as being a field configuration which maintains its "shape" while propagating, and then apply the $ABCD$ law of Sec. 3.6 to a length z of the fiber. We verify our postulate and logic by demonstrating self-consistency.

If Gaussian beams are the normal modes of the waveguide, it must be described by a complex beam parameter q. We do not know the value of q, only that it exists (by the postulate). However, we also know that q will transform according to the $ABCD$ law, from (3.6.1)

$$q(z) = \frac{A(z)q(0) + B(z)}{C(z)q(0) + D(z)} \tag{4.4.16}$$

where $\mathbf{T}$ is given by

$$\mathbf{T} = \begin{bmatrix} \cos\left(\frac{z}{l_G}\right) & l_G \sin\left(\frac{z}{l_G}\right) \\ -\frac{1}{l_G}\sin\left(\frac{z}{l_G}\right) & \cos\left(\frac{z}{l_G}\right) \end{bmatrix}$$

Now, of course, the fiber does not know of our arbitrary choice of the axis and hence does not know the location of $z = 0$. Therefore the only way for the postulated beams to be the normal modes is for the unknown complex beam parameter q to be independent of z (and our choice of the origin).

$$q(z) = q_0 \qquad \text{(i.e., } q \text{ is independent of } z\text{)} \tag{4.4.17}$$

Combining (4.4.16) and (4.4.17) leads to a single equation for $1/q_0$:

$$\frac{1}{q_0} = -\frac{A - D}{2B} - j\left[1 - \left(\frac{A + D}{2}\right)^2\right]^{1/2} \div B$$

or

$$\frac{1}{q_0} = 0 - j\frac{[1 - \cos^2(z/l_G)]^{1/2}}{l_G \sin(z/l_G} = -\frac{j}{l_G} \tag{4.4.18}$$

*The fact that the index of refraction is wavelength dependent is the most serious cause of dispersion, and some of the consequences are discussed in Sec. 4.6 to 4.9.

If we substitute this answer for $1/q_0$ into (3.3.2), we have the variation of the field with a minimum of fuss.

$$E \propto \psi = \exp\left(-j\frac{kr^2}{2q_0}\right) = \exp\left(-\frac{kr^2}{2l_G}\right) \qquad (3.3.2) \rightarrow (4.4.19)$$

We can define a characteristic spot size for this newly found Gaussian beam by

$$w^2 = \frac{2l_G}{k} \qquad (4.4.20)$$

which is identical to the result from (4.4.14) and thus the field of the $\text{TEM}_{0,0}$ mode is the same.

The fact that the spot size in this Gaussian beam does *not* change with z is due to the focusing properties of the medium. As anticipated in Chapter 2, the natural divergence of the beam is exactly balanced by the convergence of the medium and thus leads to a constant beam size in the fiber.

The fact that q is independent of z leads to a very simple solution for the parameter $P(z)$ of (3.3.2). If we repeat the derivation of (3.3.4a) and (3.3.4b) for a quadratic index variation given by (2.12.7), we obtain

$$P'(z) = -j\frac{1}{q(z)} \qquad (3.3.2) \rightarrow (4.4.21)$$

But since $1/q(z) = -j(1/l_G)$ by (4.4.18) and is independent of z, the solution is trivial.

$$P(z) = -\frac{z}{l_G} \qquad (4.4.22)$$

Hence, the electric field has the following approximate form:

$$E(r, z) = E_0 \exp\left(-\frac{kr^2}{2l_G}\right) \exp\left[-j\left(k - \frac{1}{l_G}\right)z\right] \qquad (4.4.23)$$

Thus we have shown that the postulate of a Gaussian beam within this fiber is self-consistent. Note, however, that this beam is somewhat different than those encountered in Chapter 3. For instance, freely propagating beams expand with z, whereas here the spot size remains independent of z. In free space, the longitudinal-phase factor [see Eq(3.3.14)] depends on z, whereas the corresponding term in (4.4.23) is independent of z. Those differences are because the beam is being continuously focused to exactly balance the natural tendency to spread.

There is a small difference in the phase constant β found from the two approaches: compare (4.4.15) to (4.4.23) for the $\text{TEM}_{0,0}$ mode.

$$\beta = k\left[1 - \frac{2}{kl_G}\right]^{1/2} \quad \Leftrightarrow \quad \beta = k\left[1 - \frac{1}{kl_G}\right]$$

"Exact" (4.4.15) $\qquad$ $ABCD$ law (4.4.23)

The result from the $ABCD$ law is the first two terms of a Taylor series expansion of the exact relationship. However, the quantity $1/kl_G = 1/(2198) = 4.5 \times 10^{-4}$ and thus two

terms are quite adequate. The fact that we obtain such a precise answer with such a simple theory as the $ABCD$ law should inspire confidence in other calculations using it.

4.5 PERTURBATION THEORY

Quite often in optics, we are faced with an intractable theoretical problem involving propagation through an inhomogeneous medium, encounter boundaries that do not fit common coordinate systems, or run into media that are slightly nonlinear. While we may have a problem, the electromagnetic wave solves the differential equations (instantly) and propagates along its merry way down the guide. Fortunately, there is a way to obtain a very good (but approximate) answer by using the theory presented here.

We wish to describe the propagation of a confined and guided mode, a field configuration whose relative shape in the cross-sectional plane is independent of z and which vanishes at a sufficiently large distance from the axis. Such a field can be expressed as $\psi(x, y)e^{-j\beta z}$, where ψ is the function describing the amplitude of a cartesian component of the field and $e^{-j\beta z}$ expresses the phase retardation along z. The wave equation becomes

$$\nabla_t^2\psi + \left[\frac{\omega^2}{c^2}\epsilon_r(x, y) - \beta^2\right]\psi = 0$$

or

$$\nabla_t^2\psi + [k^2(x, y) - \beta^2]\psi = 0 \tag{4.5.1}$$

where $k^2(x, y) = (\omega/c)^2 n^2(x, y) = (\omega/c)^2\epsilon_r(x, y)$; $n(x, y)$ = local index; $\epsilon_r(x, y)$ = relative dielectric constant; and the subscript "t" indicates that the vector operation is in the transverse plane. Even though the index/dielectric constant may be a function of the transverse coordinates, the phase constant is not. The mode moves along z as a complete package according to $e^{-j\beta z}$. There are special cases of the variation of $n^2(x, y)$ for which a rigorous analytic solution exists, one of which was given in Sec. 4.4. While that particular case is of some practical importance, most of the real variations of $n^2(x, y)$ are virtually intractable unless a tour-de-force of mathematics or numerical computation is employed. An easier approach is to address the problem as a series of "guesses" and rely on the fact that the unknown propagation constant β is a minimum when the correct function ψ is chosen.

We start by manipulating (4.5.1) with a few exact operations before introducing the approximation route. Multiply it by the (unknown) function ψ, integrate over the cross section of the guide, and solve for β^2.

$$\iint \psi\nabla_t^2\psi\, dA + k_0^2 \iint \epsilon_r(x, y)\psi^2\, dA - \beta^2 \iint \psi^2\, dA = 0$$

or

$$\beta^2 = \frac{k_0^2 \iint \epsilon_r(x, y)\psi^2\, dA + \iint \psi\nabla_t^2\psi\, dA}{\iint \psi^2\, dA} \tag{4.5.2}$$

where $k_0^2 = \omega/c$.

Equation (4.5.2) can be written in many forms depending upon our preferences and ultimate purpose. If we use the vector identity

$$\nabla_t \cdot (u\nabla_t v) \equiv (\nabla_t u \cdot \nabla_t v) + u\nabla_t^2 v$$

or

$$\psi\nabla_t^2\psi = -(\nabla_t\psi)^2 + \nabla_t \cdot [\psi\nabla_t\psi]$$

the second integral in the numerator of (4.5.2) can be converted to

$$\iint \psi\nabla_t^2\psi \, dA = -\iint (\nabla_t\psi)^2 \, dA + \oint \psi\nabla_t\psi \cdot \mathbf{n} \, dl \tag{4.5.3}$$

where the divergence theorem is used to convert the second term to a line integral along the perimeter of the guide. Along this path the wavefunction $\psi = 0$ and our equation becomes

$$\beta^2 = \frac{k_0^2 \iint \epsilon_r(x, y)\psi^2 \, dA - \iint (\nabla_t\psi)^2 \, dA}{\iint \psi^2 \, dA} \tag{4.5.4}$$

This is the central equation of this section, but it appears to be quite useless: The desired unknown quantity β is related to the integral of still another unknown function ψ. Have we made progress?

The answer is yes because (4.5.4) is a (variationally) correct expression for the propagation constant, which implies some very important and comforting facts:

1. It is correct in the sense that if we were "lucky" and guessed the correct wavefunction ψ_c, then a precise answer would result.
2. If we make a first-order error in the choice (or guess) of ψ, then the first-order variation* in β is zero. Thus, the last fact occurs.
3. The value of β^2 computed by (4.5.4) will be a minimum when ψ is chosen correctly, and, thus by choosing a sequence of functions involving adjustable constants, we can improve the approximation to the correct eigenfunction by minimizing β^2.

These points can be best illustrated with an example.

Example: An optical fiber with a parabolic index

Almost all fibers have an index of refraction that decreases with radius in the manner shown in Fig. 4.2 so as to confine the field near the axis. The case of the graded index fiber is one that can be and has been solved in Sec. 4.4. However, let us ignore that prior work and apply (4.5.4) together with the three characteristics just noted to guide our work.

Let us guess that $\psi(r)$ will have a Gaussian character with an unknown spot size w.

First guess:	$\psi = e^{-(r/w)^2}$	w is unknown
with	$\epsilon_r(x, y) = n_0^2[1 - (r/l_G)^2]$	(l_G is known)

*A proof follows the procedure outlined by R.E. Collin, *Field Theory of Guided Waves*, New York: McGraw-Hill, 1960), pages 128–129. An integral is added to the numerators of (4.5.2) and (4.5.4), whose value is zero when the correct wavefunction is used and which makes the first-order variation in β vanish independent of the variation in ψ.

Now the task amounts to performing some elementary integrations over the circular cross section

$$\psi = e^{-r^2/w^2} \qquad \psi^2 = e^{-2r^2/w^2} = e^{-u}$$

where $u = 2r^2/w^2 \qquad du = 4r\,dr/w^2 \qquad dA = 2\pi r\,dr = \pi w^2/2\,du$

$$\nabla\psi = \frac{-2r}{w^2} e^{-r^2/w^2} \qquad (\nabla\psi)^2 = \frac{4r^2}{w^4} e^{-2r^2/w^2} = \frac{2}{w^2} ue^{-u}$$

$$(r/l_G)^2\psi^2 = (w/2l_G)^2 2(r/w)^2 e^{-2r^2/w^2} = [w^2/2l_G^2]ue^{-u}$$

Equation (4.5.4) becomes (with $k = \omega n_0/c$)

$$\beta^2 = k^2 - \frac{\dfrac{\pi w^2}{2}\left[k^2\dfrac{w^2}{2l_G^2}\displaystyle\int_0^\infty ue^{-u}\,du - \frac{2}{w^2}\int_0^\infty ue^{-u}\,du\right]}{\dfrac{\pi w^2}{2}\displaystyle\int_0^\infty e^{-u}\,du}$$

or $$\beta^2 = k^2 - \frac{k^2w^2}{2l_G^2} - \frac{2}{w^2} \tag{4.5.5}$$

Now, choose the value of w^2 which minimizes (4.5.5).

$$w^2 = \frac{2l_G}{k} \qquad \text{[minimizes (4.5.5)]} \tag{4.5.6}$$

If we refer back to (4.4.14), we discover that this is the precise answer for w^2 and also for the propagation constant when (4.5.6) is substituted back into (4.5.5).

$$\beta^2 = k^2 - \frac{k}{l_G} - \frac{k}{l_G} = k^2 - \frac{2k}{l_G} = k^2\left(1 - \frac{2}{kl_G}\right) \tag{4.5.7}$$

which is precisely the same answer found from the differential equation approach of Sec. 4.4 [compare (4.5.7) to (4.4.12) for the $m = p = 0$ case]. It is even more accurate than that found from the $ABCD$ law. Of course, these exact answers are a result of a lucky guess for the functional form but the Rayleigh-Ritz procedure* can be used to obtain a convergent sequence of better approximations when one is not so lucky.

A major use of (4.5.4) is when we have an exact solution for a tractable problem and the practical one can be considered as a perturbation δn. Thus

$$\beta^2 = \frac{k^2\iint[(n+\delta n)/n_0]^2\psi^2\,dA - \iint(\nabla_t\psi)^2\,dA}{\iint\psi^2\,dA} \tag{4.5.8}$$

Let β_0 be the solution when $\delta n = 0$ with the known wave function ψ_0. Then, β becomes $\beta_0 + \delta\beta$ in response to the change δn in the index and $\beta^2 = \beta_0^2 + 2\beta_0(\delta\beta)$.

$$\beta_0^2 = \frac{k^2\iint[n/n_0]^2\psi_0^2\,dA - \iint(\nabla_t\psi_0)^2\,dA}{\iint\psi_0^2\,dA} \tag{4.5.9}$$

*This consists of expanding ψ in a series of orthogonal functions $\psi_0 = \Sigma A_n\phi_n$ where $\{\phi_n\}_{r\to\infty} = 0$ and minimizing the resultant mess by setting $\partial\beta/\partial(A_n) = 0$.

Subtract 4.5.9 from 4.5.8, recognize that $\iint \psi^2\, dA \approx \iint \psi_0^2\, dA$, and that $\beta^2 \simeq \beta_0^2 + 2\beta_0\delta\beta$

$$2\beta_0\delta\beta = \frac{k^2 \iint \left\{[(n+\delta n)/n_0]^2\psi^2 - [n/n_0]^2\psi_0^2\right\} dA - \iint \left\{(\nabla_t\psi)^2 - (\nabla_t\psi_0)^2\right\} dA}{\iint \psi_0^2\, dA}$$

For these small changes in the index, $(n+\delta n)^2 \approx n^2 + 2(\delta n)n$ and the wavefunction can be approximated by the old one, $\psi \approx \psi_0$, making the last integral in the numerator zero.

$$\boxed{\delta\beta = \frac{k^2}{2\beta_0} \frac{\iint [2n\delta n/n^2]\psi_0^2\, dA}{\iint \psi_0^2\, dA} = \frac{\omega^2}{c^2}\frac{n}{\beta_0} \frac{\iint \delta n \psi_0^2\, dA}{\iint \psi_0^2\, dA}} \tag{4.5.10}$$

where δn is the perturbation in the index. This is the main reason for examining contrived problems in great detail (and not, as some students suspect, for their harassment). Quite often, real practical problems can be considered as perturbations on the contrived or academic ones inflicted on the students. A very important application of this formula will be found in Sec. 4.8, where a slight intensity nonlinearity in the index compensates for the material dispersion in a fiber to permit soliton propagation.

4.6 DISPERSION AND LOSS IN FIBERS: DATA

Fiber-optic waveguides have such little loss that they have started to replace traditional metal-based guides such as twisted pairs, coaxial cables, and microwave radio relay links as these latter technologies become overloaded, wear out, or are no longer cost effective. An example of fiber performance is shown in Fig. 4.6, where the transmission (in dB/km) is plotted as a function of wavelength for the particular fiber. Notice that there is only 0.2 dB per kilometer loss for this particular fiber (or a transmission of 95.5% per 0.6 mile) at $\lambda_0 = 1.55\ \mu$m. This is far better than can be achieved in free space propagation given the uncertainties in weather. Indeed, this minimum in loss accounts for the frantic pace of research and development on semiconductor laser sources and optical amplifiers for that wavelength.

However, a minimum loss is not the only criteria for a communication system using a binary pulse code for information transfer. While fidelity is important, perfect transmission at an infinitesimal rate is worthless. Error-free transmission at high speed is the goal. A major limitation to the data rate is the material dispersion causing some wavelengths to travel at a different velocity than others.

Consider Fig. 4.7, which illustrates the effect very clearly, but uses wavelengths not normally employed for fiber optic communications. Two separate semiconductor lasers were used to generate the two pulses shown in Fig. 4.7(a), at the two wavelengths of 824 nm and 803 nm. The pulses are separated by 3.2 ns at the input to the fiber, and, of course, it takes a finite time delay to get to the end of the fiber. If this were a perfect transmission system, then there would still be 3.2 ns between the pulses, but this separation has increased to 5.6 ns.

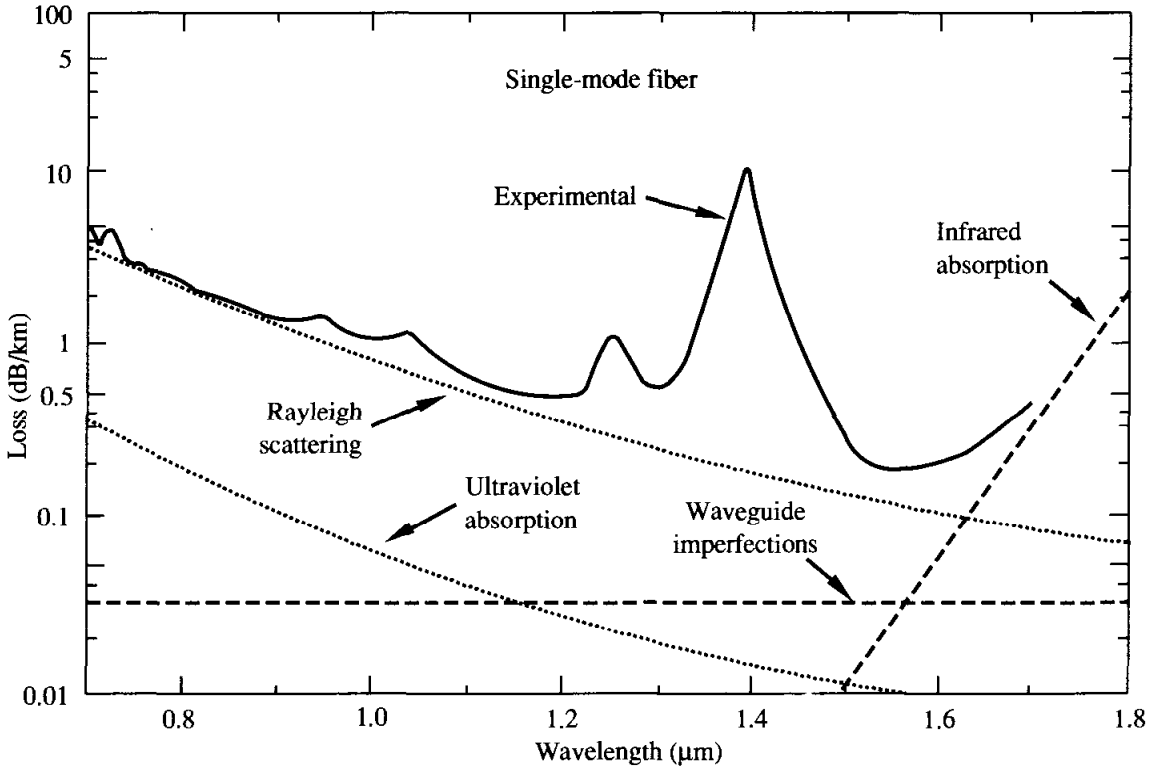


FIGURE 4.6. The transmission loss for a single mode GeO doped SiO_2 fiber. The diameter of the core is $2a = 9.4\mu$m with an index step between the core and the cladding of 0.0028. Data from T. Meya et al. (Ref. 21).

The two wavelengths travel at different velocities. While this example is exaggerated by the use of two widely separated wavelengths, the deleterious effects are also seen in the pulse shapes; that is, the durations of the output pulses are longer than the input. Because of the "bit" or pulsed nature of the information, there is a small spectral spread in λ around 824 and 803 nm with each spectral component traveling at a slightly different velocity and leads to the broadening of the pulse shape.

Electromagnetic energy travels at the group velocity given by

$$v_g = \left(\frac{\partial \beta}{\partial \omega}\right)^{-1} \tag{4.6.1}$$

If the phase constant were simply related to the index of refraction by $\beta = \omega n/c$ and if the index did not depend upon frequency or wavelength, then $v_g = c/n$, which is also the phase velocity. Neither of the "ifs" are obeyed (except for a vacuum) and thus v_g is frequency (or wavelength) dependent.

The phase constant β is dependent upon the dimensions, construction of the fiber, and the index by a fairly complicated expression. For the graded index case of Sec. 4.4, the phase constant of the $\text{TEM}_{m,p}$ mode was given by

$$\beta = k\left[1 - \frac{2(1+m+p)}{kl_G}\right]^{1/2} \triangleq \frac{\omega}{c} n_{\text{eff}}(\omega) \tag{4.6.2}$$

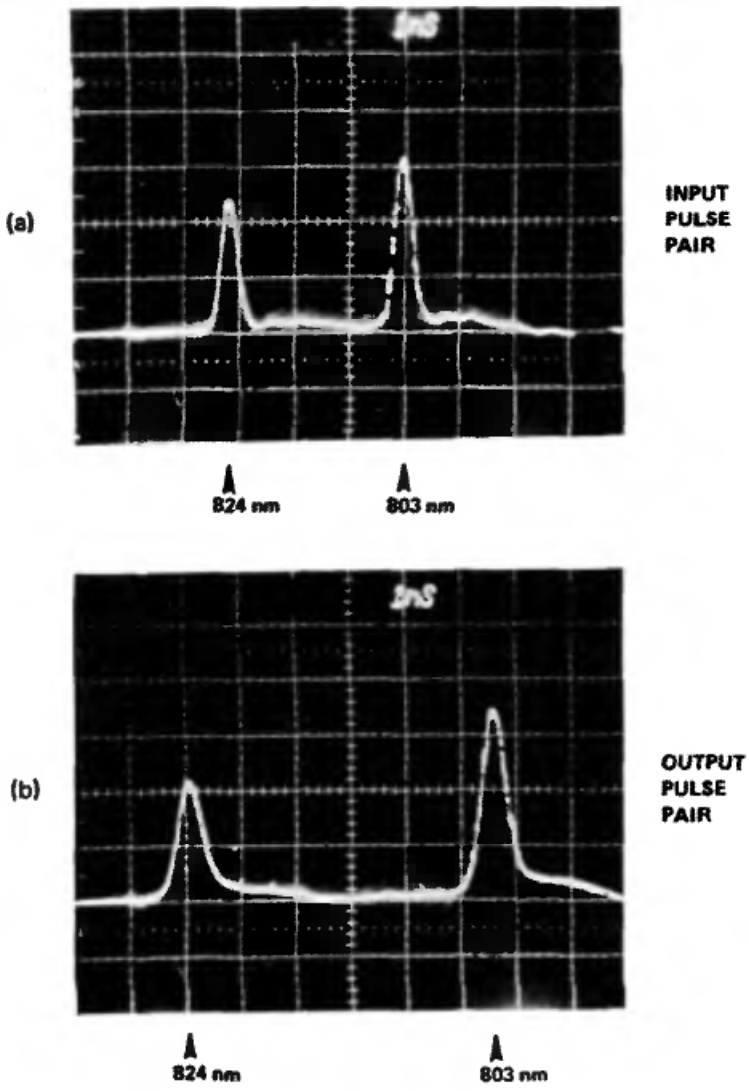


FIGURE 4.7. Dispersion effects in an optical fiber: (a) input pulse pair; (b) output pulse pair. (From Ref. 11.)

where
$$k = \frac{\omega}{c} n(\omega) \tag{4.6.3}$$

and the *effective index* is defined to give the correct value for the phase constant. Thus the group velocity depends upon the fact that a confined mode is propagating, and, thus, v_g depends upon the mode numbers m and p and most importantly depends on the inherent dependence of material index $n(\omega)$ on the frequency. The former is called waveguide or modal dispersion with the more important latter effect being called material dispersion.

We expand β in a Taylor series around the nominal central frequency ω_0 (or λ_0) by

$$\beta(\omega) = \beta(\omega_0) + \left.\frac{\partial \beta}{\partial \omega}\right|_{\omega_0} (\omega - \omega_0) + \frac{1}{2} \left.\frac{\partial^2 \beta}{\partial \omega^2}\right|_{\omega_0} (\omega - \omega_0)^2 + \cdots \tag{4.6.4a}$$

$$= \frac{\omega}{v_p} + \frac{\omega - \omega_0}{v_g(\omega_0)} + \frac{1}{2} \left(-\frac{1}{v_g^2(\omega_0)} \left.\frac{\partial v_g}{\partial \omega}\right|_{\omega_0} \right) (\omega - \omega_0)^2 \tag{4.6.4b}$$

The presence of the second derivative of β causes different spectral components to travel with different velocities and thus have different group delays. This is usually presented in the form of dispersion in arrival times in practical units of picosecond per nm of bandwidth per km of length. Figure 4.8 presents data on dispersion for a single mode fiber 11 km long and shows that the minimum occurs at $\lambda_0 \sim 1.32\mu$m.

To put this data into proper perspective, consider the limitations on using a LED source as the transmitter in an optical communication link of 11 km. Such an incoherent LED source might have a large spectral width (~5 nm) and thus experience a large dispersion in the arrival times of the various parts of a pulse, with some of it appearing in the time slot reserved for the next bit if the clock rate were too fast or the dispersion were too large. If the central wavelength were 0.8 μm, then the net dispersion would be 5 nm $\cdot$ 10^3 ps/nm = 5 ns. Thus, there is a high probability of error for any link with a data rate faster than 100 Mbit/s. If the source were a pulse-code-modulated laser at 1.32 μm occupying a spectral bandwidth on the order of that required to satisfy the Fourier transform requirements (usually 0.1 nm is more than adequate allowance), then a much faster data rate of ~50 Gbit/s can be envisioned were it not for the fact that handling information at that rate is difficult. Of course, 11 km is not that long a distance, but the effect scales directly with length, and hence we could envision a 5 Gbit/s data rate over 100 km by using a laser at 1.32 μm while suffering ~50 dB transmission loss (from Fig. 4.7). If the

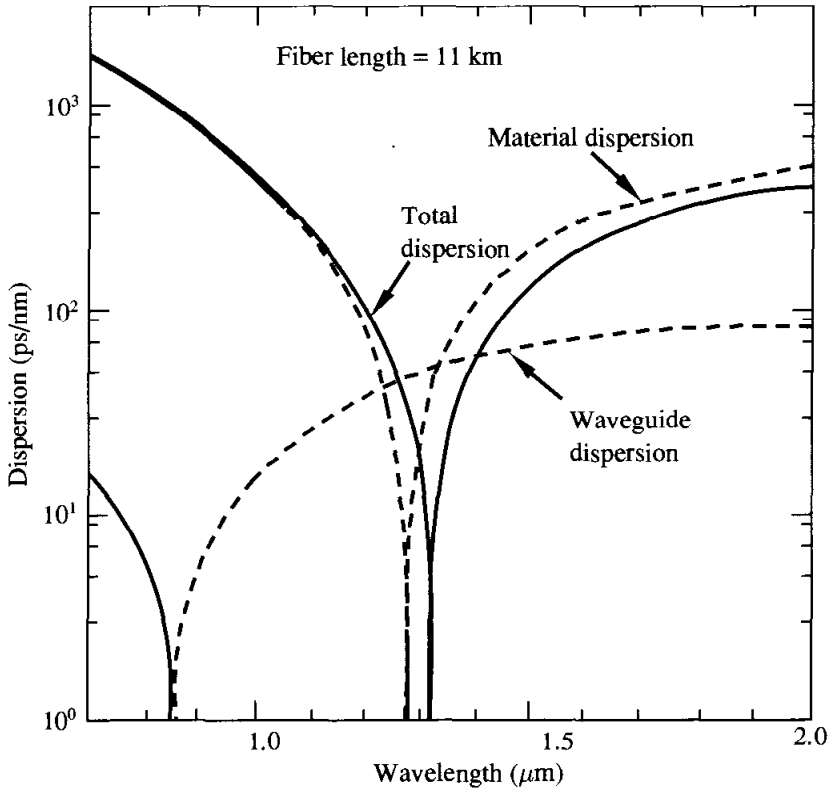


FIGURE 4.8. The various contributions to the dispersion in a single mode fiber. The dominant effect is due to the dependence of the index of the material on wavelength, but there is some due to the modal properties of the waveguide. (Data from [19])

power launched into the fiber were 20 mW(+13 dBm), then the received power would be −37 dBm (0.2 μW). In a sample time slot of 10^{-10} sec, 2×10^{-17} Joules or 133 photons would be received, adequate for a good detector.

Even though a communication system at that wavelength must live with a higher loss (0.5 dB/km), the dispersion is also much less and hence the data *rate* can be larger than at 1.55 μm if nothing else were pertinent. There are two issues that suggest that 1.55 μm will remain a dominant fiber optic wavelength: (1) There are convenient optical amplifiers at this wavelength (to be discussed in Chapter 10) so that even the small loss can be canceled, and (2) the fiber will support a "soliton" pulse (see Sec 4.8) in which the slight fiber nonlinearity to the optical pulses compensates for the dispersion permitting essentially distortion-free pulse transmission. This is and will continue to be a significant research and development topic.

4.7 PULSE PROPAGATION IN DISPERSIVE MEDIA: THEORY

As the bit rate for optical communications has escalated, it has become imperative to have analytic tools available to predict the propagation, the dispersion, and any characteristic pulse shapes. That is the goal of this section.

We start by reminding ourselves of the connection between the time and frequency domain provided by the Fourier transform pairs.

$$f(z, t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(z, \omega)\, e^{j\omega t}\, d\omega \tag{4.7.1}$$

where

$$F(z, \omega) = \int_{-\infty}^{+\infty} f(z, t) e^{-j\omega t}\, dt \tag{4.7.2}$$

We will presume that the electromagnetic energy is in the form of a guided mode in a fiber whose electric field can be expressed as a product of the transverse variations $\psi(x, y)$ and the function of z and t as just indicated. We can be a bit more specific since most of the optical energy will be centered in a narrow band around ω_0, and thus we use an envelope representation for $E(z, \omega)$

$$F(z, \omega) = E(z, \omega - \omega_0) e^{-j\beta(\omega_0)z} \tag{4.7.3a}$$

where $E(z, \omega - \omega_0)$ is the Fourier transform of the *envelope* of the electric field, $E(z, t)$. Insert this representation into (4.7.1) and realize that spectral representation of the envelope appears in a narrow band of frequencies around ω_0, which can be extended to $\pm\infty$ with little approximation.

$$f(z, t) = \left\{ \frac{1}{2\pi} \int_{-\infty}^{+\infty} E(z, \omega - \omega_0) e^{j(\omega-\omega_0)t}\, d(\omega - \omega_0) \right\} e^{j[\omega_0 t - \beta(\omega_0)z]} \tag{4.7.3b}$$

where $d\omega_0 = 0$ has been added to $d\omega$ and a multiplicative factor of $1 = [\exp(-j\omega_0 t)] \cdot [\exp(+j\omega_0 t)]$ has been used. Thus the brace is the Fourier transform of the envelope with

respect to z and t and the complete field is represented by

$$f(z, t) = e(z, t) = E(z, t)e^{j(\omega_0 t - \beta(\omega_0)z)} \tag{4.7.4}$$

This relation should appeal to our background in elementary wave propagation. If the envelope $E(z, t)$ were simply E_0, then (4.7.4) states that the constant amplitude wave (E_0) at a single frequency (ω_0) propagates as $\exp[-j\beta(\omega_0)z]$, where $\beta(\omega_0)$ is the phase constant evaluated at the central frequency ω_0. If the envelope is a short pulse, then there will be spectral components at $\omega \neq \omega_0$, which propagate with a (possibly) different phase constant $\beta(\omega)$. The phase constant differs from $k(\omega)$ and reflects the fact that the mode may be guided so that $\beta(\omega)$ must be determined from a solution to the wave equation. The function $F(z, \omega)$ obeys the conventional transmission line relation

$$\frac{\partial F(z, \omega)}{\partial z} = -j\beta(\omega)F(z, \omega) \tag{4.7.5}$$

which merely states that each spectral component of the pulse at ω propagates with a phase constant $\beta(\omega)$. All of these should be considered a reminder of the basics used in all wave propagation problems and there should be no surprises nor difficulties in applying them to any of the usual problems encountered.

When the pulse width τ_p is so short that the spectral bandwidth $\delta\omega \sim 1/\tau_p$ is a significant fraction of ω_0, then we must account for the fact that each spectral component of $E(z, \omega)$ has its own phase constant $\beta(\omega)$. To obtain an equation for the scalar envelope function $E(z, \omega - \omega_0)$, we use (4.7.3a) for the left-hand side of (4.7.5) and expand the phase constant on the right in a Taylor series around the central frequency ω_0.

$$\begin{aligned}
\frac{\partial E(z, \omega)}{\partial z} &= \frac{\partial}{\partial z}[E(z, \omega - \omega_0)e^{-j\beta(\omega_0)z}] \\
&= \left[-j\beta(\omega_0)E(z, \omega - \omega_0) + \frac{\partial E(z, \omega - \omega_0)}{\partial z}\right] e^{-j\beta(\omega_0)z} \\
&= -j\left\{\beta(\omega_0) + \delta\beta + \frac{\partial\beta}{\partial\omega}(\omega - \omega_0) + \frac{1}{2}\frac{\partial^2\beta}{\partial\omega^2}(\omega - \omega_0)^2\right\} E(z, \omega - \omega_0)e^{-j\beta(\omega_0)z}
\end{aligned}$$

where the square brackets on the second line is the expansion of $\partial E/\partial z$, the braces on the third line is the Taylor series expansion of β around the center frequency ω_0, and all derivatives of β are to be evaluated at $\omega = \omega_0$. We have also allowed for the possibility that there might be a perturbation induced by the amplitude of the field itself; that perturbation is denoted by $\delta\beta$ and will be discussed in detail in the next section. After canceling the common factor of $\exp[-j\beta(\omega_0)z]$ and deleting $-j\beta(\omega_0)E$ from both sides, we obtain

$$\frac{\partial E}{\partial z} = -j\delta\beta E - j\beta_1(\omega - \omega_0)E + j\frac{\beta_2}{2}(\omega - \omega_0)^2 E \tag{4.7.6}$$

where the subscripts on β represent the order of the derivative with respect to ω. While the angular frequency ω is the most convenient theoretical variable, the wavelength $\lambda = 2\pi c/\omega$

is the practical mode of specification. Hence there are various terminologies used to describe the same phenomena. The following illustrates the multiplicity of names*

$$\beta \triangleq \frac{\omega}{c} n(\omega) \triangleq \frac{\omega}{v_p(\omega)} \tag{4.7.7a}$$

where $n(\omega)$ = index; $v_p(\omega) = c/n(\omega)$ = phase velocity

$$\beta_1 \triangleq \frac{\partial \beta}{\partial \omega} = \frac{1}{c}\left[n + \omega \frac{dn}{d\omega}\right] = \frac{1}{c}\left[n(\lambda) - \lambda \frac{dn}{d\lambda}\right] = \frac{n_g}{c} = \frac{1}{v_g} \tag{4.7.7b}$$

where v_g = group velocity and n_g = group index.

$$\beta_2 \triangleq \frac{\partial^2 \beta}{\partial \omega^2} = \frac{1}{c}\left[2\frac{\partial n}{\partial \omega} + \omega \frac{d^2 n}{d\omega^2}\right] = -\frac{1}{v_g^2(\omega)} \frac{dv_g}{d\omega} = \frac{\lambda}{2\pi c^2}\left[\lambda^2 \frac{d^2 n}{d\lambda^2}\right] = \frac{\lambda^2}{2\pi c} D \tag{4.7.7c}$$

where β_2 = group velocity dispersion and D = delay time dispersion = $(\lambda/c) \cdot (d^2 n/d\lambda^2)$ with D usually specified in units of ps (pico-second) per km of length per nm of bandwidth used by the pulse at λ (see Fig. 4.8 for an example).

All can be derived from (4.7.7a) by using the chain rule of differentiation and the fact that $\partial n/\partial \omega = (\partial n/\partial \lambda)(\partial \lambda/\partial \omega) = -(\lambda^2/2\pi c)(\partial n/\partial \lambda)$.

We go back to the time domain by the use of (4.7.1): multiply each term of (4.7.6) by $(1/2\pi)e^{j(\omega-\omega_0)t} d(\omega - \omega_0)$, integrate, and recognize that the inverse transform of $[j(\omega - \omega_0)]^n E(z, \omega - \omega_0)$ is just the nth time derivative, $\partial^n E(z, t)/\partial t^n$ to obtain

$$\frac{\partial E(z, t)}{\partial z} + \beta_1 \frac{\partial E(z, t)}{\partial t} = j\frac{\beta_2}{2} \frac{\partial^2 E(z, t)}{\partial t^2} - j\delta\beta E(z, t) \tag{4.7.8}$$

where it is assumed that the perturbation does not depend upon ω to a significant degree.

A shift in variables makes (4.7.8) a bit more palatable. Let

$$\tau = (t - z/v_g) \tag{4.7.9}$$

and $z' = z$, which is equivalent to a transformation to a coordinate system moving at the group velocity (so that $d\tau = 0$ when $dt - dz/v_g = 0$). By using the chain rule of differentiation (again) yields

$$\frac{\partial}{\partial z} = \frac{\partial}{\partial z'}\left[\frac{\partial z'}{\partial z} = 1\right] + \frac{\partial}{\partial \tau}\left[\frac{\partial \tau}{\partial z} = \frac{-1}{v_g}\right] = \frac{\partial}{\partial z'} - \frac{1}{v_g}\frac{\partial}{\partial \tau} = \frac{\partial}{\partial z'} - \beta_1 \frac{\partial}{\partial \tau}$$

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial z'}\left[\frac{\partial z'}{\partial t} = 0\right] + \frac{\partial}{\partial \tau}\left[\frac{\partial \tau}{\partial t} = 1\right] = \frac{\partial}{\partial \tau} \quad \text{and} \quad \frac{\partial^2}{\partial t^2} = \frac{\partial^2}{\partial \tau^2}$$

*The index should have an "effective" as a subscript to indicate a frequency/wavelength dependence on n due to both the waveguide dimensions as well as the material dispersion.

Hence, (4.7.8) becomes*

$$\boxed{\frac{\partial E}{\partial z'} = +j\frac{\beta_2}{2}\frac{\partial^2 E}{\partial \tau^2} - j\delta\beta E \quad \text{or} \quad \frac{\partial^2 E}{\partial \tau^2} + j2\beta_2^{-1}\frac{\partial E}{\partial z'} = \frac{2\delta\beta}{\beta_2}E} \tag{4.7.10}$$

This is a fundamental result and needs discussion. In the next section we will discuss some of the interesting consequences of a finite value of $\delta\beta$, but, for now, we set the last term equal to zero to simplify things and to establish some of the consequences.

If $\beta_2 = 0$ or there is no group velocity dispersion, then (4.7.8) states that $\partial E/\partial z + \beta_1 \partial E/\partial t = 0$. This means that *any* envelope function can propagate any distance z away from the input plane without distortion. The only medium where $\partial v_g(\omega)/\partial\omega = 0$ is a vacuum and thus the standard buzz words about any function of the form $f(t - z/c)$ satisfying the wave equation applies. This is, of course, just a re-iteration of Sec. 1.3 and every elementary wave course.

For all other media where $\partial v_g(\omega)/\partial\omega \neq 0$, pulse shapes become distorted as they propagate. Notice that j appears on the right-hand side of (4.7.10) indicating that the distortion appears in the phase of the envelope E. A specific case of a Gaussian envelope function can be analyzed with ease after rewriting (4.7.10) to emphasize its similarity to (3.2.8) encountered earlier.

$$\frac{\partial^2 E}{\partial \tau^2} + j2\beta_2^{-1}\frac{\partial E}{\partial z'} = 0 \tag{4.7.11}$$

Equation (4.7.11) should be compared to (3.2.8) for the propagation of a one-dimensional (1-D) Gaussian beams.

$$\frac{\partial^2 \Psi}{\partial y^2} - j2k\frac{\partial \Psi}{\partial z} = 0 \tag{3.2.8}$$

There are some obvious similarities and differences: Equation (3.2.8) describes a characteristic Gaussian spatial beam whereas (4.7.11) describes a Gaussian temporal pulse, and (2) the free space propagation constant k is always positive for (3.2.8) whereas β_2^{-1}, the variation of the group velocity with frequency, can be either positive (normal dispersion) or negative (anomalous dispersion). The characteristic temporal pulse for a dispersive medium is a Gaussian of the form $\exp[-(t/\tau_p)^2]$ as we shall now demonstrate.

The details of the mathematical solution parallels that presented in Sec. 3.3 so only the differences and pitfalls are noted here. Let

$$E(z, \tau) = E_0 \exp\left\{-j\left[P(z) + \frac{\beta_2^{-1}\tau^2}{2q(z)}\right]\right\} \tag{4.7.12}$$

Substitute (4.7.12) into (4.7.11) and equate coefficients of τ^0 and τ^2 to zero, to find that $P'(z) = +j/q(z)$ and $q'(z) = -1$ by following the steps between (3.3.1) and (3.3.4a).

*Much of the literature cited will express (4.7.10b) with $\partial E/\partial z'$ having a negative sign because of the assumption of the field variation as $\exp[j(\beta z - \omega_0 t)]$ and hence the negative sign does not appear in (4.7.5).

Note that there is a difference in sign. The acceptable solutions for q and P are

$$q(z) = -\left\{z - jl_0[\text{sgn}(\beta_2^{-1})]\right\} \tag{4.7.13a}$$

$$e^{-jP(z)} = \frac{1}{[1 + (z/l_0)^2]^{1/4}} \exp\left[-j1/2 \tan^{-1} \frac{z}{l_0} (\text{sgn}(\beta_2^{-1})\right] \tag{4.7.13b}$$

where the sign of $\beta_2^{-1} = \text{sgn}(\beta_2^{-1})$ must be included in the integration of $q' = -1$ so as to guarantee that the envelope vanish at $\tau = \pm\infty$. The plane $z = 0$ is defined by $q(0) = jl_0$ and not by the coordinates of the input plane. Thus, the Gaussian pulse has the following functional form in the local time coordinates.

$$\begin{aligned} E = {} & \frac{E_0}{[1 + (z/l_0)^2]^{1/4}} \\ & \times \left\{\exp\left[-\frac{|\beta_2^{-1}|\tau^2}{2l_0} \cdot \frac{1}{1 + (z/l_0)^2}\right]\right\} \rightarrow \text{amplitude} = \exp[-\tau^2/\tau_p^2(z)] \\ & \times \left\{\exp\left[+j\left(\omega_0 t + \frac{\beta_2^{-1}\tau^2}{2l_0} \frac{z/l_0}{1 + (z/l_0)^2}\right)\right]\right\} \rightarrow \text{“chirp”} \\ & \times \exp\left[-j\left\{\beta(\omega_0)z + \tan^{-1}\left[\frac{z}{l_0} \text{sgn}(\beta_2^{-1})\right]\right\}\right] \end{aligned} \tag{4.7.14}$$

where the time harmonic variation $\exp[j\omega_0 t]$ and axial propagation variation $\exp[-j\beta(\omega_0)z]$ have been re-inserted. Note that the magnitude of $|\beta_2|$ appears in the amplitude of the pulse expression and thus the sign in the exponent is always negative. However β_2 appearing in the “chirp” line carries its own sign and can be either positive or negative.

The parameter l_0 is again a characteristic length for the increase of the pulse width by a factor of $\sqrt{2}$. It is evaluated by matching the pulse width τ_{p0} at $z = 0$, as predicted by the top line to that of the input pulse where the input is given by*

$$E(z = 0, t) = E_0 e^{-(t/\tau_{p0})^2} e^{j\omega_0 t}$$

for $z = 0$ where $\tau = t - z/v_g = t$.

$$\tau_{p0}^2 = 2|\beta_2|l_0 \qquad \text{or} \qquad l_0 = \frac{\tau_{p0}^2}{2|\beta_2|} \tag{4.7.15a}$$

and thus the envelope can be written as a Gaussian pulse whose width is expanding with z.

$$\tau_p^2(z) = \tau_{p0}^2(1 + (z/l_0)^2) = \tau_{p0}^2[1 + 4(\beta_2 z/\tau_{p0}^2)^2] \tag{4.7.15b}$$

*To minimize the volume of symbols, we have assumed a Fourier transform limited pulse, one whose bandwidth around the central frequency ω_0 is dictated by (4.7.1). For the general case, we should consider the case where ω_0 is replaced by $(\omega_0 + \Delta\omega t/T)$, permitting a “chirped” signal. This complicates the evaluation of l_0, since $z = 0$ is no longer the input plane but a (possibly fictitious) plane where q equals an imaginary quantity.

where τ_{p0} is the e^{-1} pulse-width at the input. If β_2 is very small, around 1.32 μm (see Fig. 4.8) or the input pulse width is long, then the characteristic distance l_0 for the increase of τ_p is also very long. But, the losses at 1.32 μm are larger than at 1.55 μm (see Fig 4.6), and a long pulse width means a limited data rate. If we use a short pulse required of a fast data rate and the "wrong" wavelength, then the distance is severely limited as the following example illustrates.

Example

Suppose a laser at 1.5 μm is used in a digital link on the fiber illustrated in Fig. 4.8 with a "1" bit represented by a Gaussian pulse with $\tau_p = 1$ ps (10^{-12}s), and of course, a "0" by the absence of energy in a similar time slot. From Fig. 4.8, the dispersion parameter $D \approx (110 \text{ ps/nm}) \div 11 \text{ km} = 10 \text{ ps/nm/km}$. Now $\beta_2 = (\lambda^2 D/2\pi c)$ by (4.7.7c) and thus the characteristic distance for the increase of τ_p by $\sqrt{2}$ is $z = l_0$ or

$$l_0 = \frac{\tau_{p0}^2 2\pi c}{2\lambda^2 D}$$

by (4.7.15a).

$$l_0 \frac{(1 \text{ ps})^2 \cdot 2 \cdot \pi \cdot (c = 300\ \mu\text{m/ps})}{2 \cdot (\lambda = 1.5\ \mu\text{m}) \cdot (\lambda = 1500 \text{ nm}) \cdot (D = 10 \text{ ps/nm/km})} = 0.0419 \text{ km} = 41.9 \text{ m}$$

which indicates that fiber optics is hardly worth the effort with this pulse width. If the length of the fiber link were 1 km, then the pulse width has expanded to

$$\tau_p^2(z = 1 \text{ km}) = \tau_{p0}^2[1 + ((z = 1 \text{ km})/l_0)^2] = 571\tau_{p0}^2$$

or $\tau_p(z = 1 \text{ km}) = 23.9$ ps. Hence this rate is limited to a data rate of $\sim 1/2\tau_p(1 \text{ km})$ or 20 Gbits/s. If the fiber link were 10 km, then the maximum data rate would be 2.0 Gbit/s. Note that the loss of 0.2 dB/km is *not* the limiting factor since even in the longer link, there is only a 2 dB loss ($T = 63\%$).

One of the differences between the Gaussian pulses and beams is the "1/4 power" of the premultiplier in the amplitude expression in (4.7.14) whereas (3.3.14) has the square root of a similar term. The difference arises because of three-dimensional (3-D) character of the beams (x, y, z) as opposed to the two-dimensional (2-D) pulses (t, z). The 1/4 power is necessary to conserve energy.

The second line of (4.7.14) describes the evolution of the phase of the "carrier" of field as it crosses a plane z. The phase is given by

$$\phi = \omega_0 t + \frac{(t - z/v_g)^2}{2\beta_2 l_0} \frac{z/l_0}{1 + (z/l_0)^2} = \omega_0 t + \frac{(t - z/v_g)^2}{\tau_{p0}^2 \operatorname{sgn}(\beta_2)} \frac{z/l_0}{1 + (z/l_0)^2}$$

The instantaneous frequency of the carrier at that plane is by definition $\omega(t) = d\phi/dt$.

$$\omega \triangleq \frac{d\phi}{dt} = \omega_0 + \frac{2(t - z/v_g)}{\tau_{p0}^2 \operatorname{sgn}(\beta_2)} \cdot \frac{z/l_0}{1 + (z/l_0)^2} \tag{4.7.15c}$$

At $z = 0$, the "chirp" provided by the second term disappears, indicating a Fourier-transform limited pulse as was assumed at the start. Since the pulse width increases with z, we would expect that the spectral bandwidth might decrease. However, because of the

"chirp" caused by the dispersion, the spectrum broadens to exceed that required to transmit the larger pulse width and thus is no longer a Fourier-transform limited signal. The order in which the parts of the spectrum arrive at z depends on the sign of β_2.

If β_2 is positive (normal dispersion), then the frequency of the wave increases with time indicating that the higher frequencies (or "blue" wavelengths) appear later in time or in the trailing part of the pulse. This is what would be expected: if $\partial^2\beta/\partial\omega^2 = \beta_2$ is positive then $(\partial v_g(\omega)/\partial\omega < 0)$ by (4.7.7c), indicating that the blue-shifted frequencies have a slower group velocity than the red-shifted ones. Hence the higher frequencies or "blue" arrives last. If $\beta_2 = \partial^2\beta/\partial\omega^2$ is negative $(\partial v_g(\omega)/\partial\omega > 0)$ corresponding to anomalous dispersion, then the situation is reversed with the high frequencies (blue) appearing on the leading edge and the lower ones (red) trailing. The two cases are sketched in Fig. 4.9.

It is important to agree with this line of reasoning because the $j\delta\beta\ E$ term, discarded after (4.7.11), can compensate for the dispersion introduced by the material medium if the wavelength is chosen to be in the anomalous dispersion region, which, for silica fibers, means that $\lambda > 1.27\mu$m. The example presented earlier should convince us that the dispersion is a most serious limitation and thus the solution posed in the next section on solitons may play a very important role in future fiber optic communications.

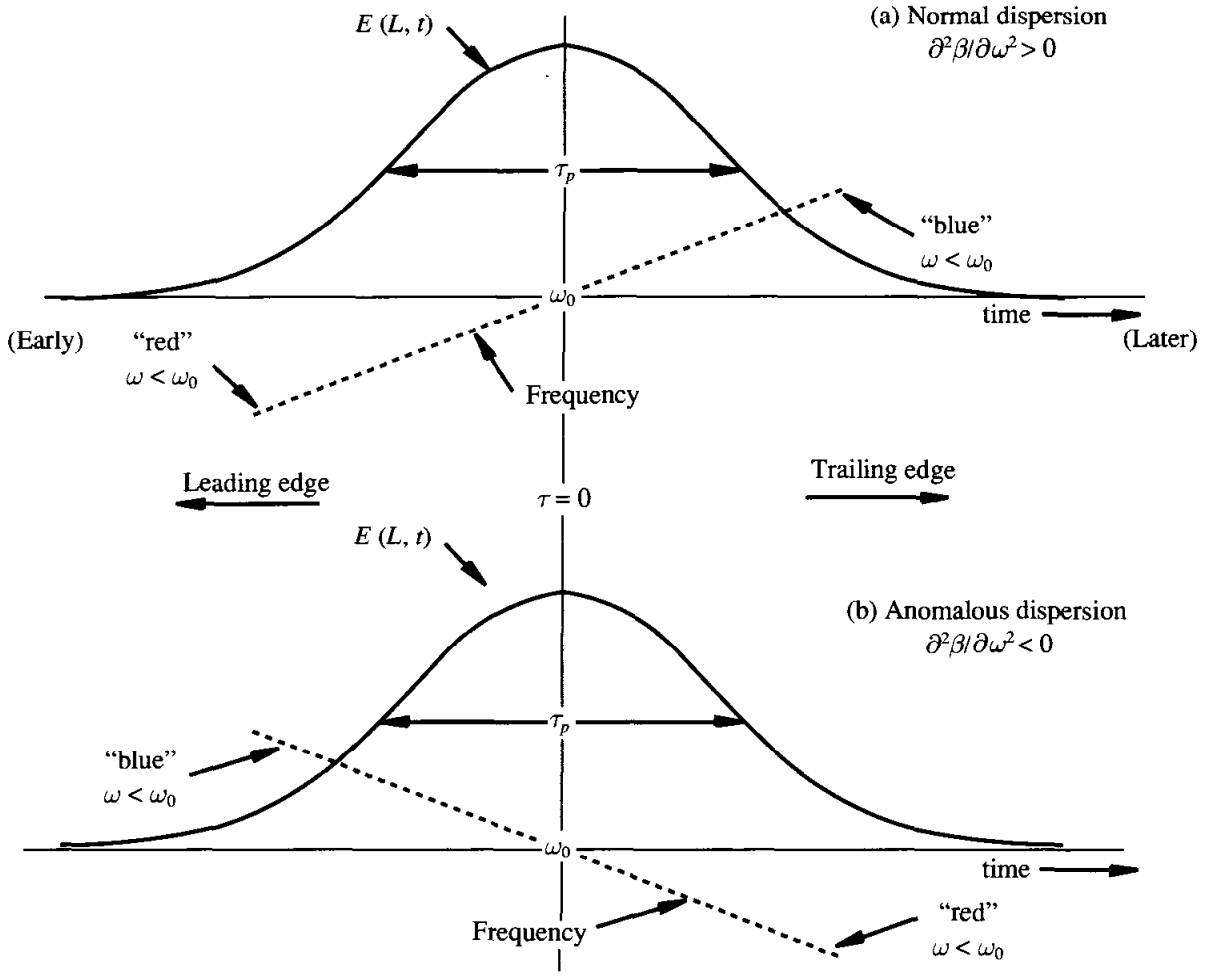


FIGURE 4.9. The "chirp" of pulses caused by the dispersive character of the medium.

4.8 OPTICAL SOLITONS

The use of optical solitons for fiber-optic communication is a rather recent addition to the technology and promises to cure many of the difficulties noted in the previous section. Under some rather well-defined circumstances, an optical nonlinearity (optical Kerr effect) can be used to compensate for the dispersion inherent in the materials and guidance of the mode and to permit high data rate over long distances. Fortunately, the existence of solitons does not require exotic devices or new technology. It does require a nonlinear medium whose index of refraction depends (slightly) on the local intensity

$$n(\omega, I) = n_0(\omega) + n_2' I(t) = n_0 + n_2|E|^2 \tag{4.8.1}$$

where a typical value* for $n_2' = 3.2 \times 10^{-16}$ cm^2/watt (for silica and is positive) given by Mollenauer and Stolen [37] in a review of soliton theory and experiments. Thus the intensity must be rather large before the nonlinear term is even 10^{-8} of the linear index, $n_0 = 1.5$. ($I = 10^{-8} \times 1.5/3.2 \times 10^{-16} = 47$ MW/cm^2). This sounds large, but the effective area of a mode in a fiber is $\sim 10^{-8}$ cm^2, and thus a power on the order of ~ 1 watt can change the index by 1 part in 10^8. For many applications, we would ignore such a small effect, but here this small change gets multiplied by a large number (the optical frequency times a length of many kilometers) and thus the product cannot be dismissed without further examination.

Let us start by making some rather imprecise statements to get the flavor of the phenomena. Consider a pulsed wave with an envelope $E(z, t)$ with a nominal center frequency ω_0 and phase constant $\beta(\omega_0)$ in a linear medium.

$$e(z, t) = E(z, t) \exp[j(\omega_0 t - \beta(\omega)z] \tag{4.8.2}$$

where $\beta(\omega) = \omega n(\omega, I)/c$. If we allow the index to be intensity dependent, and thus dependent upon time and space through $|E(z, t)|^2$, then $\beta(\omega_0)$ has the same characteristics. Hence, the exponential term has an extra time dependent factor.

$$\beta(\omega_0) \rightarrow \frac{\omega_0[n_0(\omega) + n_2' I(t)]}{c}$$

$$\exp[j(\omega_0 t + \beta(\omega_0)z)] \rightarrow \exp\left[+j\left\{\omega_0 t - \frac{\omega_0}{c}[n_0 + n_2' I(t)]z\right\}\right]$$

and thus the nominal instantaneous center frequency is "chirped" by virtue of the time dependent intensity in addition to that discussed in Sec. 4.7.

$$\omega(t) = \frac{d}{dt}\left\{\omega_0 t - \frac{\omega_0}{c}[n_0 + n_2' I(t)]z\right\}$$

$$= \omega_0 - z\frac{\omega_0}{c} n_2' \frac{dI(t)}{dt} = \omega_0 + \delta\omega(t)$$

If, for instance, $I(t) = I_0 \operatorname{sech}^2(t/\tau)$, then $dI(t)/dt = -(2I_0/\tau)\operatorname{sech}^2(t/\tau)\tanh(t/\tau)$ and thus the "chirp" $\delta\omega(t)$ has one sign on the leading edge and the opposite one for the

*The value of n_2' is most useful for the quick estimation of power as was just done, whereas n_2 (unprimed) is used for the theory to follow. Since $I = n_0[E^2/\eta_0]$, $n_2 = (n_0/\eta_0)n_2' = 1.2 \times 10^{-18}$ (cm/V)^2.

trailing part:

$$\delta\omega(t) = 2z\frac{\omega_0}{c}n_2'\frac{I_0}{\tau}\,\mathrm{sech}^2(t/\tau)\tanh(t/\tau)$$

where $t = 0$ is the center of the envelope. The fact that n_2' is multiplied by $\omega_0 z/c = 2\pi z/\lambda_0$, a very large number for $z \Rightarrow$km and $\lambda_0 \Rightarrow (\mu\text{m})$, means that this nonlinear effect can be significant. Such a pulse and its instantaneous frequency is sketched in Fig. 4.10 and should be compared to the "chirp" shown in Fig. 4.9 caused by the dispersion in the fiber.

A comparison between Figs. 4.9 and 4.10 indicates an intriguing possibility that the chirp caused by dispersion can be compensated by that caused by the intensity—the solutions that achieve this balance are called "solitons."* A close comparison of Figs. 4.9 and 4.10 shows that such a compensation can only be achieved if β_2 is negative (i.e., one operates in the region of anomalous dispersion). Fortunately, this requires the long wavelengths $\lambda > 1.27\ \mu$m where the fiber loss is very small.

We now must evaluate the quantity $\delta\beta$, which was set equal to 0 for the examples of the last section. Since the magnitude of the change in the index is small, we can use the perturbation approach of Sec. 4.6. For instance, if the complete field in the fiber is given by

$$e(x, y, z, t) = E(z, t)\psi(x, y)\exp[+j(\omega_0 t - \beta(\omega_0)z)]$$

then (4.5.10) provides an expression for $\delta\beta$ in terms of the perturbation in dielectric constant.

$$\delta\beta = \frac{\omega_0^2}{c^2}\frac{n_0}{\beta(\omega_0)}\frac{\iint(\psi^2)\delta n\,dA}{\iint(\psi^2)\,dA}$$

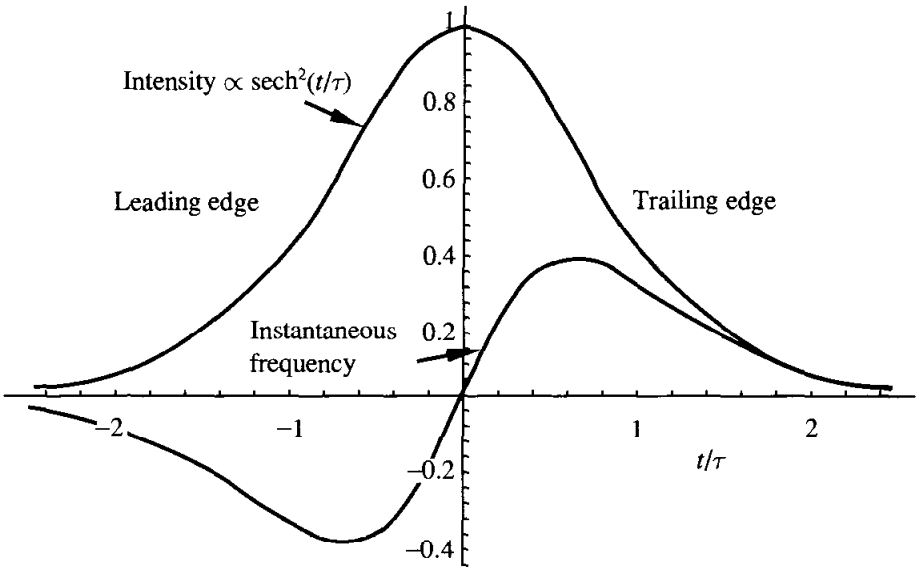


FIGURE 4.10. A sketch of the chirp introduced by the intensity [$\mathrm{sech}^2(t)$] dependent part of the index of refraction. Note that the higher frequencies (blue) appear on the trailing edge of the pulse.

*We will consider the solutions that propagate without any change in shape, the so-called fundamental soliton. Depending upon the input conditions, the actual shape will be periodic along z, but to show this fact requires more mathematics than is warranted here.

where $\delta n = n_2|E(z, t)|^2|\psi(x, y)|^2$. Thus, (4.7.8) becomes

$$\boxed{\frac{\partial E(z,t)}{\partial z} + \beta_1 \frac{\partial E(z,t)}{\partial t} = j(\beta_2/2)\frac{\partial^2 E(z,t)}{\partial t^2} - ja|E(z,t)|^2 E(z,t)} \tag{4.8.3}$$

where

$$a = \frac{\omega_0^2}{c^2}\frac{n_0}{\beta(\omega_0)}\frac{\iint n_2|\psi|^4\,dA}{\iint |\psi|^2\,dA} = \frac{\omega_0^2}{c^2}\frac{n_0 n_2}{\omega_0 n_0/c}\frac{\iint |\psi|^4\,dA}{\iint |\psi|^2\,dA} \simeq \frac{\pi n_2}{\lambda_0} \tag{4.8.4}$$

where the ratio of the two integrals has been replaced by 1/2, an exact relationship for a Gaussian beam but a very good approximation for any "bell-shaped" distribution of the field in the transverse plane.

To reduce the complexity of the equation we go to normalized quantities and use a transformation similar to that used in Sec. 4.7. Let

Let: $$T = \frac{t - z/v_g}{\tau_p} \tag{4.8.5a}$$

a time normalized to τ_p in a coordinate system moving at the group velocity v_g

Let: $$z' = \frac{1}{\tau_p^2}|\beta_2|z \triangleq \frac{\pi}{2}\frac{z}{z_0} \tag{4.8.5b}$$

where

$$z_0 = \frac{\pi^2 c\tau_p^2}{\lambda^2 D(\lambda)} = 0.322\frac{\pi^2 c(\Delta t_{1/2})^2}{\lambda^2 D(\lambda)}$$

z_0 = a normalized distance* which depends on the group velocity dispersion and defines the so-called "soliton period" z_0. The quantity $D(\lambda)$ is the delay dispersion such as that plotted in Fig. 4.8, and $\Delta t_{1/2} = 2\tau_p[ln(1 + \sqrt{2})]$ and is equal to the FWHM of the $\text{sech}^2(t/\tau_p)$ pulse, which is the functional form of a soliton.

Let: $$F = \tau_p\left\{\frac{a}{|\beta_2|}\right\}^{1/2} E \qquad \text{a normalized electric field} \tag{4.8.5c}$$

Equation (4.8.3) becomes (with the assumption that $\beta_2 = \partial^2\beta/\partial\omega^2$ is negative)

$$\boxed{j\frac{\partial F}{\partial z'} = \frac{1}{2}\frac{\partial^2 F}{\partial T^2} + |F|^2 F} \tag{4.8.6}$$

If the normalized variables z' and T are interchanged and the last term dropped, then (4.8.6) is precisely of the form of Schrödinger's equation. Hence the complete equation is called the Nonlinear Schrödinger equation (NLS).† This equation describes phenomena

*Most of the literature uses the last form for the definition of z_0.

†Many authors do choose the interchanged letter symbols to emphasize the relationship to Schrödinger's equation, while others use s for T, ξ for z', and u for F.

with intriguing and truly amazing characteristics, so amazing that it is worthwhile deriving the so-called "fundamental soliton" solution.

Let
$$F = e^{-jz'/2}\Gamma(T) \tag{4.8.7}$$

This substitution transforms (4.8.6) to

$$e^{-jz'/2}\left\{\frac{1}{2}\Gamma(T) = \frac{1}{2}\frac{d^2\Gamma(T)}{\partial T^2} + \Gamma^3(T)\right\} \tag{4.8.8}$$

which can be integrated by multiplying by $(2d\Gamma/dT)\cdot dT$

$$\int_{-\infty}^{T}\Gamma\frac{d\Gamma}{dT}dT = \int_{-\infty}^{T}\frac{d\Gamma}{dT}\left(\frac{d^2\Gamma}{dT^2}dT\right) + 2\int_{-\infty}^{T}\Gamma^3\frac{d\Gamma}{dT}dT$$

$$\frac{1}{2}\Gamma^2(T) = \frac{1}{2}\left(\frac{d\Gamma}{dT}\right)^2 + \frac{1}{2}\Gamma^4(T) \quad \text{or} \quad \left(\frac{d\Gamma}{dT}\right)^2 = \Gamma^2(1-\Gamma^2) \tag{4.8.9}$$

where the obvious boundary condition that $\Gamma(-\infty) = 0$ has been used. Now invoke the next obvious boundary condition for a "pulse" solution: Γ must be a maximum (i.e., $d\Gamma/dT = 0$) at the center of the pulse ($T = 0$). Hence

$$\Gamma(0) = 1 \tag{4.8.10}$$

The last statement should shock the reader since it states that there is a "unique" amplitude to this solution. There are more surprises in store.

We can integrate (4.8.9) from $T = 0$ (where $\Gamma = 1$) to T by using the negative square root (since the amplitude obviously decreases from a maximum).

$$\int_{1}^{\Gamma(t)}\frac{d\Gamma}{\Gamma(1-\Gamma^2)^{1/2}} = -\int_{0}^{T}dT \tag{4.8.11}$$

which can be integrated by [Math Handbook (14.241)]

$$-ln\left\{\frac{1+(1-\Gamma^2)^{1/2}}{\Gamma}\right\} = -T$$

or (4.8.12)

$$\Gamma(T) = \frac{1}{\cosh T} = \operatorname{sech} T$$

and thus the normalized electric field is

$$\boxed{F_1 = e^{-jz'/2}\operatorname{sech}(T)} \tag{4.8.13a}$$

The actual electric field is given by

$$\boxed{E = \left\{\frac{|\beta_2|}{a}\right\}^{1/2}\frac{1}{\tau_p}\operatorname{sech}\left[\frac{t - z/v_g}{\tau_p}\right]\exp\left[-j\frac{\pi z}{4z_0}\right]} \tag{4.8.13b}$$

and thus the intensity will have a $\text{sech}^2(t/\tau_p)$ variation as has been anticipated.

There are some rather interesting "conservation laws"(an infinite number according to Zakharov and Shabat):

1. The "area" of this fundamental soliton is given by

$$\text{"area"} = \int_{-\infty}^{+\infty} |F_1|\, dT = \int_{-\infty}^{+\infty} \text{sech}\, T\, dT = 2\tan^{-1}(e^T)\Big|_{-\infty}^{+\infty} = \pi \quad (4.8.14a)$$

which is independent of τ_p, and therefore this fundamental soliton is sometimes called a "pi" pulse, although such a terminology is usually reserved for quantum phenomena, which is to be discussed in Chapter 14.

2. The "normalized energy" in this pulse is

$$\int_{-\infty}^{+\infty} F_1 \cdot F_1^*\, dT = \int \text{sech}^2\, T\, dT = \tanh T\Big|_{-\infty}^{+\infty} = 2 \quad (4.8.14b)$$

Two interesting transformation laws follow:

3. We can also show that a function of the form

$$F_2(T, Z') = A[\text{sech}(AT)] \exp\left[-j\frac{A^2 z'}{2}\right]$$

is also a solution and has the same "area" independent of the amplitude A.

$$\text{"area"} = \int_{-\infty}^{+\infty} |F_2|\, dT = A\int_{-\infty}^{+\infty} \text{sech}(AT)\, dT = \int_{-\infty}^{+\infty} \text{sech}(AT)\, d(AT) = \pi$$

4. Another solution is of the form:

$$F_3(T, z') = F_2(T + z'/u, z') \exp\left[j\left\{\frac{1}{u}\left(T + \frac{z'}{2u}\right)\right\}\right] \quad (4.8.14c)$$

Hasegawa and Tappert [25] have shown that these pulses are amazingly stable and reasonably easy to excite. Computer experiments indicate that if we use enough energy in the pulse to exceed the normalized value of (4.8.14b), then the pulse actually contracts as it propagates to reappear in its original width after the characteristic length z_0.

Let us take an example to put some life in the dull mathematics. Our example uses some of the data presented in the paper by Hasegawa and Kodama [35], but the calculations are modified to make the arithmetic tolerable for a textbook.

We consider a single mode silica fiber ($n_0 = 1.5$) carrying an optical power confined to the central core so that the function $\psi(x, y)$ can be approximated by a Gaussian. (It is not a Gaussian but any bell-shaped curve can be approximated by one.)

$$\psi(x, y) = \psi(r) = \exp[-r^2/w^2]$$

where w is the characteristic scale length for the mode. Thus the parameter a given by (4.8.4) can be evaluated.

$$a = \frac{\omega_0^2}{c^2} \cdot \frac{n_0 n_2}{\beta(\omega_0) = \omega_0 n_0 / c} \cdot \frac{\iint \left\{ |\psi|^4 = \exp[-4r^2/w^2] \right\} r\, dr\, d\phi}{\iint \left\{ |\psi|^2 = \exp[-2r^2/w^2] \right\} r\, dr\, d\phi}$$

The quotient of the integrals is 1/2, a result peculiar to the Gaussian mode assumption but an excellent approximation for any "bell-shaped" mode structure, and a becomes

$$a = \frac{\pi n_2}{\lambda_0}$$

where $n_2 = 1.2 \times 10^{-22} m^2/V^2 = 1.2 \times 10^{-18}\ \text{cm}^2/V^2$.

Equation (4.8.10) states that the normalized peak electric field $|F| = |\Gamma(0)| = 1$. Thus, we use (4.8.5c) to evaluate the peak electric field in normal units of V/m or V/cm and that requires the value for $\beta_2 = \partial^2\beta/\partial\omega^2$, which, in turn, comes from experiment.

$$\beta_2 = \frac{\lambda}{2\pi c^2} \left\{ \lambda^2 \frac{d^2 n}{d\lambda^2} \triangleq f(\lambda) \right\} \qquad \text{Recall Eq. 4.7.7c}$$

where Hasegawa and Kodama [35] approximate the dimensionless function $f(\lambda)$ by using the data in Marcuse's [32] paper.

$$f(\lambda) = -5.3 \times 10^{-2} \left[1 - \frac{1.27 \mu m}{\lambda_0(\mu m)} \right]$$

which applies only to fiber where $\beta_2 = 0$ at $\lambda = 1.27\ \mu$m such as that shown in Fig. 4.8. If the wavelength were 1.3 μm, then $f(\lambda) = 1.223 \times 10^{-3}$ and $\beta_2 = -2.81 \times 10^{-29}\ \text{sec}^2/\text{cm}$. Thus the peak electric field is

$$E_{\text{peak}} = \left\{ \frac{\beta_2 \lambda_0}{\pi n_2} \right\}^{1/2} \frac{\{|F| = \Gamma(0) = 1\}}{\tau_p} = \frac{3.11 \times 10^4}{\tau_p(\text{ps})}\ V/\text{cm}$$

The intensity implied by such a field is very large

$$I_{\text{peak}} = n_0 \left\{ \frac{E_{\text{peak}}^2}{2\eta_0} \right\} = \frac{n_0 \beta_2 \lambda_0}{2\pi n_2} \frac{1}{\tau_p^2} \frac{1}{\eta_0} = \frac{1.93 \times 10^6}{\tau_p^2(\text{ps})}\ (\text{W/cm}^2)$$

but the effective area of the mode, $\pi w^2/2$, is also small. Hence the peak power is quite reasonable.

$$P_{\text{peak}} = \iint I_{\text{peak}} \cdot r\, dr\, d\phi = \frac{n_0 \beta_2 \lambda_0}{2\pi n_2} \frac{1}{\tau_p^2(\text{ps})} \frac{1}{\eta_0} \frac{\pi w^2}{2}$$

$$= \frac{0.76}{\tau_p^2}\ (\text{in watts/cm}^2;\ \text{if } w = 5\ \mu\text{m})$$

Thus, a 1 watt peak signal is quite adequate to generate a 1 ps soliton.

The normalized energy contained in the soliton is 2 by (4.8.14b) but in real life units we have

$$p(t) = P_{\text{peak}} \operatorname{sech}^2(t/\tau_p)$$

and hence

$$W = \int_{-\infty}^{+\infty} p(t)\,dt = (P_{\text{peak}})\tau_p \tanh(t/\tau_p)\Big|_{-\infty}^{+\infty} = 2P_{\text{peak}}\tau_p$$

or: $$W = \frac{n_0\beta_2\lambda_0}{2\pi n_2}\frac{w_s^2}{\eta_0}\frac{1}{\tau_p} = 1.52 \text{ pJ(for } \tau_p = 1 \text{ ps, } w = 5\ \mu\text{m)} = 9.5 \times 10^6 \text{eV}$$

At $\lambda = 1.3\ \mu$m, $hv/e = 0.9537$ eV/photon and thus this soliton only carries $\sim 10^7$ photons.

By careful design of the fiber, Gambling, Matsumara and Ragdale [27] have shown that the wavelength at which $\beta_2 = 0$ can be shifted closer to the magic 1.55 μm value so as to take advantage of the minimum loss exhibited by fibers at that wavelength.

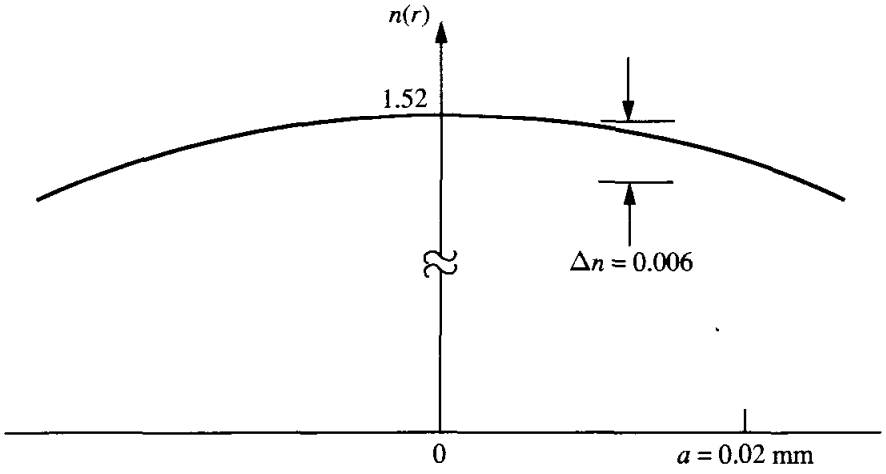
Lonngren's book [44] identifies two other nonlinear equations that admit solitary solutions, the Korteweg-deVries equation and the Sine-Gordon equation, and points out (page 219) that the initial observation of a soliton was observed as a water wave (and chased on horseback!) by John Scott Russell in 1834. Solitons have some truly amazing characteristics and it is not clear that all of the useful properties have been identified.

PROBLEMS

4.1. A typical fiber with a quadratic variation of index $n(r) = n_0 - \Delta n(r/a)^2$ has the following parameters: $n_0 = 1.5$, $\Delta n = 8 \times 10^{-3}$, and $a = 20\ \mu$m. Assume that all modes with numbers $(m, p \leq 20)$ are excited more or less uniformly at $z = 0$ and with a common phase. Calculate the phase shift of the (m,p) mode relative to (0,0) mode at $z = 1$ km. Estimate the distribution of phases of the various modes relative to $m = p = 0$.

4.2. Consider an optical fiber whose core has a parabolic index variation in the manner shown below. The index of refraction on the axis, $n(r = 0) = n_0 = 1.52$; the diameter of the core is 40 μm; and the change in index is $\Delta n = 0.006$. Assume $\lambda_0 = 1.315\ \mu$m.

- **(a)** What is the scale length l in the formula for the index?
- **(b)** What is the numerical value of the spot size in this fiber?
- **(c)** Suppose the $\text{TEM}_{0,0}$ and $\text{TEM}_{0,1}$ modes were excited with equal powers and with a common phase at $z = 0$. Find the distance z_1 where the two modes are 180°out of phase.
- **(d)** Make a careful sketch of the total intensity, $E \cdot E^*/2\eta_0$, as a function of $(x, y = 0, z = 0)$ and for $(x, y = 0, z = z_1)$. Assume that both modes have the same sense of polarization.



(e) What is the value of the index at $r = w$?

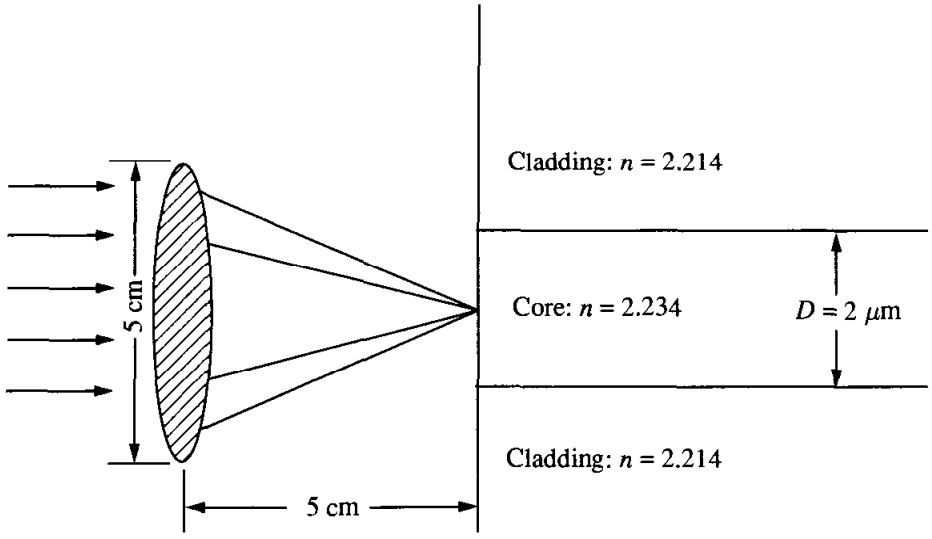
4.3. Find the normalized coordinates $(x/w, y/w)$ at which the electric field of the $TEM_{1,0}$ mode is a maximum.

4.4. The complex beam parameter in a parabolic index fiber is given by

$$\frac{1}{q} = -\frac{(A - D)}{2B} - j\frac{\left\{1 - [(A + D)/2]^2\right\}^{1/2}}{B}$$

which was derived from the $ABCD$ matrix via a quadratic equation. However the plus or minus sign in front of the imaginary part of this definition has been changed to a simple minus sign. For what physical reason or reasons can we drop the positive portion of the plus or minus sign before the imaginary part of $1/q$? (Do not just say that the plus sign is physically impossible. Explain why it is impossible.)

4.5. A fiber is excited in the manner shown below (obviously not drawn to scale).



(a) If the incoming wave uniformly fills a circular lens, what fraction of the power can be captured and guided by the fiber? (Assume a round fiber and use geometric optics for this part of the problem.) If this were a slab fiber excited by a cylindrical lens, then what fraction of the incident power would be captured and guided?

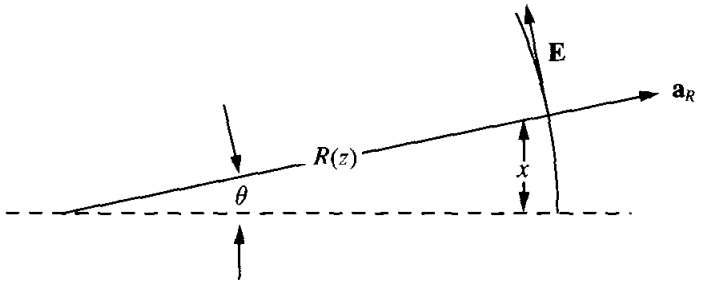
(b) What is the R# for this fiber?

(c) If only a single TE (or TM) mode is to be guided by the slab fiber, what is the shortest wavelength (i.e., highest frequency) that can be used?

(d) How thick must the cladding be in order for the field at the outside radius to be 10^{-6} of that at the edge of the core?

(e) What is the phase constant β (in cm^{-1}) for the TE_0 mode?

4.6. The phase front of a Gaussian beam is spherical with a radius of curvature $R(z)$ (3.3.11) and the actual electric field is perpendicular to this radius (i.e., $\mathbf{E} \cdot \mathbf{a}_R = 0$)



(a) Show that the z component of $\mathbf{E}$ can be expressed in terms of the transverse component E_t[i.e., (3.3.14)]$\times \tan\theta$, where $\tan\theta = x/R(z)$.

(b) Find the value of x which makes $|E_z|$ a maximum and evaluate that maximum in terms of the transverse field.

(c) The radius of curvature $R(z)$ is infinite at $z = 0$ and $z = \infty$. Hence, there is some place in between these extremes where E_z is a maximum. Find the ratio (z/z_0) that maximizes E_z and compare that maximum to $|E_t|$. Evaluate for $n = 1$, $\lambda_0 = 1\mu\text{m}$, and $w_0 = 1\text{mm}$.

4.7. Derive a formula for $(\beta_1 = \partial\beta/\partial\omega)$ and $(\beta_2 = \partial^2\beta/\partial\omega^2)$ for the $TEM_{0,0}$ mode in the graded index fiber discussed in the text.

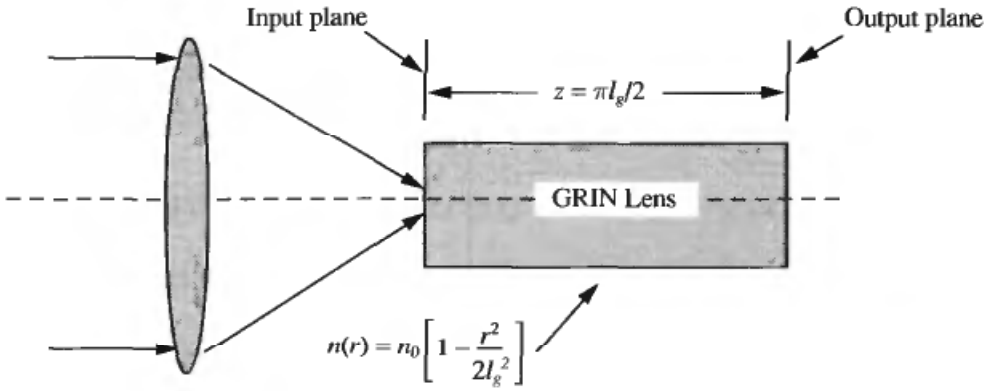
4.8. The ray matrix for a GRIN lens is given below.

$$T = \begin{bmatrix} \cos(z/l) & \dfrac{l}{n_0}\sin(z/l) \\ -\dfrac{n_0}{l}\sin(z/l) & \cos(z/l) \end{bmatrix}$$

where

$$n(r) = n_0\left[1 - \frac{r^2}{2l^2}\right]$$

where l is a parameter specifying the variation of the index with respect to the radial coordinate, and z is the length of the lens. Such a lens of length $z = \pi l/2$ is excited by an H-G beam characterized by a planar wavefront and an input spot-size parameter $\pi w_{01}^2/\lambda$ at $z = 0$ using geometry shown on the diagram below. What are the characteristics of the output beam (in terms of the input parameters)? R_2 =?; $\pi w_2^2/\lambda$ =?



4.9. The purpose of this problem is to estimate the conditions under which a high power quasi-CW beam will *self-focus* and lead to catastrophic damage in an otherwise uniform dielectric.

(a) Start with the wave equation with an index that is a function of the local intensity or field [c.f. (4.8.1)], $n = n_0 + n_2' I = n_0 + n_2|E|^2$, assume a wave propagating as $E = E_0\psi(x, y, z)\exp[-j\beta z]$, and expand the wave equation in the manner used in Sec. 3.2 to derive

$$\nabla_t^2\psi - 2\beta\frac{\partial\psi}{\partial z} + \frac{\partial^2\psi}{\partial z^2} + \left\{\left(\frac{\omega n_0}{c}\right)^2\left[1 + \frac{2n_2}{n_0}E_0^2|\psi|^2\right] - \beta^2\right\}\psi = 0 \tag{P4.1}$$

(b) If we choose $\beta^2 = (\omega n_0/c)^2 = k^2$ and make the usual paraxial approximation [c.f. (3.2.7) → (3.2.8)], show that (P4.1) becomes

$$\nabla_t^2\psi - j2k\frac{\partial\psi}{\partial z} + \left[2k^2\frac{n_2}{n_0}E_0^2|\psi|^2\right]\psi = 0 \tag{P4.2}$$

(c) Assume $\nabla_t^2\psi = \partial^2\psi/\partial x^2$, a one-dimensional case for (P4.2) and go to normalized variables defined by $u = x/w$; $z' = z/kw$; and $F = kw[2n_2E_0^2/n_0]^{1/2}\psi$. Show that the differential equation obeyed by F is identical to (4.8.6) provided the above variables are used. Once this is done, the solutions indicated by (4.8.14b) and (4.8.14c) can be used. However there is still the limitation of a 1-D beam assumption.

(d) Return to (P4.1) and assume a Taylor series for the term $|\psi|^2 = \exp[-2r^2/w^2] \approx 1 - 2r^2/w^2$ (but only for that term) and let β^2 =

$k^2(1 + 2n_2E_0^2/n_0) \approx k^2$ to obtain

$$\nabla_t^2\psi - j2k\frac{\partial\psi}{\partial t} - \left[4\frac{k^2}{w^2}\frac{n_2E_0^2}{n_0}r^2\right]\psi = 0 \qquad \text{(P4.3)}$$

(e) This equation has the same functional form as that for propagation in a parabolic graded index medium with $n(r) = n_0(1 - r^2/2l^2$. Derive an expression for the scale length l^2 for the self-focusing medium. (Ans.: $l^{-2} = 4n_2E_0^2/w^2n_0$.)

(f) We also know expressions for the elements in the $ABCD$ matrix for a graded index medium [c.f. (2.12.11)]. Assume $l \gg d$ in (2.12.11) and evaluate T. *Show that*

$$T \sim \begin{bmatrix} 1 - \dfrac{2n_2(E_0d)^2}{w^2n_0} & d \\ -\dfrac{4n_2E_0^2d}{w^2n_0} & 1 - \dfrac{2n_2(E_0d)^2}{w^2n_0} \end{bmatrix}$$

and evaluate the focal length f in terms of the power (not intensity) carried by the Gaussian beam.

$$\left(\text{Ans.: } f = \frac{\pi}{8}\frac{n_0w^4}{n_2'Pd}\right)$$

(g) The ray matrix approach suffers because w is a function of z and an unknown one at that. Hence the development of a ray matric from (2.12.6) is not straightforward. However, we can assume $\psi = \exp\left\{-j[P(z) + kr^2/2q(z)]\right\}$ as was done in Sec. 3.3 and solve for the *critical field* or intensity so that the focussing effect balances the tendency for the beam to spread and thus $w \neq f(z)$. What is this critical value of the intensity?

$$\left(\text{Ans.: } I(0) = \frac{\lambda^2}{4\pi^2n_0w^2}\cdot\frac{1}{n_2'}\right)$$

4.10. Calculate the numerical aperture of a step index fiber with $n_1 = 1.48$ and $n_2 = 1.46$.

4.11. Suppose a GaAs laser oscillates at three equal amplitude closely spaced frequencies around *8000* Å producing a pulse 5 ns wide. The spacing between these frequencies is 50 GHz.

(a) Use the data in Fig. 4.8 to estimate differences in arrival times of the lowest and highest frequencies for a fiber of 11 km length.

(b) Use the data of Fig. 4.6 to estimate the attenuation in 11 km of the low loss fiber.

4.12. Consider the propagation of a Gaussian pulse in a nondispersive but slightly nonlinear medium where $n = n_0 + n_2|e(t)|^2$

$$e(t) = E_0\left\{\exp\left[-\frac{(t - z/v_g)^2}{2\tau_p^2}\right]\right\}e^{j\omega_0t}$$

where the braces are the input envelope. Use (4.8.3) to show that this pulse will develop a "chirp" by the following procedure.

(a) Use the coordinate transformation indicated by (4.7.9) to convert (4.8.3) to

$$\frac{\partial E}{\partial z'} = ja|E|^2E = j\frac{k}{2}\frac{n_2E_0^2}{n_0}\exp\left[-\left(\frac{\tau}{\tau_p}\right)^2\right]E$$

(b) Assume $E \sim A(\tau)\exp +j\phi(z')$, $A(\tau)$ = real, and solve for $A(\tau)$, $\phi(z')$.

(c) The frequency "chirp" appears as a correction to the above equation in the form $e(t) = \{\text{new envelope}\}\exp[j\omega(t)t]$. What is $\omega(t)$?

REFERENCES AND SUGGESTED READINGS

1. A.G. Fox and T. Li, "Resonant Modes in a Maser Interferometer," *Bell Syst. Tech. J. 40*, 453–488, Mar. 1961.
2. H. Kogelnik, "Imaging of Optical Modes—Resonators with Internal Lenses," *Bell Syst. Tech. J. 44*, 455–494, Mar. 1963.
3. D. Marcuse, *Light Transmission Optics* (Princeton, N.J.: D. Van Nostrand, 1972).
4. D. Marcuse and S.E. Miller, " Analysis of a Tubular Gas Lens," *Bell Syst. Tech. J. 43*, 1759-1787, July 1964.
5. D. Marcuse, "Theory of a Thermal Gradient Gas Lens," *IEEE Trans. MTT-13*, 734–739, Nov. 1965.
6. E.A.J. Marcateli and R.A. Schmeltzer, "Hollow Metallic and Dielectric Waveguides," *Bell Syst. Tech. J. 43*, 1783–1809, July 1964.
7. P.K. Tien, "Integrated Optics and New Wave Phenomena in Optical Waveguides," *Rev. Mod. Phys. 49*, 361–420, Apr. 1977.
8. Special Issue on Light Wave Communications, *Phys. Today 29*(5), May 1976.
9. W.S. Boyle, "Light Wave Communications," *Sci. Am. 237*, Aug. 1977.
10. D. Botez and G.J. Herskowitz, "Components for Optical Communications Systems: A Review," *Proc. of IEEE 68*, 689–731, June 1980. *Note:* The bibliography of this reference contains 640 references.
11. D.L. Franzen and G.W. Day, "Measurement of Propagation Constants Related to Material Properties in High-Bandwidth Optical Fibers," *IEEE J. Quant. Electron. QE-15*, 1409–1414, Dec. 1979.
12. K.A. Jones, *Introduction to Optical Electronics* (New York: Harper and Row, 1987).
13. R.G. Hunsperger, *Integrated Optics: Theory and Technology*, Springer Series in Optical Sciences (New York: Springer-Verlag, 1984).
14. H.A. Haus, *Waves and Fields in Optoelectronics* (Englewood Cliffs, N.J.: Prentice Hall, 1984), Chap. 6.
15. G. Keiser, *Optical Fiber Communications* (New York: McGraw-Hill, 1983).
16. A.H. Cherin, *An Introduction to Optical Fibers* (New York: McGraw-Hill, 1983.)
17. D.. Marcuse, *Theory of Dielectric Optical Waveguides* (New York: Academic, 1974).
18. E.A. Lacy, *Fiber Optics* (Englewood Cliffs, N.J.: Prentice Hall).

19. J. Yamada, "High Speed Optical Pulse Transmission at 1.29 μm Wavelength using Low-loss Single Mode Fibers." *IEEE J. Quant. Electron. QE-14*, 791, 1978.
20. R.E. Collin, *Field Theory of Guide Waves* (New York: McGraw-Hill, 1960) pages 128–129.
21. T. Miya, Y. Terunuma, T. Hosaka, and T. Miyashita, "Ultimate Low-Loss Single Mode Fiber at 1.55 μm," *Electron. Lett. 15*, 106,1976.
22. J.C. Slater, *Microwave Electronics* (New York: D. Van Nostrand, Inc., 1950).
23. H.A. Haus, "Higher Ordered Trapped Light Beam Solutions," *Appl. Phys. Lett. 8*, 128–129, 1966.
24. V.E. Zakharov and A.B. Shabat, "Exact Theory of Two-Dimensional Self-Focusing and One-Dimensional Self-Modulation of Waves in Nonlinear Media," *Sov. Phys. JETP 34*, 62–69, 1972.
25. A. Hasegawa and F. Tappert, "Transmission of Stationary Nonlinear Optical Pulses in Dispersive Dielectric Fibers, 1, Anomalous Dispersion," *Appl. Phys. Lett. 23*, 142–144,1973.
26. J. Satsuma and N. Yajima, "Initial Value Problems of One-Dimensional Self-Modulation of Nonlinear Waves in Dispersive Media," *Suppl. Progr. Theor. Phys. 55*, 284–306, 1974.
27. W.A. Gambling, H. Matsumara, and C.M. Ragdale, "Zero Total Dispersion in Graded-Index Single-Mode Fibers," *Electron. Lett. 15*, 474–476, 1979.
28. H. Tsuchiya and N. Imoto, "Dispersion-Free-Single-Mode Fiber in 1–5 μm Wavelength Region," *Electron, Lett. 15*, 476–478, 1979.
29. L. Jeunhomme, "Dispersion Minimization in Single-Mode Fibers Between 1–3 μm and 1–7 μm," *Electron. Lett., 15*, 478–479, 1979.
30. D.M. Bloom, L.F. Mollenauer, C. Lin, D.W. Taylor, and A.M. DelGaudio, "Direct Demonstration of Distortionless Picosecond-Pulse Propagation in Kilometer-Length Optical Fibers," *Opt. Lett. 4*, 297–299, 1979.
31. T. Miya, Y. Terunuma, T. Hosaka, and T. Miyashita, "Ultimate Low-Loss Single-Mode Fiber at 1–55 μm," *Electron. Lett. 15*, 106–108, 1979.
32. D. Marcuse, "Pulse Distortion in Single-Mode Fibers," *Appl. Opt. 19*, 1653–1660, 1980.
33. A. Hasegawa, "Self-Confinement of Multimode Optical Pulse in a Glass Fiber," *Opt. Lett. 5*, 416–417, 1980.
34. L.F. Mollenauer, R.H. Stolen, and J.P. Gordon, "Experimental Observation of Picosecond Pulse Narrowing and Solitons in Optical Fibers," *Phys. Rev. Lett. 45*, 1095–1098, 1980.
35. A. Hasegawa and Y. Kodama, "Signal Transmission by Optical Solitons in Monomode Fiber," *Proc. IEEE 89*, 1145–1150, 1981.
36. R.H. Stolen, L.F. Mollenauer, and W.J. Tomlinson, "Observation of Pulse Restoration at the Soliton Period in Optical Fibers," *Opt. Lett 8*, 186–188, 1983.
37. L.F. Mollenauer and R.H. Stolen, "Solitons in Optical Fibers," *Fiberopt. Tech.* 193–198, April 1983.
38. L.F. Mollenauer, R.H. Stolen, and J.P. Gordon, "Extreme Picosecond Pulse Narrowing by Means of Soliton Effect in Single-Mode Optical Fibers," *Opt. Lett. 8*, 289–291, 1983.
39. J.P. Gordon, "Interaction Forces Among Solitons in Optical Fibers," *Opt. Lett. 8*, 596–598, 1983.
40. L.F. Mollenauer and R.H. Stolen, "The Soliton Laser," *Opt. Lett. 9*, 13–15, 1984.
41. A. Hasegawa, "Numerical Study of Optical Soliton Transmission Amplified Periodically by the Stimulated Raman Process," *Appl. Opt. 23*, 3302–3309, 1984.
42. R. Boyd, "Processes Resulting From the Intensity-Dependent Refractive Index," in *Nonlinear Optics* (New York: Academic, 1992).

43. K. Lonngren and A. Scott, *Solitons in Action* (New York: Academic, 1978).
44. Karl E. Lonngren, *Introduction to Physical Electronics*, Chapter 11 (Boston: Allyn and Bacon, 1988).
45. Hermann A. Haus, "Molding Light into Solitons", *IEEE Spectrum*, 48–53, March 1993.
46. H.C. Panish, Jr., and M.B. Panish, *Heterostructure Lasers Part A: Fundamental Principles* (New York: Academic Press, 1978).



Optical Cavities

5.1 INTRODUCTION

A most important part of any laser is the feedback system. In the simplest case, this consists of two mirrors with the active medium located between them. It is the purpose of this chapter to relate the field parameters of the characteristic cavity modes to the geometry of the feedback structures.

5.2 GAUSSIAN BEAMS IN SIMPLE STABLE RESONATORS

In Chapter 3, Gaussian beams were discussed in exhaustive detail. We may ask, How are the parameters of the Gaussian beam determined by a real-life cavity?

To answer this question, consider Fig 5.1, which shows the expansion of the $TEM_{0,0}$ mode [i.e., the variation of $w(z)$] and also the curvature of the phase front, $R(z)$. We choose our coordinate system such that the minimum spot size occurs at $z = 0$, where the wave front has an infinite radius of curvature. While we do *not* know the value of w_0, we do know that it exists. But if we did know w_0, we *could* predict everything about the beam.

Now suppose that we choose a set of mirrors and adjust their positions and curvature so that their surfaces exactly *match* the surfaces of constant phase of the beam. For instance, suppose that we choose a flat mirror $R_1 = \infty$ and a curved mirror of radius R_2. Since the

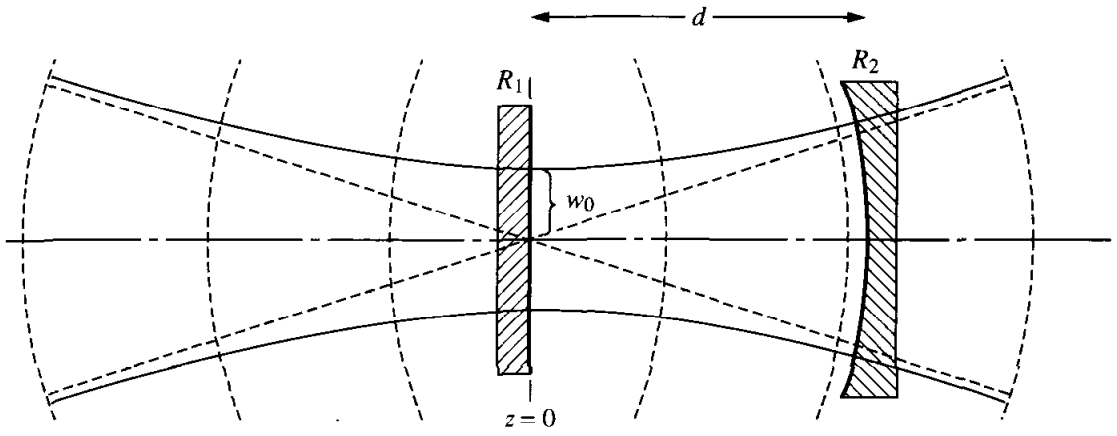


FIGURE 5.1. Expansion of a Gaussian beam.

rays associated with this Gaussian beam impinge perpendicular to the mirror surface, they will be reflected back on themselves and return to the other. Thus we have obtained a self-consistent description of a normal mode of this cavity. We need three formulas from Chapter 3 to relate the beam parameters to the mirror specifications:

$$R(z) = z\left[1 + \left(\frac{z_0}{z}\right)^2\right] \tag{3.3.11}$$

$$w^2(z) = w_0^2\left[1 + \left(\frac{z}{z_0}\right)^2\right] \tag{3.3.10}$$

with

$$z_0 = \frac{\pi n w_0^2}{\lambda_0} \tag{3.3.7}$$

Thus we choose the value of w_0 such that the equiphase surfaces do coincide with our choice of mirrors. For the example shown in Fig. 5.1, the flat mirror obviously *matches* the phase surface at $z = 0$. We now *force* the phase surface to match the curved mirror at $z = d$.

$$R(z) = z\left[1 + \left(\frac{z_0}{z}\right)^2\right]$$

or (5.2.1)

$$R(d) = R_2 = d\left[1 + \left(\frac{z_0}{d}\right)^2\right]$$

Solving this equation for the unknown z_0 is trivial:

$$z_0 = \frac{\pi w_0^2}{\lambda} = (dR_2)^{1/2}\left(1 - \frac{d}{R_2}\right)^{1/2} \tag{5.2.2}$$

In obtaining this last result, we have used (3.3.7) to relate w_0 to z_0 and have assumed $n = 1$.

Note that z_0 and w_0 are real quantities provided that $0 \leq d/R_2 \leq 1$. Outside this region, (5.2.2) yields nonsensical answers, or, to put it bluntly, it is impossible to choose, in a self-consistent manner, the parameters of a Gaussian beam to match the resonator surfaces.

Within these limits we have achieved a solution and we now know the value of the minimum spot size w_0, (5.2.2), and where it is located; that is at the flat mirror. From this information we can predict what the Gaussian beam parameters are at any point in space. For instance, the spot size at the spherical mirror is found by

$$\frac{\pi w^2(d)}{\lambda_0} = \frac{\pi w_0^2}{\lambda_0}\left[1 + \left(\frac{d}{z_0}\right)^2\right] \tag{5.2.3}$$

Substituting (5.2.2) into (5.2.3) yields the desired information.

$$\begin{aligned}\frac{\pi w^2(d)}{\lambda_0} &= (dR_2)^{1/2}\left(1 - \frac{d}{R_2}\right)^{1/2}\left[1 + \frac{d^2}{dR_2(1 - d/R_2)}\right] \\ &= \frac{(dR_2)^{1/2}}{(1 - d/R_2)^{1/2}}\end{aligned} \tag{5.2.4}$$

We could have started with a more complicated problem with two curved mirrors, as shown in Fig. 5.2. In this problem we must adjust the position of the plane $z = 0$ so that the radius of curvature of the beam going to the right matches the surface of R_2 and simultaneously let this same beam expand to the left to mate with R_1. Here we run into a slight notation problem: The wave front on the left of $z = 0$ has a (mathematical) negative radius curvature, but we know that the mirror R_1 was ordered from the optics company with positive (focusing) properties. Thus we must be sensible and wide awake in our application of the mathematics. We treat all distances z_1, z_2 as positive numbers and let the radii of curvature of the mirrors carry their own sign.

Since we do not know where $z = 0$ is located, we also do not know the z coordinates of the mirrors. But we do know that the sum $z_1 + z_2 = d$. Thus the following equations are pertinent:

$$z_1 + z_2 = d \tag{5.2.5}$$

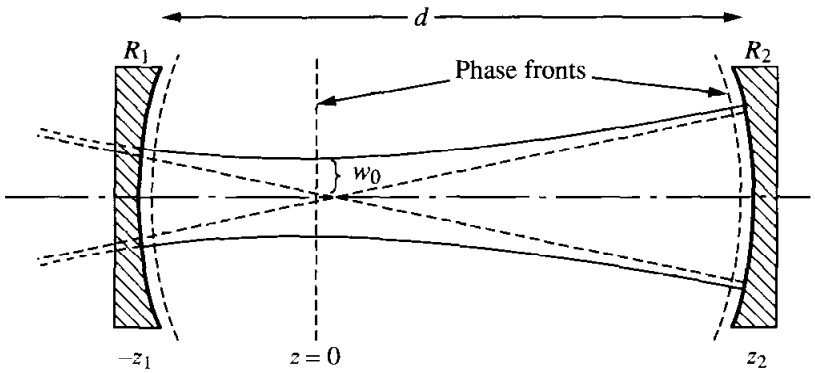


FIGURE 5.2. Simple stable cavity.

$$R(z_2) = R_2 = z_2\left[1 + \left(\frac{z_0}{z_2}\right)^2\right] \tag{5.2.6}$$

$$R(z_1) = -R_1 = -z_1\left[1 + \left(\frac{z_0}{z_1}\right)^2\right] \tag{5.2.7}$$

We now have the unpleasant and uninteresting task of solving three nonlinear equations in three unknowns. The solutions for the desired quantities are as follows:

$$z_0^2 = \left(\frac{\pi w_0^2}{\lambda_0}\right)^2 = \frac{d(R_1 - d)(R_2 - d)(R_1 + R_2 - d)}{(R_1 + R_2 - 2d)^2} \tag{5.2.8}$$

$$\begin{aligned} z_1 &= \frac{d(R_2 - d)}{R_1 + R_2 - 2d} \\ z_2 &= \frac{d(R_1 - d)}{R_1 + R_2 - 2d} \end{aligned} \tag{5.2.9}$$

This procedure of forcing the phase fronts to coincide with the mirror surface is always successful for stable cavities but may become somewhat tedious for complex systems. Fortunately, there is a very straightforward way of handling all stable cavities, complex or simple. This is shown in the next section, where the *ABCD* law of Sec. 3.6 is used.

5.3 APPLICATION OF THE *ABCD* LAW TO CAVITIES

We can find the characteristic modes of an optical cavity by the application of the *ABCD* law, after we have established a definition of a characteristic mode in a cavity.

> A cavity mode is a field distribution that reproduces itself in relative shape and in relative phase after a round trip through the system.

This definition is so basic and so obvious that its importance is sometimes overlooked. It applies equally well to low-frequency resonators, *LC* circuits, microwave cavities, and optical cavities.

In Fig. 5.3(a), a simple short-circuited coaxial cable is used to construct a resonator. The dominant TEM mode in this system has a $1/r$ field, which propagates down to the short, reflects, comes back to the input port, reflects, and starts all over again. Because of the losses in the cable, the amplitude of the field starting the second trip is smaller than the first. But its relative amplitude $(1/r)$ shape is the same, and the relative phase is the same.

The same ideas are there for the $TE_{1,0}$ microwave cavity mode shown in Fig. 5.3(b). If we launch a $TE_{1,0}$ mode down the waveguide, it propagates to the short, back to the other short, and back to the start. Again, the relative shape and phase of the field are the same after a round trip. There is an apparent difference between the optical cavity in Fig. 5.3(c) and that shown in Fig. 5.3(a) or Fig. 5.3(b). The fields in the optical system change in "size" as they traverse the cavity, whereas those in the low-frequency cavities do not. This

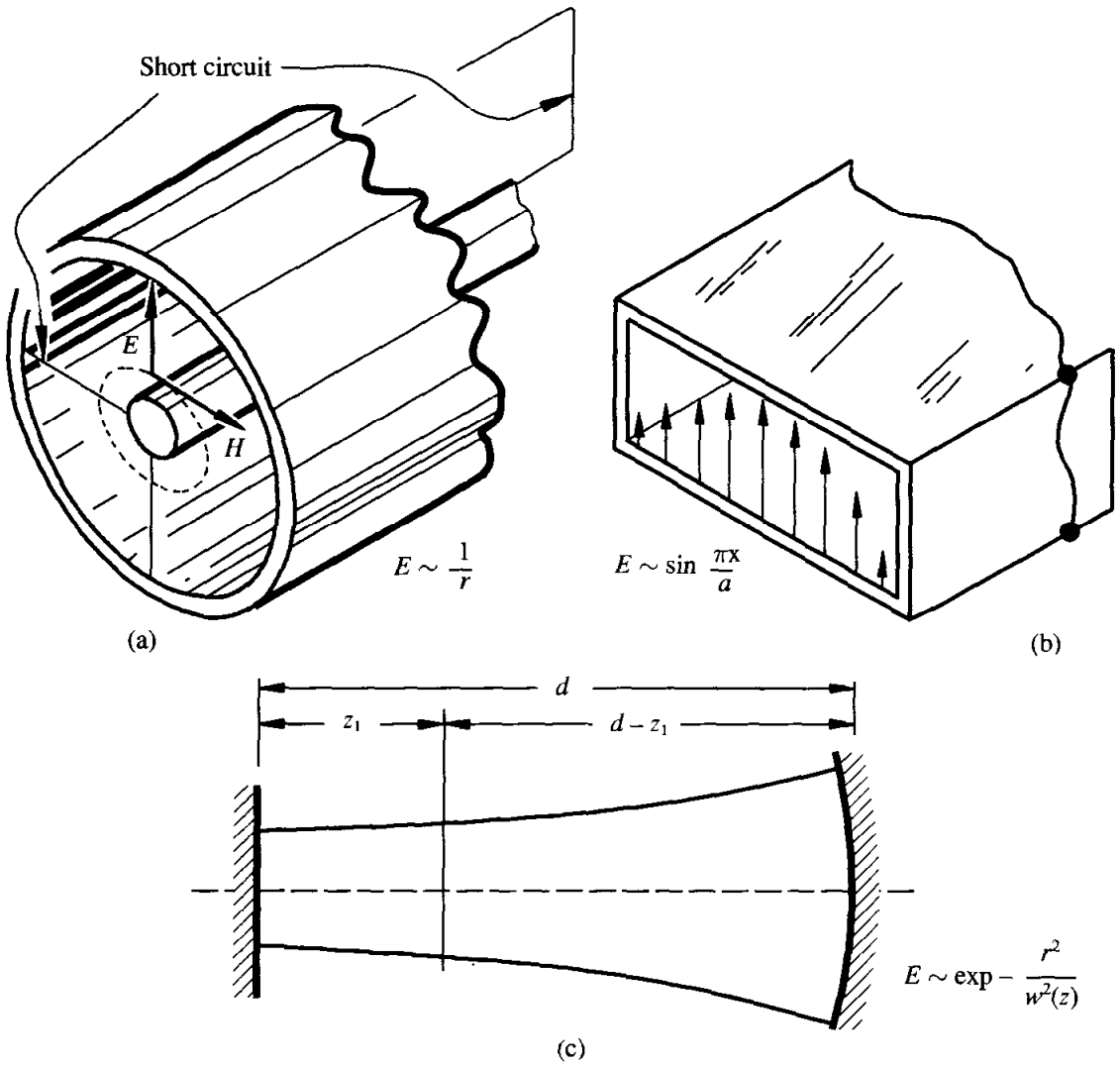


FIGURE 5.3. Typical cavities.

is merely a consequence of the *guided* waves in Figs. 5.3(a) and 5.3(b) as opposed to a freely propagating wave in Fig. 5.3(c). All three fields fit the definition of a cavity mode.

Indeed, in their classic work, Fox and Li [1] utilized this definition to let a computer pick out a normal mode of a cavity. The mathematical detail used by them is more involved than is justified here, and the computer programming is out of place; however, the physics of their reasoning is perfect for our consideration.

They started with an *assumed* field distribution on the mirror surface M_1. That distribution was probably wrong, but, as we shall see, the procedure is self-correcting. They then used Huygen's principle to predict the field distribution at mirror M_2. This involves a messy integral from an analytic standpoint but is straightforward numerically and involves merely computer time to find the field distribution at M_2. Then Huygen's principle is used once again to find the field at M_1, which is most probably different than what was started at the beginning of this paragraph.

Then they began the whole process all over again with the field that survived one complete round trip. After many such round trips through the cavity (and the DO loop), a field distribution is found that doesn't change significantly with additional round trips. At that point, they claim they had achieved a solution for the characteristic mode in the cavity.

Rather than using a computer, we will use the $ABCD$ law to follow a Gaussian beam through a round trip. In particular:

1. We *assume* that the Hermite-Gaussian beams are the characteristic modes of the optical cavity. For this to be true:
2. We require the complex beam parameter to repeat itself after a round trip.

Consider the cavity shown in Fig. 5.3 (analyzed previously in Sec. 5.2). We assume that there is a complex beam parameter $q(z_1)$, which describes the field on an arbitrary plane inside the cavity. We *do not* know its value at this stage, but we do know that it exists. (We hope!) We determine the value of q *at the plane* z_1 by forcing the q to transform into *itself* after a round trip. This will be true if

$$q(z_1 + \text{round trip}) = q(z_1) \tag{5.3.1}$$

Since the left-hand side can be found from the $ABCD$ law, we find that

$$q(z_1) = \frac{Aq(z_1) + B}{Cq(z_1) + D} \tag{5.3.2}$$

where A, B, C, and D are the elements of the transmission matrix for the unit cell chosen for the problem.

We now solve (5.3.2) for q, or rather $1/q(z_1)$:

$$Cq^2 + Dq = Aq + B$$

or

$$B\left(\frac{1}{q}\right)^2 + 2\left(\frac{A-D}{2}\right)\left(\frac{1}{q}\right) - C = 0 \tag{5.3.3}$$

Therefore

$$\frac{1}{q} = -\frac{A-D}{2B} \pm \frac{1}{B}\left[\left(\frac{A-D}{2}\right)^2 + BC\right]^{1/2} \tag{5.3.4}$$

Now we use the fact that $AD - BC = 1$ and manipulate (5.3.4) to conform to the definition of $1/q$.

$$\frac{1}{q(z_1)} = -\frac{A-D}{2B} - j\frac{\left[1 - \left(\frac{A+D}{2}\right)^2\right]^{1/2}}{B} \tag{5.3.5}$$

Thus the radius of curvature of the beam and the spot size at the plane z_1 are given by

$$R(z_1) = -\frac{2B}{A-D} \tag{5.3.6}$$

$$\frac{\pi n w^2(z_1)}{\lambda_0} = \frac{B}{\left[1 - \left(\frac{A+D}{2}\right)^2\right]^{1/2}} \tag{5.3.7}$$

It is most important to recognize that these beam parameters are found *at the plane* z_1, where the unit cell starts (and stops). The unit cell for the case shown in Fig. 5.3(c) is shown in Fig. 5.4. The transmission matrix of the unit cell shown is

$$\mathbf{T} = \begin{bmatrix} 1 & d + z_1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & d - z_1 \\ -\frac{1}{f} & 1 - \frac{d - z_1}{f} \end{bmatrix} \tag{5.3.8}$$

Note that if the unit cell is started at the flat mirror (i.e., $z_1 = 0$), it is symmetric and $A = D$. Thus we have automatically found the position of the beam waist ($z_1 = 0$) where the phase front is planar or $R = \infty$. This, of course, mates to the mirror surface in the same manner as was found in Sec. 5.2. The beam waist is

$$\left.\begin{aligned} \frac{\pi w_0^2}{\lambda_0} &= \frac{d + d(1 - d/f)}{[1 - (1 - d/f)^2]^{1/2}} = \frac{2d(1 - d/R)}{[4(d/R)(1 - d/R)]^{1/2}} \\ \text{or} \quad \frac{\pi w_0^2}{\lambda_0} &= (dR)^{1/2}\left(1 - \frac{d}{R}\right)^{1/2} \end{aligned}\right\} \tag{5.3.9}$$

which is the same as (5.2.2).

It is important to recognize that any and all cavities can be analyzed by the method just outlined. The essential steps in the analysis follow:

1. Postulate that Hermite-Gaussian beams are the normal modes of the cavity.
2. Formulate an equivalent transmission system for this cavity showing *at least* one complete round trip.

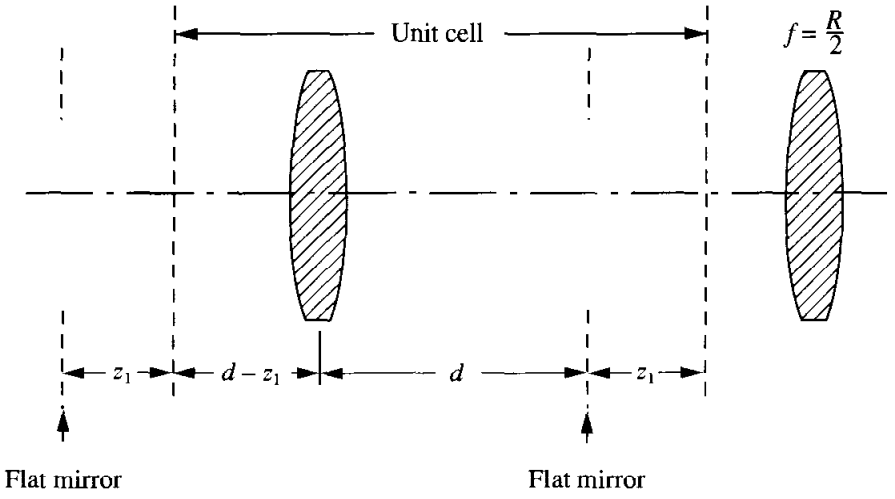


FIGURE 5.4. Equivalent-lens waveguide for the cavity shown in Fig. 5.3(c).

3. Identify a unit cell. Is the cavity stable?
   a. The starting point is arbitrary. However, the parameters determined from (5.3.6) and (5.3.7) correspond to R and w at the corresponding planes in the cavity.
   b. Considerable arithmetic can be avoided by an intelligent choice of a unit cell.
4. Force the complex beam parameter to transform into itself after a round trip by use of the $ABCD$ law.
5. Evaluate R and w from (5.3.6) and (5.3.7).

Although the foregoing steps are logical and straightforward, nonsense can result from a blind application of (5.3.6) and (5.3.7). For instance, if $(A + D)/2$ is greater than 1, these formulas predict an imaginary radius of curvature and an imaginary spot size, which is ridiculous. Thus the theory applies for stable cavities only. No information is obtained for unstable systems.

5.4 MODE VOLUME IN STABLE RESONATORS

Let us return to (5.2.2) and (5.2.4) for the spot sizes on the mirrors of the cavity analyzed in Sec. 5.2.

$$\frac{\pi w_0^2}{\lambda_0} = (dR_2)^{1/2}\left(1 - \frac{d}{R_2}\right)^{1/2} \tag{5.2.2}$$

$$\frac{\pi w^2(d)}{\lambda_0} = \frac{(dR_2)^{1/2}}{(1 - d/R_2)^{1/2}} \tag{5.2.4}$$

The relative variation of these spot sizes as a function of mirror separation is shown in Fig. 5.5. A few minutes of arithmetic in evaluating (5.2.2) or (5.2.4) should convince you that these spot sizes are quite small for visible wavelengths. For instance, let $\lambda = 632.8$ nm, $d = 1$ m, $R_2 = 20$ m, with $R_1 = \infty$. The spot size w_0 is only 0.94 mm (and R_2 is a reasonably "flat" mirror).

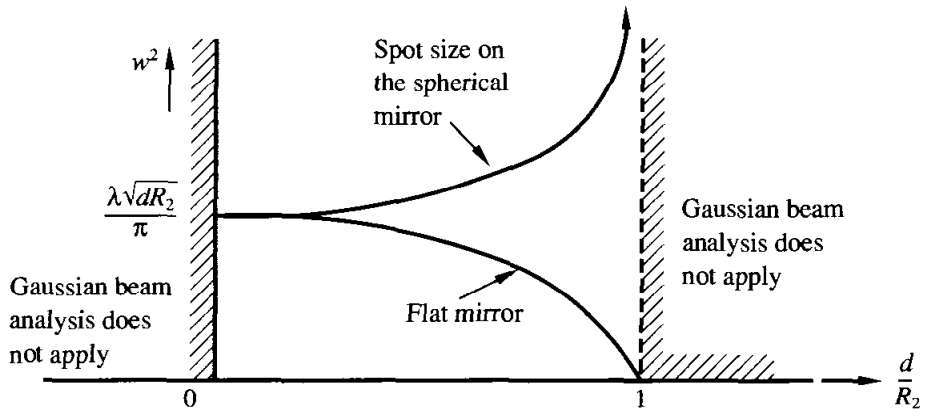


FIGURE 5.5. Spot sizes in a hemispherical cavity.

As we will see later, the active atoms in a laser interact with the square of the electric field; hence, it behooves us to compute the effective mode volume of the Gaussian beam. Knowing this mode volume, we can quickly estimate the *number* of atoms that must be present and radiating to generate a given optical power.

We define the mode volume in terms of the peak electric field in the cavity by

$$E_0^2 V = \int_0^d \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} E(x, y, z) E^*(x, y, z)\, dx\, dy\, dz \tag{5.4.1}$$

where E_0 is the peak value of the electric field within the cavity, obviously occurring at the beam waist and on the axis. The complex-conjugate multiplication eliminates all phase factors, leaving only transverse variation of the field described by the Hermite-Gaussian functions and the z variation of $w(z)$. For a single mode, we have

$$E_0^2 V_{m,p} = E_0^2 \int_0^d \frac{w_0^2}{w^2(z)} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} H_m^2\left(\frac{2^{1/2}x}{w}\right) e^{-2x^2/w^2} \times H_p^2\left(\frac{2^{1/2}y}{w}\right) e^{-2y^2/w^2}\, dy\, dx\, dz \tag{5.4.2}$$

This equation can be rearranged and put in a more conventional form by the substitution

$$u = \frac{2^{1/2}x}{w} \quad \text{or} \quad \frac{2^{1/2}y}{w}$$

$$V_{m,p} = \int_0^d \frac{w_0^2}{2}\, dz \left[\int_{-\infty}^{+\infty} H_m^2(u) e^{-u^2}\, du\right]\left[\int_{-\infty}^{+\infty} H_p^2(u) e^{-u^2}\, du\right] \tag{5.4.3}$$

Now

$$\int_{-\infty}^{+\infty} e^{-u^2} H_m^2(u)\, du = 2^m m! \pi^{1/2}$$

Hence, the mode volume is given by

$$V_{m,p} = \frac{\pi w_0^2}{2} d(m!p!2^{m+p}) \qquad (Note : 0! = 1) \tag{5.4.4}$$

The first factor in (5.4.4) has the satisfying interpretation: $\sim$ area$(\pi w_0^2/2)$ $\times$ length(d), whereas the last is the modification of this basic volume for the higher-order modes.

If we take the example considered previously and restrict our attention to the $\text{TEM}_{0,0}$ mode, the mode volume is quite small:

$$w_0 = 0.94 \text{ mm} \qquad R_2 = 20 \text{ m}$$

$$d = 1 \text{ m} \qquad R_1 = \infty$$

Therefore

$$V_{0,0} = 1.38 \text{ cm}^3$$

With this number we can quickly estimate the number of atoms that can possibly interact with the mode and thus contribute to the laser output power. For instance, suppose that we had a pressure of 0.1 torr for neon for this example, with each atom being excited (on the average) of 10 times per second (by the gas discharge) and thus producing a photon at 632.8 nm. What is the maximum power that we could expect from this laser?

$$\text{Energy per photon} = h\nu = \frac{hc}{\lambda_0} = 3.14 \times 10^{-19}\ \text{J} = 1.96\ \text{eV}$$

$$\times$$

$$\text{Number of neon atoms} = 0.1(3.54 \times 10^{16})V_{0,0} \qquad = 4.88 \times 10^{15}$$

$$\times$$

$$\left\{\begin{array}{l}\text{Average excitation per atom} \\ \| \\ \text{Average emission per atom}\end{array}\right\} \qquad = 10\ \text{sec}^{-1}$$

$$\|$$

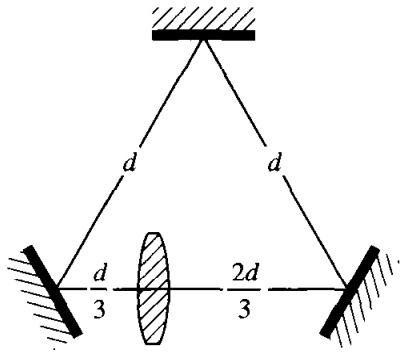
$$\text{Power} \qquad = 15.3\ \text{mW}$$

This is typical for a laser. There are only a couple ways to increase this power. We could increase the length d to make the mode volume large. But there is a practical limit: A 10-m-long laser would be most unwieldy. If we could excite the atoms at a faster rate, the power would be higher. But as we will see later, the 10(second)$^{-1}$ rate is already optimistic. Thus we are left with changing the mode volume.

We could go to the higher-order modes, and for some applications this is a viable method of extracting more power. But as was pointed out in Sec. 3.6, we are still limited by the divergence of the fundamental mode. Unstable resonators have a much larger mode volume and can be used with high gain lasers. These are covered in Chapter 12.

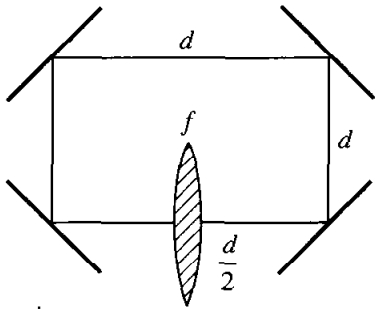
PROBLEMS

5.1. Consider the optical cavity shown in the accompanying diagram. Assume that it is stable.

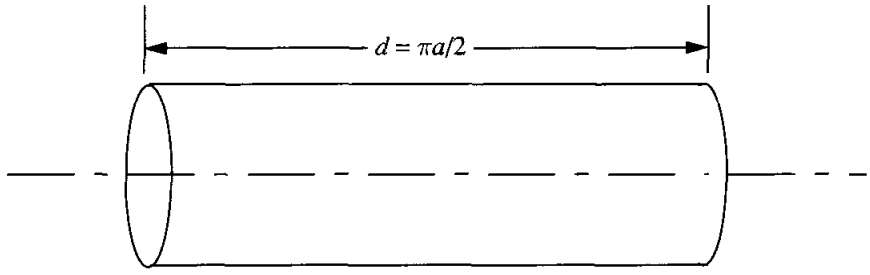


(a) Sketch an equivalent-lens waveguide.
(b) Compute the minimum spot size for a Gaussian beam and identify the place in the cavity where the beam achieves this minimum.
  (1) Identify the appropriate unit cell that makes the $ABCD$ matrix symmetric.
  (2) Identify the plane $z = 0$ in the cavity.
  (3) What is the $ABCD$ matrix for the unit cell?
(c) What is the formula for the minimum spot size?
(d) Show that this formula is valid for stable cavities only.

5.2. Find the spot size and the radius of curvature on the lens for the cavity shown below. The following procedure must be followed:
(a) Show an equivalent-lens waveguide with a unit cell starting just after the lens and proceeding in a counterclockwise fashion.
(b) For what values of d/f is this cavity stable?
(c) If $d = 20$ cm, $f = 40$ cm, and $\lambda_0 = 6000$ Å, find R and w (at the lens).
(d) Identify the plane in the cavity where the spot size is a minimum.



5.3. The GRIN lens shown on the diagram below consists of a *gr*aded *in*dex fiber with $n(r) = n_0(1 - r^2/2a^2)$ of length $d = \pi a/2$. $(a \gg r)$



(a) Find the $ABCD$ matrix of this lens. (CAUTION: Do not ignore the air-dielectric interfaces.)
(b) If we consider this short section of the fiber a cavity, what are the spot sizes at the entrance and exit planes?

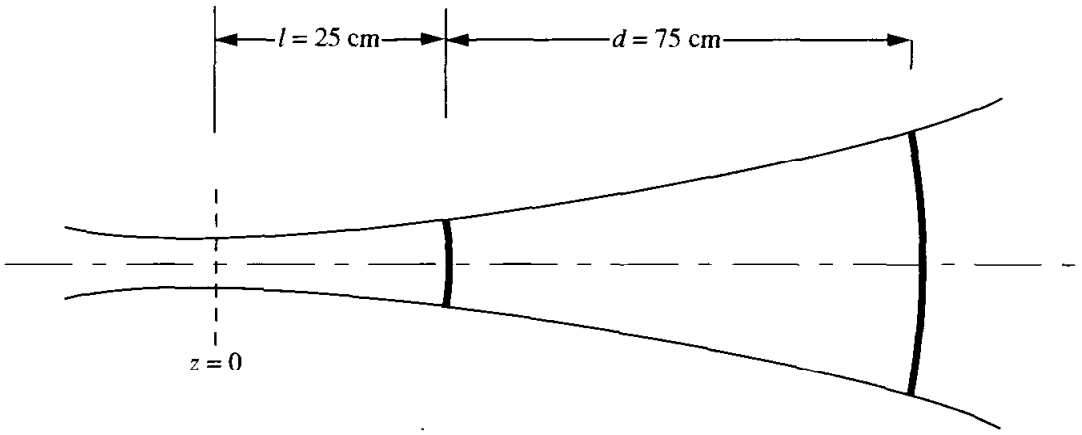
5.4. (a) Construct a graph similar to Fig. 5.5 for the cavity of Fig. 5.2 ($R_1 = R_2$).
(b) Use the expansion law for a Gaussian beam to find the spot sizes on the mirrors.

(c) If $R_1 = R_2$ in Fig. 5.2, find the distance that maximizes the mode volume.

5.5. (a) If $d/2$ of the space adjacent to the flat mirror of Fig. 5.1 were filled with a negative gas lens [see (2.12.12)], show that the cavity is stable for $d = R$.

(b) Find the spot size on the mirrors with $d = R$, and the parameter L of (2.12.12) equal to 100 m.

5.6. In the stable optical cavity shown in the diagram below, the plane $z = 0$ occurs at a distance 25 cm to the left of M_1 with the beam parameter $z_0 = 125$ cm. The distance between the two mirrors is 75 cm.



(a) Find a formula for the resonant frequency of the $\text{TEM}_{m,p,q}$ mode.

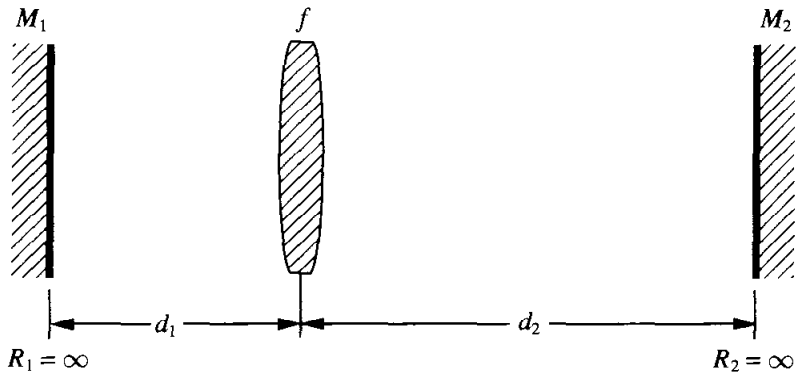
(b) Find the difference between the resonance frequency of the $\text{TEM}_{1,2,q}$ and $\text{TEM}_{0,0,q}$ modes.

(c) Find the radius of curvature for the mirrors M_1 and M_2.

5.7. Consider the optical cavity consisting of two flat mirrors with a converging lens as shown in the accompanying diagram.

(a) What are the stability limits for this cavity? Express your answer in the form of an inequality involving the ratio of d_1/f and d_2/f.

(b) Construct a stability diagram expressing this inequality.



5.8. Find the spot sizes at the mirrors M_1 and M_2 of the cavity shown in Problem 5.7.

5.9. **(a)** Sketch the e^{-1} points of the field for the $TEM_{0,0}$ mode for the cavity of Problem 5.7.

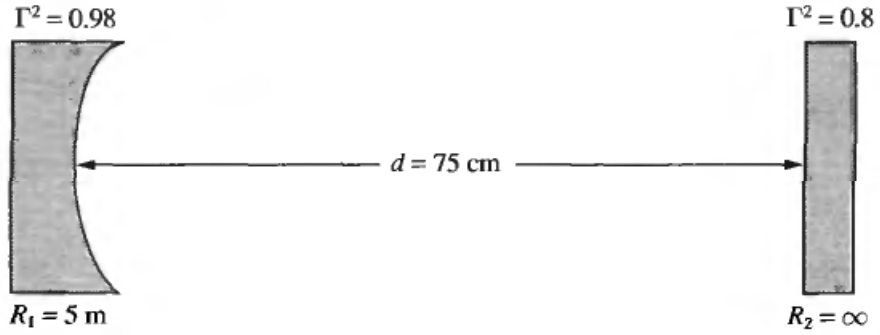
(b) Find the far-field divergence angle of the radiation emerging from the two mirrors. (Label this angle on the sketch.)

5.10. Find a formula for the spot size of a $TEM_{0,0,q}$ mode at the spherical mirror of Fig. 5.1 by following the procedure below.

(a) Show an equivalent-lens waveguide for this cavity and identify a unit cell such that the $ABCD$ law will yield the spot size on the spherical mirror directly.

(b) Use the $ABCD$ law to find a formula for $\pi w_s^2/\lambda_0$.

5.11. A CO_2 laser operating at $\lambda_0 = 10.6\ \mu$m uses the following optical cavity.



(a) Use the $ABCD$ law to find an expression and numerical value for the spot size on the spherical mirror.

(b) Evaluate the photon lifetime and Q of the cavity.

(c) Sketch the intensity (dot) pattern of a $TEM_{2,3}$ mode.

5.12. Suppose we required that the beam parameter repeat after two round trips in the cavity rather than one. What is the formula for complex beam parameter?

5.13. Consider the problem of "mode-matching" the Gaussian beam from a laser to the normal mode of the (semi-)confocal cavity shown below.



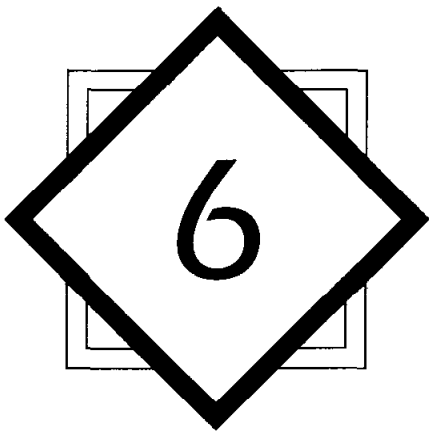
We require that z_{01} be transformed into z_{02} with the proper choice of a lens of focal length f and distances d_1, d_2. What are the design equations for the choice of the lens and distances?

REFERENCES AND SUGGESTED READINGS

1. A. G. Fox and T. Li, "Resonant Modes in a Maser Interferometer," *Bell Syst. Tech. J.* 40, 453-488, Mar. 1961.

2. H. Kogelnik and T. Li, "Laser Beams and Resonators," *Appl. Opt.* 5, 1550-1567, Oct. 1966.

3. A. Yariv, *Introduction to Optical Electronics* (New York: Holt, Rinehart & Winston, 1976), Chap. 4.
4. A. Yariv, *Quantum Electronics*, 2nd ed. (New York: John Wiley & Sons, 1975), Chap. 7.
5. A. E. Siegman, *Lasers* (Mill Valley, Calif.: University Science Books, 1986), Chaps. 16-21.



Resonant Optical Cavities

6.1 GENERAL CAVITY CONCEPTS

Until now we have implied that the cavity is an integral part of any laser because of the feedback that it provides. This is indeed true, but there is more than just raw feedback involved. As we see in this chapter, the cavity provides the ultimate in frequency-determining properties of the laser. It is most important to understand the *classical electromagnetic problem* of a cavity so as to appreciate why a laser always oscillates at a cavity resonance, why the *photon lifetime* is so important to the threshold of oscillation condition, and why the cavity *"filters"* the spontaneous emission from the atoms in the cavity so that stimulated emission takes place to generate a coherent electromagnetic wave.

6.2 RESONANCE

Resonance of an electromagnetic wave at optical frequencies is no different than resonance of any other system, be it mechanical or electrical. There is always an interchange of energy in such a system between potential and kinetic forms in the case of a mechanical system, with attendant friction losses, or between electric and magnetic energy with resistive losses in an electromagnetic problem.

Quite often, the phenomenon of resonance gets lost in the mathematics when analyzing a low-frequency system; fortunately, a much simpler physical picture emerges when we consider systems where the wavelength is much less than the physical dimensions of the components.

To make the problem as familiar as possible, we consider the excitation of the cavity shown in Fig. 6.1 by an external source, such as a tunable laser or a variable-frequency oscillator. To keep the arithmetic to a minimum, we consider all waves, incident on the cavity from the left, inside the cavity, or transmitted through it to the right to be uniform plane waves of limited spatial extent transverse to the direction of propagation. Our task is to relate the fields, running wave intensities, and stored energy on the inside of the cavity to those quantities that we can measure on the outside.

Let us follow a wave as it bounces back and forth between the two mirrors. Consider the initial field at the plane just to the right of M_1, labeled by $\mathbf{E}_0$. It propagates to M_2 and back to the starting plane and experiences an amplitude change of $\Gamma_1 \cdot \Gamma_2$ and a phase factor $\exp[-jk2d]$ as it travels that round trip and thus generates the field labeled $\mathbf{E}_1^+$, which experiences the same trials and tribulations as $\mathbf{E}_0$, and in turn generates $\mathbf{E}_2^+$, and so on. At every point along the path from M_1 to M_2, the fields $\mathbf{E}_1^+$, $\mathbf{E}_2^+$, and so on are to be added to $\mathbf{E}_0$ to which we assign the reference phase of 0°. This phasorial addition is shown in Fig. 6.2 where, because there is an assumed lagging phase angle, we have assumed that the round trip phase shift (RTPS), $2\theta = 2kd$, is almost but not quite an integral multiple of 2π radians. That deficiency is labeled by ϕ and is related to kd by

$$2\theta = 2kd = q2\pi - \phi \tag{6.2.1}$$

where q is now an integer (not the complex beam parameter). Note that if the angle ϕ is significant, the total field propagating to the right inside the cavity is merely the difference between the origin and the spiral of phasors, quite similar to the straight-line distance between the beginning and end of a coiled rope. Not much at all. But, if that rope is uncoiled, the distance becomes much larger.

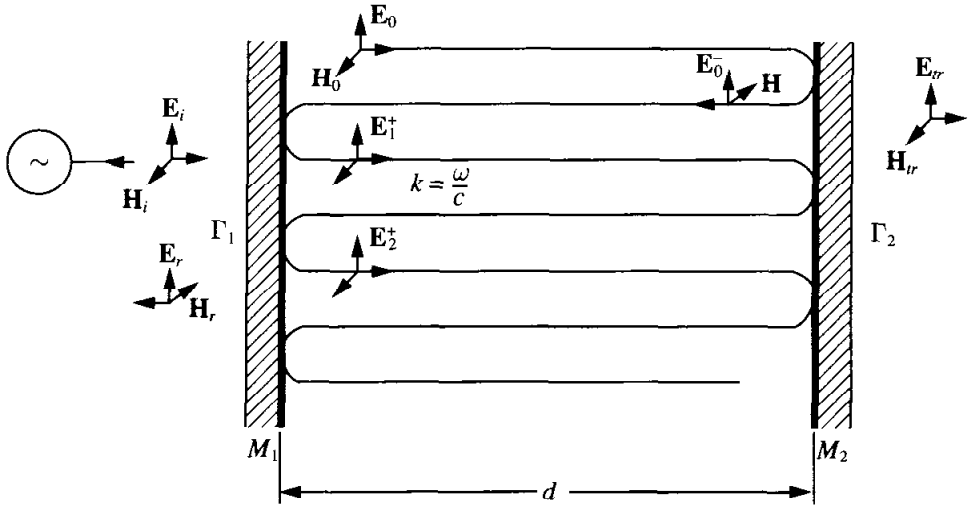


FIGURE 6.1. Optical cavity.

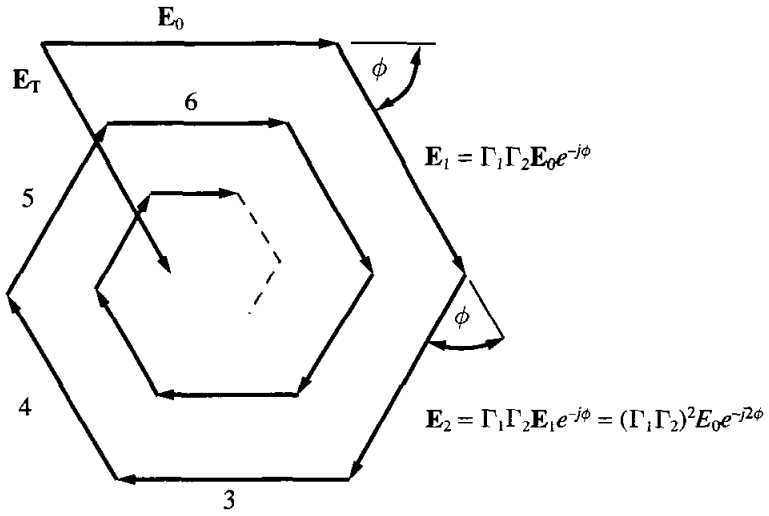


FIGURE 6.2. Phasor diagram.

In a similar fashion, the *total* field $\mathbf{E}_T$ will be many times the initial value $\mathbf{E}_0$ if $\Gamma_{1,2}$ are close to 1 and $\phi = 0$. This point is most important and needs repeating many times over. The following quantities are all maximized by the simple equation $\phi = 0$: the total field traveling to the right (and thus to the left also); the magnetic fields associated with $\mathbf{E}$; the intensities of the waves; the number of photons bouncing back and forth; and the stored energy. These physical facts are characteristic of *resonance,* which is defined by

$$\boxed{\textbf{round trip phase shift} = \textbf{(RTPS)} = \textbf{2kd} = \textbf{q2}\pi} \tag{6.2.2}$$

This simple equation is most important and should be committed to memory for instant recall.

Equation (6.2.2) contains a lot of information that can be discussed in complementary ways. Since $k = \omega n/c = 2\pi/\lambda$, we can use one of the equalities to find the resonant wavelengths

$$k \cdot 2d = \frac{\omega n \cdot 2d}{c} = \frac{2\pi \cdot 2d}{\lambda} = q \cdot 2\pi \tag{6.2.3a}$$

or

$$d = \frac{q \cdot \lambda}{2} \tag{6.2.3b}$$

where $\lambda = \lambda_0/n$. This view of resonance states that there has to be an integral number of half wavelengths between the two mirrors. This implies that the integer q is a very large number for optical frequencies and reasonable size cavities as the following examples illustrate.

Example 6.1 A semiconductor laser cavity

Material	Gallium arsenide (GaAs)
Index of refraction	3.6
Length of cavity d	100 μm = 0.01 cm= 3.94×10^{-4} in.

Wavelength region of interest	8000 Å

Thus

$$q = \frac{nd}{(\lambda_0/2)} = \frac{3.6 \cdot 10^{-2}\ \text{cm}}{0.4 \cdot 10^{-4}\ \text{cm}} = 900$$

We can turn this problem around and ask, What is the wavelength for $q = 899$ and 901?

$$\frac{\lambda_0}{2} = \frac{nd}{899} \quad \text{or} \quad \lambda_0 = 8008.898\ \text{Å for } q = 899$$

$$\frac{\lambda_0}{2} = \frac{nd}{901} \quad \text{or} \quad \lambda_0 = 7991.121\ \text{Å for } q = 901$$

Example 6.2 A gas laser

Material	Helium-neon gas at 1 torr
Index of refraction	1.000+
Length of cavity	20 cm=8 in.
Wavelength region of interest	6328 Å

Thus $q = 632, 111$ for $\lambda_0 = 6328.0025$ Å

and $q = 632, 110$ for $\lambda_0 = 6328.0125$ Å

Note that there is only 0.010 Å difference between two adjacent wavelengths for the large gas laser, a difference beyond the resolution capabilities of most monochromators. Obviously, it was somewhat silly to carry out the calculation of wavelength to eight significant digits when the distance is only given to two. However, the point is that the wavelengths are very close together even for the small semiconductor and extremely so for a large gas laser.

Equation (6.2.2) can also be interpreted in terms of frequency ν.

$$k \cdot 2d = \omega \cdot \frac{2nd}{c} = 2\pi\nu \cdot \frac{2nd}{c} = q(2\pi)$$
$$\nu = q \cdot \frac{c}{2nd} \tag{6.2.4}$$

Because q is restricted to integer values, there are only discrete frequencies which obey the resonance condition. The separation between those frequencies is given by

$$\nu_{q+1} - \nu_q = \frac{c}{2nd} \tag{6.2.5}$$

For reasonable-size cavities, this separation yields numbers that should be comfortable to most. For the helium-neon laser cavity of example 2, this difference is 750 MHz. Although the frequency difference is important (it is called the free spectral range), do not lose sight of the resonant frequency given by $\nu = q(c/2d) = 473.7553$ THz for $\lambda_0 = 6328.002$ Å, which is very high compared to those "comfortable" values.

Stimulated emission, which is the key issue with any laser, is always proportional to the energy (E^2); hence that stimulation will always be a maximum at a cavity resonance.

6.3 SHARPNESS OF RESONANCE: Q AND FINESSE

It should be obvious from Fig. 6.2 that $\phi = 0$ (or RTPS $= q \cdot 2\pi$) corresponds to the maximum internal field, but there would be a significant amplitude for a "small" deviation from the exact resonance. The issues to be addressed here are (1) how small, (2) what is the ratio of the fields at resonance to that at anti-resonance, and (3) what is the frequency selectivity of the cavity.

There are three interrelated characteristic parameters associated with a cavity that describe the resonance phenomenon: Q (quality), F (finesse), and τ_p (photon lifetime). To derive an explicit relationship between the resonance, these quantities, and the construction of the cavity, we need an analytic description of the fields inside the cavity and their relationship to those exciting the cavity.

The total electric field on the right side of M_1 and traveling to the right (indicated by the superscript "+") is given by the field E_0, which is that transmitted through the mirror from the source, plus the fields that have made (1 to N) round trips to M_2, back to M_1, and starting the (2 to $N+1$) trip. The amplitudes of those fields are simply related to E_0 by the field reflection coefficient of each mirror. The phase of the Nth component, E_N, is delayed with respect to E_0 by N times the round trip phase shift of $2kd = 2\theta$ and N times the phase contributed by each mirror.* While this last contribution can be important, let us ignore it for now in the interest of simplicity.

$$E_T^+ = \Sigma E_N^+ = \underbrace{\mathbf{E}_0\{1}_{\text{incident}} + \underbrace{\Gamma_1\Gamma_2 e^{-jk\cdot 2d}}_{\text{one}} + \underbrace{(\Gamma_1\Gamma_2 e^{-jk\cdot 2d})^2}_{\text{two}} + \cdots \quad \text{round trips}$$

$$= \mathbf{E}_0\left[\frac{1}{1-\Gamma_1\Gamma_2 e^{-j2\theta}}\right] \tag{6.3.1}$$

where θ is the electrical (or optical) length of the cavity, and is equal to $\omega nd/c$.

The total field returning from M_2, traveling to the left (indicated by the superscript "−") and incident on M_1 is just Γ_2 times E_T^+ times the round-trip phase factor.

$$E_T^- = \Gamma_2 e^{-j2\theta} E_T^+ = \mathbf{E}_0\left[\frac{\Gamma_2 e^{-j2\theta}}{1-\Gamma_1\Gamma_2 e^{-j2\theta}}\right] \tag{6.3.2}$$

Both (6.3.1) and (6.3.2) state (in mathematical terms) the conclusions of Fig. 6.2; namely, that the fields are a maximum when the denominators are a minimum (when $2\theta = q \cdot 2\pi$).

The running wave intensities I^+ and I^- are simply related to $E^{(+,-)}$ by $I = E \cdot E^*/2\eta$ where the asterisk denotes complex conjugation. For the wave running to the right, we obtain

$$I^+(z=0^+) = \frac{|\mathbf{E}_0|^2}{2\eta}\left\{\frac{1}{1-\Gamma_1\Gamma_2 e^{-j2\theta} - (\Gamma_1\Gamma_2 e^{-j2\theta})^* + |\Gamma_1\Gamma_2|^2}\right\}$$

$$= I_0 \frac{1}{1-2|\Gamma_1\Gamma_2|\cos 2\theta + |\Gamma_1\Gamma_2|^2}$$

$$I^+(z=0^+) = I_0 \frac{1}{1-2|\Gamma_1\Gamma_2|[1-2\sin^2\theta] + |\Gamma_1\Gamma_2|^2}$$

*The phase of the reflection coefficient is usually a slow function of frequency, and we neglect it.

$$I^+(z = 0^+) = \frac{E_0^2}{2\eta}\left\{\frac{1}{(1 - \sqrt{R_1R_2})^2 + 4\sqrt{R_1R_2}\sin^2\theta}\right\} \tag{6.3.3}$$

where we have assumed that the field reflection coefficients are real numbers and have substituted the power reflection coefficients $R_{1,2} = |\Gamma_{1,2}|^2$. The plane $z = 0^+$ is just to the right of the surface of M_1.

The quantity $E_0^2/2\eta$ is an intensity and is simply the (power) transmission coefficient of M_1 times the incident intensity $T_1 \cdot [E_{\text{inc}}^2/2\eta]$, where $T_1 = 1 - R_1$ for lossless mirrors. In a similar fashion, the intensity transmitted through M_2 is the power transmission coefficient $T_2 = 1 - R_2$ times the intensity given by (6.3.3) since any thing that starts to the right from $z = 0^+$ impinges on M_2 at $z = d$. After converting to reflectivities and incident intensity, we obtain an expression for the intensity or power transmission coefficient through the two mirrors.

$$I_t = \left\{\frac{E_0^2}{2\eta} = T_1 I_{\text{inc}} = (1 - R_1)I_{\text{inc}}\right\}\left\{\frac{T_2 = (1 - R_2)}{(1 - \sqrt{R_1R_2})^2 + 4\sqrt{R_1R_2}\sin^2\theta}\right\}$$

or

$$\boxed{T(\theta) = \frac{I_{\text{trans}}}{I_{\text{inc}}} = \left\{\frac{(1 - R_1)(1 - R_2)}{(1 - \sqrt{R_1R_2})^2 + 4\sqrt{R_1R_2}\sin^2\theta}\right\}} \tag{6.3.4}$$

The net transmission is obviously a maximum when the denominator is a minimum, but simple examples bring home the consequences more clearly.

Example 1

Assume identical mirrors $R_1 = R_2$ and let $\sqrt{R_1R_2} = R$. At resonance $\sin^2\theta = 0$ and

$$T(q\pi) = [(1 - R)^2/(1 - R)^2] = 1 \text{ (one) independent of } R.$$

Think about that for a moment. If each mirror's reflectivity were 99%, then the wave encounters two such obstacles each with a transmission coefficient of 1%. Our first inclination would be to suggest that the overall transmission from left of M_1 to the right of M_2 would be 10^{-4}. That is wrong because it ignores the coherence properties of the wave as was used in the derivation of (6.3.4), illustrating that we must first add the fields and then square to obtain the intensity.

There is one more startling fact that is contained in the statement that $T(q\pi) = 1$ independent of the mirror reflectivities. Since the transmission coefficient of M_2 is $(1 - R_2)$, the value of I^+ must be $1/(1 - R_2)$ times the *incident* value so that the product would be I_{inc}. For the case of the reflectivity being 99%, the intensity incident upon M_2 must be 100 times larger than the intensity used to excite the cavity.

> The running wave intensities on the inside of the cavity can be much larger than those on the outside, including that used to excite the cavity.

We have not violated conservation of energy. Rather, this intensity enhancement is a manifestation of the energy storage capabilities of a cavity with highly reflecting mirrors. At antiresonance, $\theta = (q + 1/2)\pi$, and the transmission coefficient is very small.

$$T[(\theta = (q + 1/2)\pi] = \frac{(1 - R)^2}{(1 + R)^2} = 2.53 \times 10^{-5}$$

for $R = 0.99$. Thus the cavity expresses a loud and significant preference for those frequencies obeying the resonance condition.

A plot of the transmission coefficient given by (6.3.4) is shown in Fig. 6.3 where the horizontal axis is $[\theta/\pi - q]$. Thus, it can be changed by varying the frequency $\omega = 2\pi\nu$, the wavelength $\omega/c = 2\pi/\lambda_0$, a distance d, or an index n. While the vertical axis is plotted as a transmission coefficient, it can also serve as a measure of the relative fields, energy, or running intensities inside the cavity since T_2 is merely a sampling device for those quantities.

The quality factor (Q) of the cavity is a measure of the sharpness or selectivity of the resonance. If ν_0 is the frequency of one of the peaks, then Q is given by

$$\boxed{Q = \frac{\nu_0}{\Delta\nu_{1/2}} = \frac{\omega_0}{\Delta\omega_{1/2}} = \frac{\lambda_0}{\Delta\lambda_{1/2}}} \tag{6.3.5}$$

where $\Delta\nu_{1/2}$ (or $\Delta\lambda_{1/2}$) is the full width at half of the maximum (FWHM), and the horizontal axis is treated as a frequency axis for the frequency variables or wavelength (increasing in the opposite direction) for a λ variation.

It is a rather tiring exercise in trigonometry of small angles to solve for the frequencies $\nu_{+,-}$ which make the response decrease to 50% of the maximum to find an analytic expression for Q. For reasonable values of the product of the reflectivities, we can expand $\sin\theta$ around the peak:

$$\sin\theta_{+,-} = \sin\left[\frac{(\omega_{+,-})nd}{c}\right] = \pm\frac{1 - (R_1R_2)^{1/2}}{(R_1R_2)^{1/4}} \tag{6.3.6a}$$

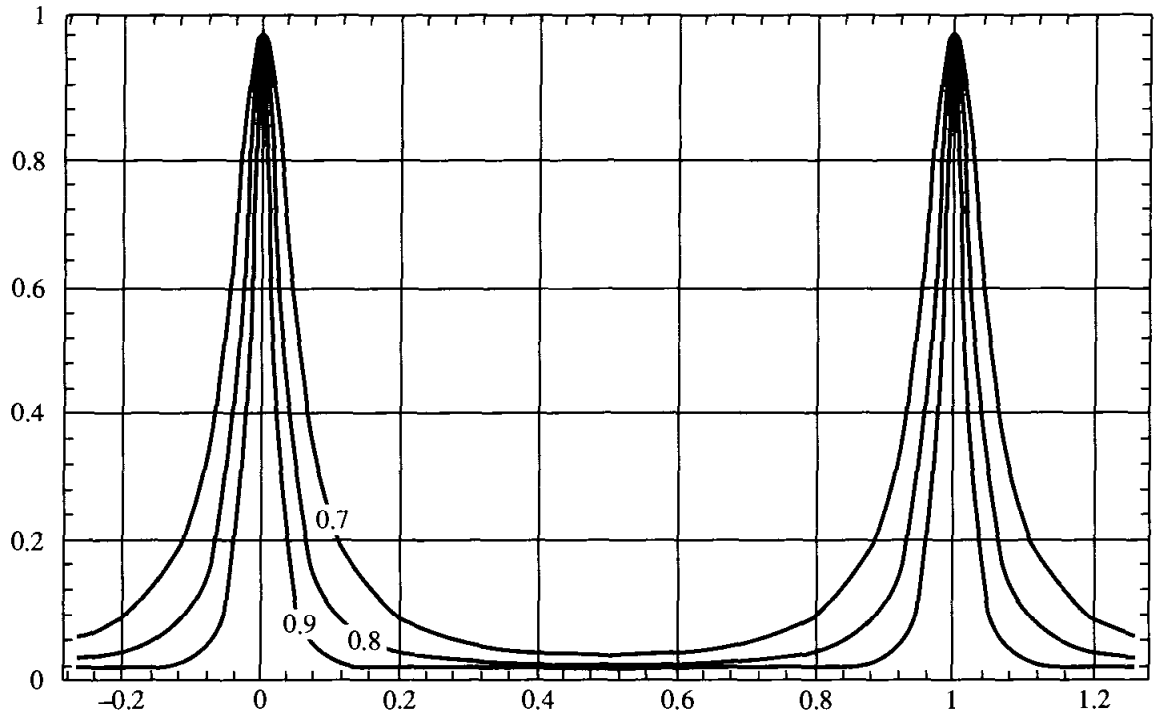


FIGURE 6.3. The transmission through a Fabry-Perot cavity as a function of the electrical length measured in units of $\theta/\pi - q$. The three curves were plotted for $R_1 = R_2 = 0.9$, 0.8, and 0.7.

and thus

$$\Delta\nu_{1/2} = \nu_+ - \nu_- = \frac{c}{2nd}\left\{\frac{1-(R_1R_2)^{1/2}}{\pi(R_1R_2)^{1/4}}\right\} \tag{6.3.6b}$$

Thus the cavity Q is given by $q(c/2nd)$ divided by $\Delta\nu_{1/2}$. We should recognize that q is merely the number of half wavelengths between the two mirrors; thus

$$Q = \frac{q(c/2nd)}{\Delta\nu_{1/2}} = \frac{2\pi nd}{\lambda_0}\frac{(R_1R_2)^{1/4}}{1-(R_1R_2)^{1/2}} \tag{6.3.7}$$

since

$$q = \frac{nd}{\lambda_0/2}$$

Those who were introduced to Q at radio and/or microwave frequencies should be comfortable with the concept of Q as a measure of the sharpness of the resonance. Unfortunately, the numerical values of Q at optical frequencies are astronomical, primarily because of the smallness of λ_0. For instance, suppose that $d = 1$ m, R_1, $R_2 = 0.99$, and $\lambda = 632.8$ nm; then $Q = 9.88 \times 10^8$. But, of course, the resonant frequency is also astronomical: $\nu = c/\lambda = 4.74 \times 10^{14}$. Only the half-width $\Delta\nu_{1/2}$ is a familiar quantity, 480 kHz.

To avoid such large numbers and yet provide a measure of the filtering properties of the cavity, one uses another term, the Finesse (F).*

$$\left.\begin{aligned} F &= \frac{\text{free spectral range}}{\text{full width at half maximum}} = \frac{c/2nd}{\Delta\nu_{1/2}} \\ \text{or} \qquad F &= \frac{\pi(R_1R_2)^{1/4}}{1-(R_1R_2)^{1/2}} \end{aligned}\right\} \tag{6.3.8}$$

For the numbers just used, this yields a more reasonable value of $F = 313$.

6.4 PHOTON LIFETIME

Closely related to the quality factor Q or the finesse F of a cavity is its photon lifetime. It is a time constant describing the build up or the decay of energy in a cavity and is one of the most useful (but simple) parameters describing a cavity. The steps used to derive an explicit formula are simple but very important because similar steps will be used at strategic points through the remainder of the text. The derivation will be the first instance of describing the physical phenomena by a *rate equation* (i.e., a differential equation in which the independent variable is time).

Consider Fig. 6.4, which depicts a simple cavity with a package of photons bouncing back and forth between the mirrors. Assume that at $t = 0$, there are N_p photons in this

*Historically, the use of the term came from those using optical instruments, whereas the Q concept came from the lower-frequency domain. At optical frequencies, the mirror system is referred to as a Fabry-Perot cavity. It is quite often used as a bandpass filter to examine very narrow portions of a spectrum.

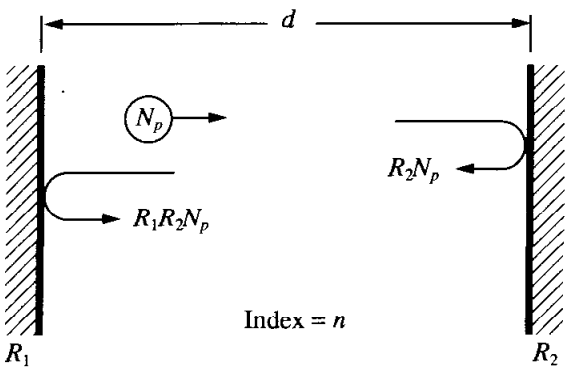


FIGURE 6.4. Decay of photons in a cavity.

package. Hence, the energy in the cavity is $h\nu N_p$. The number of photons surviving one (1) round trip is $[S = R_1R_2]$ times the initial number of photons and thus the number lost is

$$\text{number of photons lost in one round trip} = [1 - S]N_p \tag{6.4.1}$$

where S = Survival factor for the cavity [$= R_1R_2$ for the simple cavity of Fig. 6.4]. The time rate of change of photons (or energy) in the cavity is given by the negative of the number lost divided by the time for the round trip (τ_{RT}).

$$\frac{\Delta N_p}{\Delta t} \to \frac{dN_p}{dt} = \frac{N_p(t + \tau_{RT}) - N_p(t)}{\tau_{RT}} = -\frac{[1 - S]N_p}{\tau_{RT}}$$

or

$$\boxed{\frac{dN_p}{dt} \triangleq -\frac{N_p}{\tau_p}} \tag{6.4.2a}$$

where

$$\boxed{\tau_p = \frac{\tau_{RT}}{1 - S} = \text{photon lifetime}} \tag{6.4.2b}$$

Equation (6.4.2a) has a simple solution.

$$N_p(t) = N_p(0)\exp[-t/\tau_p]$$

Thus it takes on the order of $(1 - S)^{-1}$ round trips for the stored photon energy to decrease to 36.8% ($= e^{-1}$) of its initial value. For the simple cavity of Fig. 6.4, $\tau_{RT} = 2nd/c$ and S is the product of the reflectivities. If the index were lossy so that the transmission through the length $2d$ was $\exp[-\alpha 2d]$, then $S = R_1R_2\exp[-\alpha 2d]$. In other words, the survival factor includes any process that decreases the number of photons (or decreases the energy) in the cavity.

The concept of the photon lifetime is most important from a theoretical standpoint and is usually the simplest quantity to compute. It can be related to Q by using the theoretical definition* of Q.

$$Q = \frac{2\pi(\text{energy stored in the system at resonance} = W)}{(\text{energy lost in a cycle of oscillation})} \tag{6.4.3a}$$

*(The previous definition of $Q = \nu/\Delta\nu_{1/2}$ is most likely the one to be used by experimentalists using frequency domain data, whereas (6.4.3) is useful for theoretical calculations in the time domain.)

The energy lost in a cycle is the average power lost in one cycle times the period of oscillation:

$$Q = \frac{2\pi}{T} \cdot \frac{W}{\langle P \rangle} = \omega_0 \frac{W}{\langle P \rangle} \tag{6.4.3b}$$

The average power is equal to the (decreasing) time rate of change in the stored energy W.

$$\boxed{Q = \omega_0 \frac{W}{-dW/dt} \quad \text{or} \quad \frac{dW}{dt} = -\frac{\omega_0}{Q} \cdot W} \tag{6.4.3c}$$

If we rewrite W in (6.4.3c) in terms of the number of photons, $W = h\nu N_p$, then we have the promised connection between τ_p and Q.

$$\frac{dW}{dt} = -\frac{\omega_0}{Q} W \Rightarrow \frac{d}{dt}[h\nu N_p] = \frac{\omega_0}{Q}[h\nu N_p] = -\frac{[h\nu N_p]}{\tau_p}$$

Thus

$$\boxed{\tau_p = \frac{Q}{\omega_0}} \tag{6.4.4}$$

Since $\Delta\omega_{1/2} = \omega_0/Q$, we have an "uncertainty" relationship.

$$\Delta\omega_{1/2} = \frac{1}{\tau_p} \quad \text{or} \quad \boxed{\Delta\omega_{1/2}\tau_p = 1} \tag{6.4.5}$$

which should surprise no one. Once the FWHM of the resonance $\Delta\nu_{1/2} = \Delta\omega_{1/2}/2\pi$ is known, the computation of Q or the finesse F is straight forward. For the simple cavity of Fig. 6.4 we have

$$\tau_p = \frac{\tau_{RT}}{1 - S} = \frac{2nd/c}{1 - R_1R_2} \quad \text{by (6.4.2b)}$$

$$\therefore \Delta\nu_{1/2} = \frac{1}{2\pi\tau_p} = \frac{c(1 - R_1R_2)}{(2nd) \cdot 2\pi} \quad \text{by (6.4.5)}$$

$$Q = \frac{\nu_0}{\Delta\nu_{1/2}} = 2\pi\nu_0\left(\frac{2nd}{c}\right)\frac{1}{1 - R_1R_2} = \frac{4\pi nd}{\lambda_0}\left(\frac{1}{1 - R_1R_2}\right) \quad \text{by (6.3.5)}$$

$$F = \frac{FSR}{\Delta\nu_{1/2}} = \frac{c/2nd}{\Delta\nu_{1/2}} = \frac{2\pi}{1 - R_1R_2} \quad \text{by (6.3.8)}$$

These formulas are slightly different from those found at the end of Sec. 6.3, but the difference is negligible for high quality cavities. The important issues are (1) the resonant enhancement of the internal energy inside the cavity, (2) the spectral separation and selectivity of the resonances, and (3) the integration or summing time constant τ_p associated with each resonance.

While a passive system driven by an external has been assumed, the analysis is easily adapted to an (active) medium with gain on the inside or between the mirrors. For such a case, as we will see in the next chapter, these resonances can become the oscillating frequencies

of the laser. For a laser to oscillate, it must start from noise (as must all oscillators) and thus an active medium with feedback always starts with τ_p as a negative number, which indicates that the photon number is growing with time. When the oscillator reaches a steady state or a CW value, the composite photon lifetime must be infinite so that $dN_p/dt = 0$. Thus the net round trip gain must change from being bigger than 1 to becoming equal to 1 in the CW limit. The dynamics of this change will be addressed in Sec. 8.3 for a laser, a far simpler task than attempting to do so for a transistor.

6.5 RESONANCE OF THE HERMITE-GAUSSIAN MODES

In the analysis to this point in the chapter we have assumed a limited-extent uniform plane wave for the fields, whereas the more exact description leads to the Hermite-Gaussian modes. The concepts introduced previously are still applicable, but there are refinements.

Resonance is still determined by (6.2.1), and the round-trip phase must equal 2π radians. However, the phase constant for the Hermite-Gaussian modes is not $\omega n/c$ but is much more complicated. The phase shift experienced by a $\text{TEM}_{m,p}$ mode in propagating from $z = 0$ to a point $z = d$ is given by

$$\phi(d) - \phi(0) = kd - (1 + m + p)\tan^{-1}\left(\frac{d}{z_0}\right) \tag{3.5.1}$$

Thus the equation for resonance for the mirror system shown in Fig. 6.5 is as follows:

$$kd - (1 + m + p)\tan^{-1}\left(\frac{d}{z_0}\right) = q\pi \tag{6.5.1}$$

If we recall (5.2.2), which relates z_0 to the radius of curvature, R_2,

$$z_0 = (dR_2)^{1/2}\left(1 - \frac{d}{R_2}\right)^{1/2} \tag{5.2.2}$$

and perform some painful, yet trivial manipulations on (6.5.1), we obtain a more easily interpreted formula for the resonant frequencies of the $\text{TEM}_{m,p,q}$ modes.

$$\nu_{m,n,q} = \frac{c}{2nd}\left\{q + \frac{1 + m + p}{\pi}\tan^{-1}\left[\frac{(d/R_2)^{1/2}}{(1 - d/R_2)^{1/2}}\right]\right\} \tag{6.5.2}$$

or

$$\nu_{m,n,q} = \frac{c}{2nd}\left[q + \frac{1 + m + p}{\pi}\cos^{-1}\left(1 - \frac{d}{R_2}\right)^{1/2}\right] \tag{6.5.3}$$

If M_1 has a finite radius of curvature R_1, the arithmetic is most painful, but the answer is quite similar to (6.5.3):

$$\nu_{m,n,q} = \frac{c}{2nd}\left[q + \frac{1 + m + p}{\pi}\cos^{-1}(g_1 g_2)^{1/2}\right] \tag{6.5.4}$$

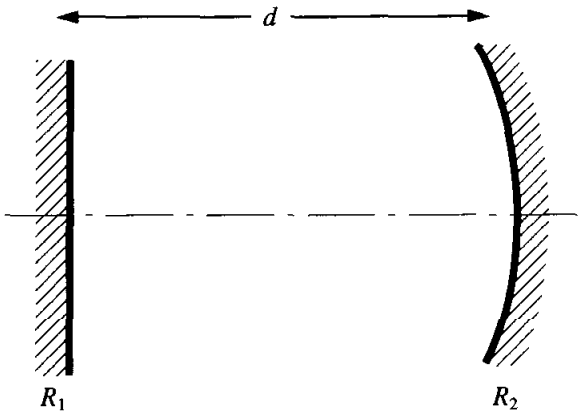


FIGURE 6.5. Optical cavity for the Hermite-Gaussian beam modes.

where

$$g_{1,2} = 1 - \frac{d}{R_{1,2}}$$

Note that the frequency separation between modes with the same transverse numbers is still $c/2nd$ as in the previous section, but there is a difference between modes with different m and p values.

For example, let $d/R_2 = \frac{1}{2}$ in Fig. 6.5; hence, the arc cosine term in (6.5.3) is $\pi/4$.

$$\nu_{m,n,q} = \frac{c}{2nd}\left(q + \frac{1+m+p}{4}\right) \tag{6.5.5}$$

Thus, as m or p increases by 1, the frequency changes by one fourth of the fundamental spacing $c/2nd$. Furthermore, the (0,0) mode is shifted from that predicted by the plane-wave analysis by this same value. These effects are shown in Fig. 6.6.

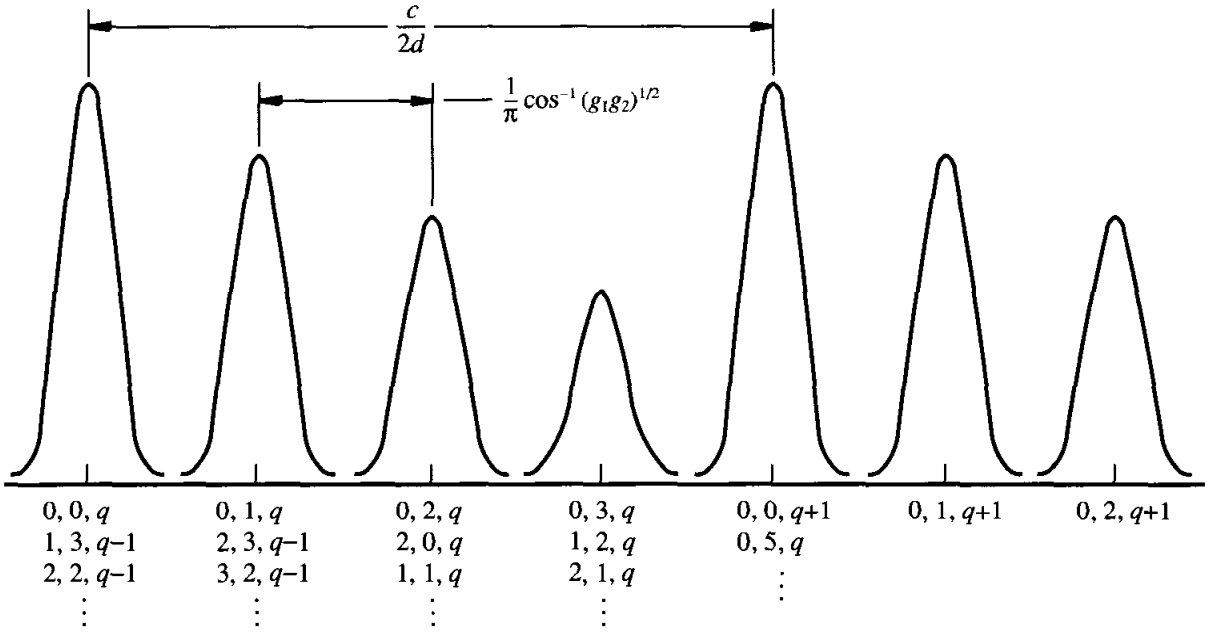


FIGURE 6.6. Frequency degeneracy in an optical cavity.

Also shown in Fig. 6.6 is the fact that two modes, with different values of m, p, and q, can have the same resonant frequency. This is an example of frequency degeneracy.

6.6 DIFFRACTION LOSSES

The Hermite-Gaussian beam modes are the characteristic values for stable infinite aperture mirrors, whereas any practical system has a finite size. Furthermore, the transverse dimensions of the active median of the laser may be considerably smaller than that of the mirrors. Indeed, this is the most logical case.

In either event, the beam modes are only approximate solutions for the cavity and are, fortunately, excellent ones for most laser cavities. But we do have to account for the fact that the modes extend to a considerable distance from the axis and may *not* hit a reflecting surface. The fraction of the incident power that is not intercepted by the mirrors is called diffraction losses.

An exact field description is quite complicated and is beyond the scope of the treatment here. However, we can generate an approximate technique by using some elementary physical reasoning, as is shown in Fig. 6.7.

Here we show a Hermite-Gaussian beam impinging on an iris, an absorbing plate with a hole in it. (Quite often, such a device is used to restrict oscillation to the dominant $TEM_{0,0}$ mode in a cavity.) If we let our imagination loose, we would estimate that this aperture would transmit only the amount of power contained in the cross-sectional area of

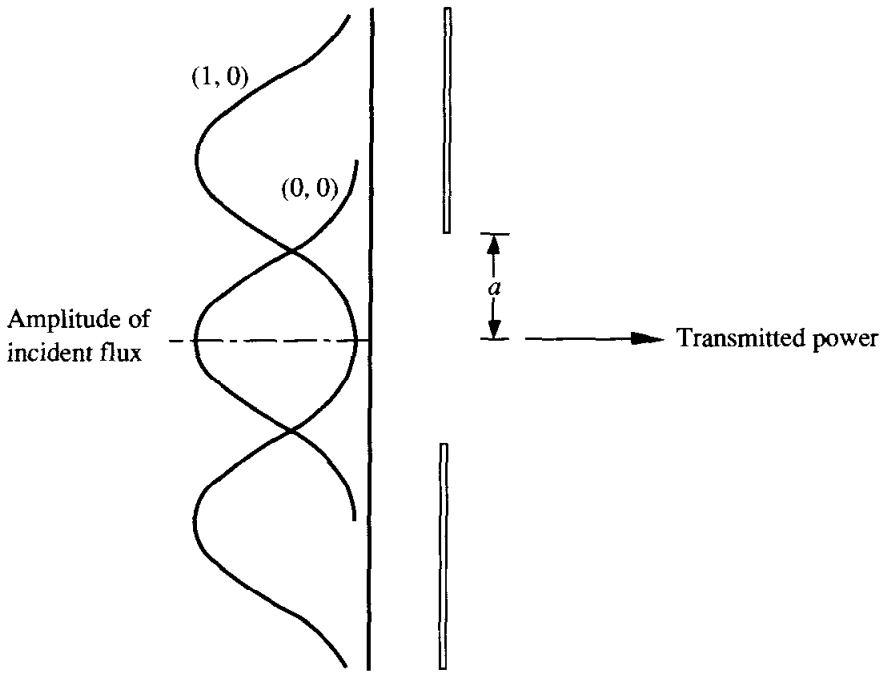


FIGURE 6.7. Transmission through an aperture.

the hole. Thus the transmission coefficient is given by

$$T = \frac{\int_{r=0}^{a}\int_{0}^{2\pi} |E(r,\phi,z)|^2 r\, dr\, d\phi}{\int_{0}^{\infty}\int_{0}^{2\pi} |E(r,\phi,z)|^2 r\, dr\, d\phi} \tag{6.6.1}$$

For the two modes illustrated in Fig. 6.7, we obtain:

$$T_{(0,0)} = 1 - \exp\left[-2\left(\frac{a}{w}\right)^2\right] \tag{6.6.2a}$$

$$T_{(1,0)} = 1 - \left[1 + 2\left(\frac{a}{w}\right)^2\right]\exp\left[-2\left(\frac{a}{w}\right)^2\right] \tag{6.6.2b}$$

By inspection of (6.6.2a) and (6.6.2b), we can see that this simple iris transmits more of the (0,0) mode than of the (1,0) mode.

We should not attempt to carry this approximate analysis too far, for it is only a zeroth-order approximation. Only when the loss is low (or the reflectivity high) does this physical approach lead to the precise numerical value. However, it is close and provides guidelines.

6.7 CAVITY WITH GAIN: AN EXAMPLE

Let us redo the material of Sec. 6.3 but now include a medium with a power gain G_0 between the two mirrors. (Thus the field is amplified by $G_0^{1/2}$.) The serious student should repeat each step of the development from (6.3.1) to (6.3.8), which amounts to replacing the round trip propagator $1 \cdot \exp[-jk \cdot 2d]$ by $G_0 \exp[-jk \cdot 2d]$, to find formulas for the power transmission and reflection coefficients. The answers are

$$T = \frac{G_0(1 - R_1)\cdot(1 - R_2)}{(1 - G_0\sqrt{R_1R_2})^2 + 4G_0\sqrt{R_1R_2}\sin^2\theta} \tag{6.7.1}$$

$$R_{\text{net}} = \left[\frac{E_r}{E_i}\right]^2 = \frac{(\sqrt{R_1} - \sqrt{R_2})^2 + 4G_0\sqrt{R_1R_2}\sin^2\theta}{(1 - G_0\sqrt{R_1R_2})^2 + 4G_0\sqrt{R_1R_2}\sin^2\theta} \tag{6.7.2}$$

These functions are sketched for various values of G_0 in Fig. 6.8 assuming that the wave impinging on the cavity contains 1 watt of optical power. Note that the reflected power does not go to zero at resonance for $G_0 > 1$; indeed, it peaks along with the transmitted wave, both being considerably larger than the incident power. All the "action" takes place near resonance, and it, too, is critically dependent on the single pass gain G_0.

The full width at half maximum of the resonance is

$$\text{FWHM} = 2\Delta\theta_{1/2} = \frac{1 - G_0\sqrt{R_1R_2}}{G_0^{1/2}\sqrt[4]{R_1R_2}} \tag{6.7.3}$$

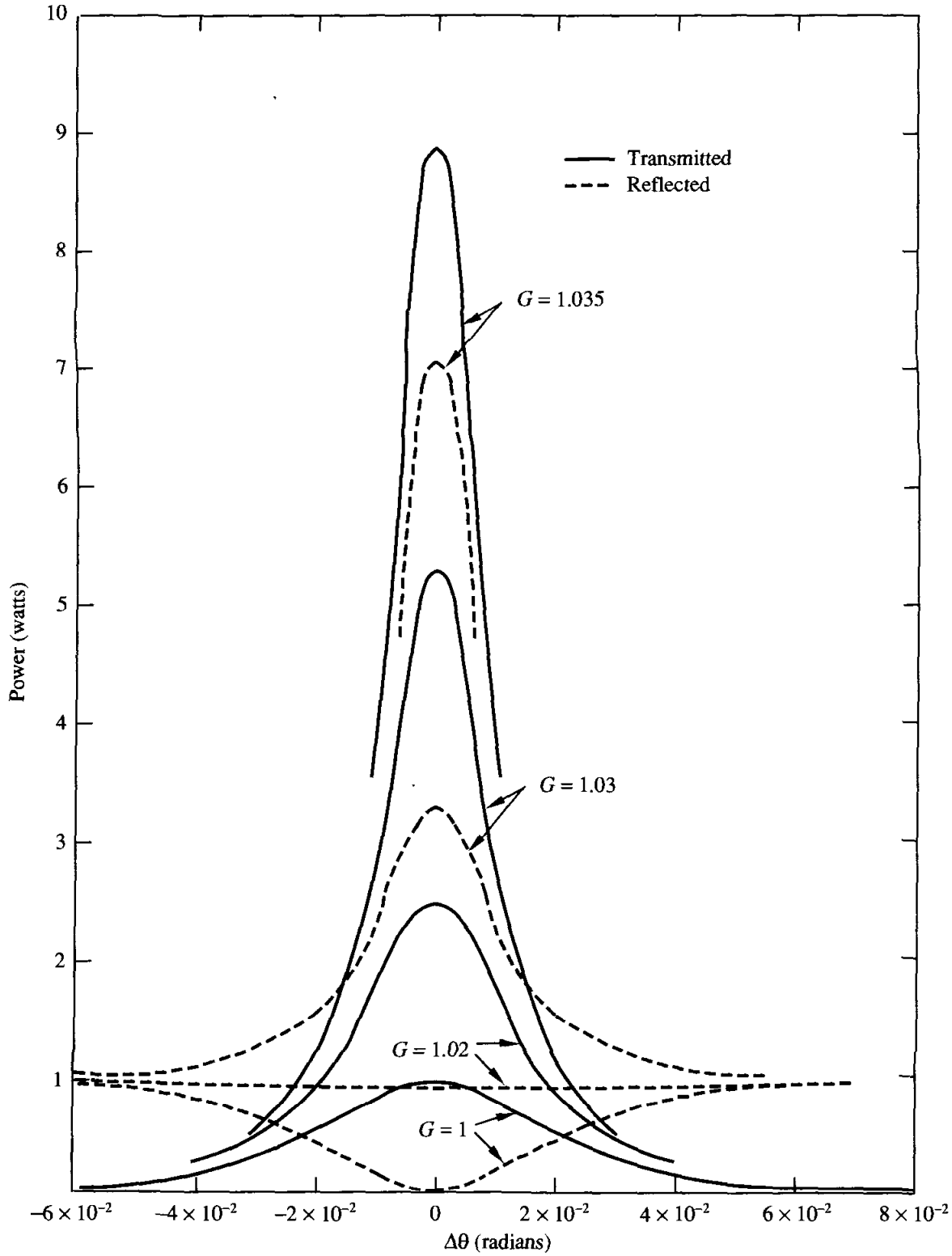


FIGURE 6.8. Response of an active cavity. The parameter G is the single pass power amplification factor.

If G_0 becomes bigger than $1/R_1R_2$, the mathematics indicates an infinite (!) output with a zero-line width. What this absurdity means is that we have an oscillator, and we can shut off the external source. Thus it is essential to discuss the physics leading to the gain G_0, and this is the subject of the remainder of the book.

PROBLEMS

6.1. The following questions refer to the optical cavity shown in Fig. 6.5 with $d = (3/4)R_2$, $\Gamma_1^2 = 0.99$, and $\Gamma_2^2 = 0.97$.

(a) Find an expression for the resonant frequencies of the $TEM_{0,0}$ modes of the cavity.

(b) If the radius of curvature is 2.0 m and the wavelength region of interest is 5000 Å, compute the following quantities:

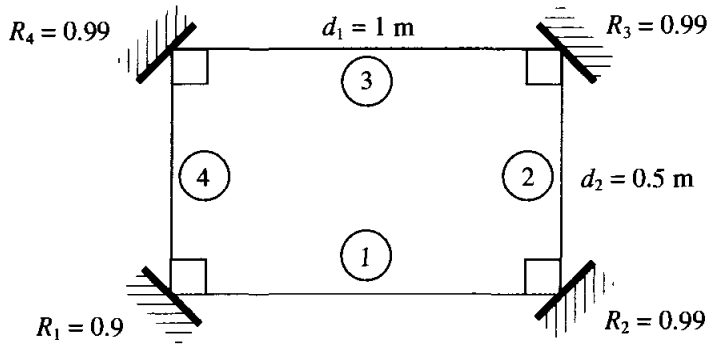
(1) Free spectral range in MHz and in Å units

(2) Cavity Q

(3) Photon lifetime in nsec

(4) Finesse

Problems 6.2 through 6.4 refer to the optical cavity shown in the accompanying diagram.



6.2. If the optical paths 1 through 4 are lossless, what is the photon lifetime of this cavity? (Ans.: 78.9 ns.)

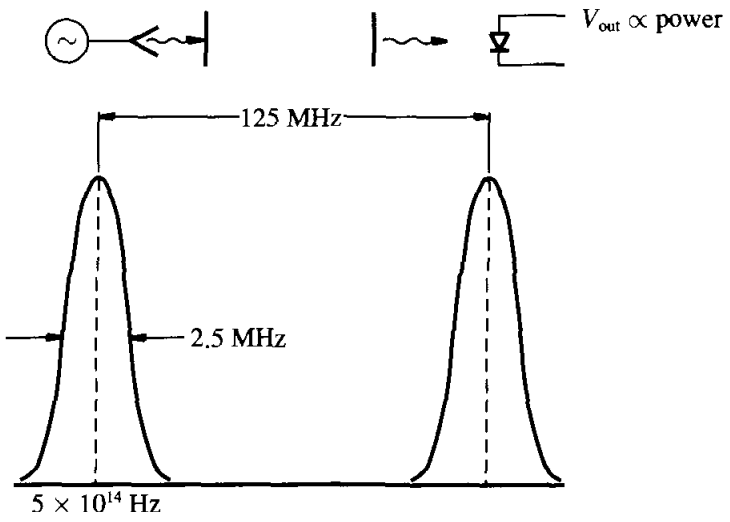
6.3. What is the cavity Q (assume that the wavelength region of interest is 5000 Å)? (Ans.: 2.97×10^8.)

6.4. **(a)** Suppose that path 1 has a transmission coefficient of 0.85 rather than 1 as in Problem 6.2. What is the new photon lifetime? (Ans.: 38.8 nsec.)

(b) Suppose that path 1 had a power gain of 1.1. What is the new photon lifetime?

(c) If we blindly plug into the formulas, τ_p becomes negative for G sufficiently large. What is the meaning of this apparent absurdity?

Problems 6.5 through 6.10 refer to the optical cavity in the accompanying diagram. It is excited by a variable-frequency source, and the detected intensity is as shown.



6.5. What is the nominal wavelength of the source?

6.6. How long is the cavity?

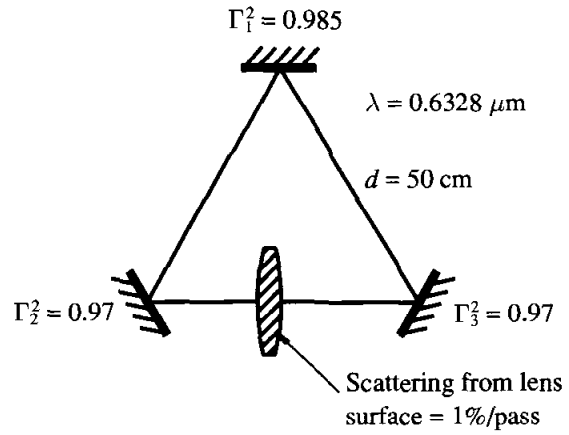
6.7. What is the finesse?

6.8. What is the Q?

6.9. What is the photon lifetime?

6.10. Suppose that the cavity is filled with an active medium with a single-pass gain of G. How large should G be to obtain oscillation?

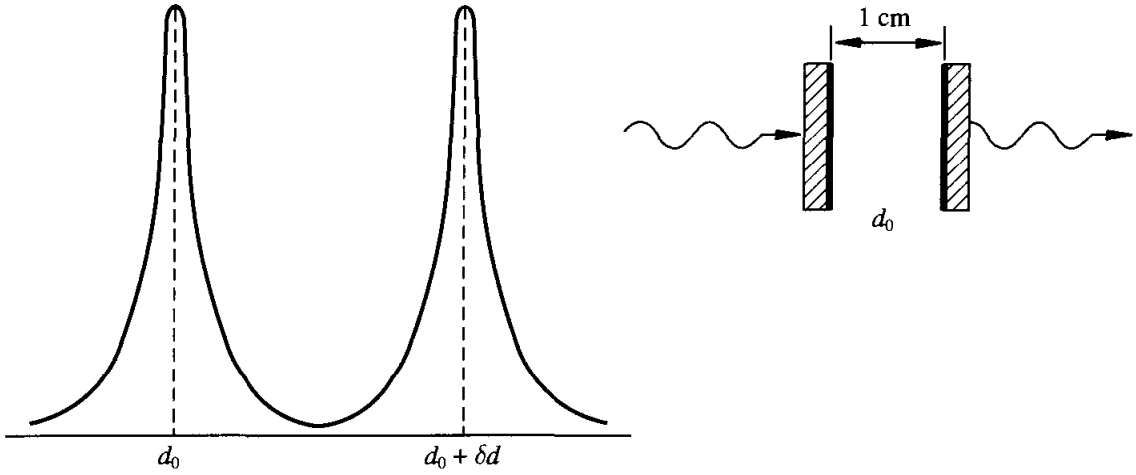
6.11. The following two questions refer to the laser cavity shown in the accompanying diagram.



(a) What is the photon lifetime? (Ans.: = 78.66 ns.)

(b) What is the cavity Q? (Ans.: 2.35×10^8.)

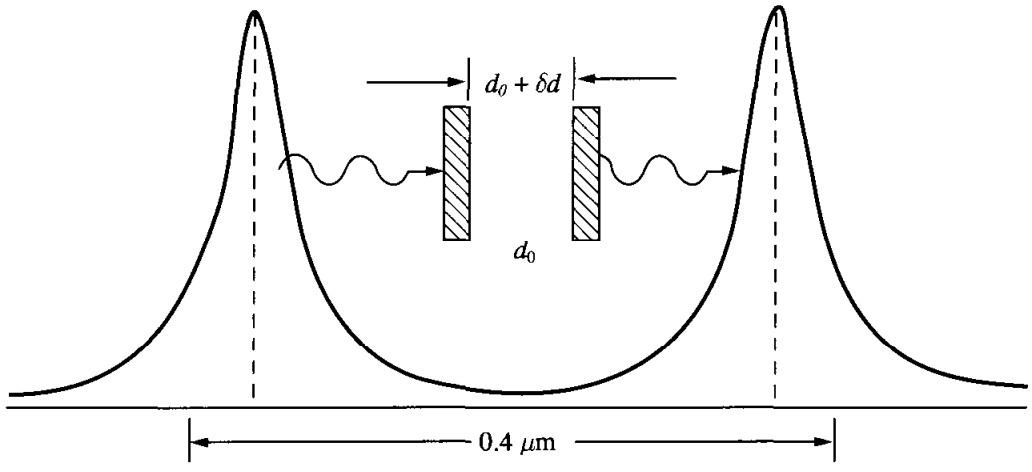
6.12. Drawn to scale on the graph below is the relative power transmission through a Fabry-Perot cavity when the distance d is increased slightly. The source is a He:Ne laser at $\lambda_0 = 6328$ Å.



(a) What is the distance δd?
(b) What is the finesse of the cavity?
(c) What is the cavity Q?

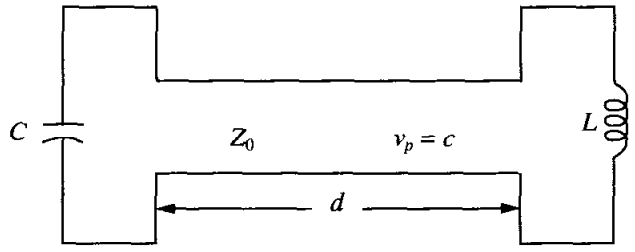
6.13. Drawn to scale in the graph below is the relative power transmitted through the cavity as the distance d is increased from its initial value of 2 cm to 2 cm + 0.5 μm. The source is a single-mode laser of wavelength λ_0.

(a) What is the wavelength of the source?
(b) What is the finesse?
(c) Find the full width at half maximum (FWHM) of the resonance and express your answer in MHz.



(d) What is the cavity Q?
(e) Find the photon lifetime.

6.14. Consider a cavity constructed by terminating a transmission line of length d with an inductance L at one end and a capacitance C at the other. Assume that the characteristic impedance of the line is Z_0 and its phase velocity is c.

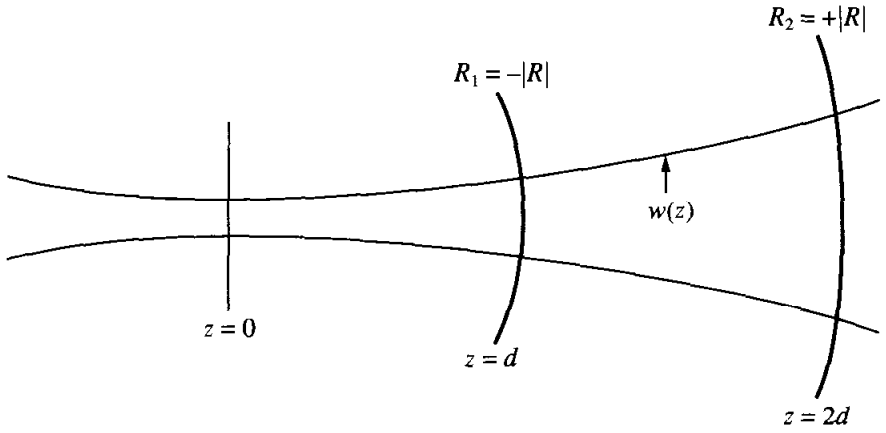


(a) Find a transcendental equation that determines the resonant frequency of this cavity by applying the condition that the round-trip phase shift is an integral number times 2π radians.

(b) Show that the equation derived for (a) reduces to $\omega_0^2 = 1/LC$ if d is sufficiently small.

[HINT: From transmission line theory, the voltage reflection coefficient is given by $\Gamma = (Z - Z_0)/(Z + Z_0)$, and thus there is a phase shift associated with Γ for complex terminations.]

6.15. Consider the optical cavity shown in the diagram below in which the variation of the spot size $w(z)$ is also shown. Note that the beam waist occurs at a distance d to the left of M_1 and that the mirrors have curvatures of opposite signs. (Assume that M_1 is so thin that it does not focus or otherwise affect the beam passing through it.)

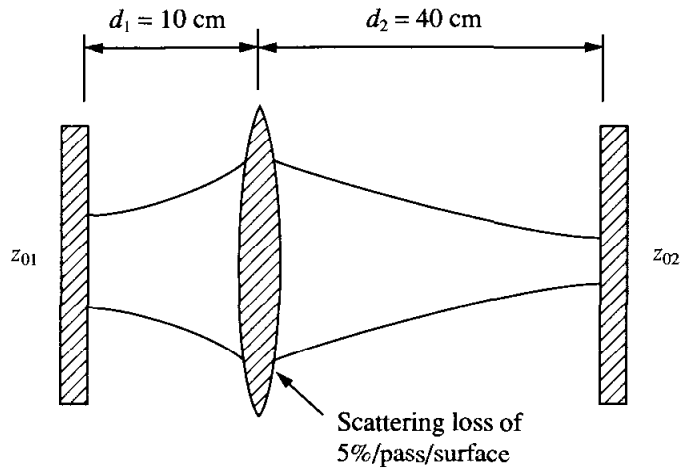


(a) Assume that the cavity is stable. Solve for the ratio d/R and evaluate the parameter z_0^2.

(b) If $d = 100$ cm and $z_0 = 100\sqrt{2}$ cm, what is the difference between the $\text{TEM}_{0,0,q}$ and $\text{TEM}_{1,0,q}$ resonant frequencies (in MHz)?

6.16. **(a)** Find an expression for the difference in the resonant frequencies of the $TEM_{0,0,q}$ and the $TEM_{m,n,q}$ modes of the cavity shown below. You may assume that the cavity is stable and the parameters z_{01} and z_{02} are known. Express your answer in terms of d_1, d_2, z_{01}, and z_{02} —(do not evaluate).

(b) What is the photon lifetime and Q of the passive cavity if $\lambda_0 = 500$ nm and $R_1 = R_2 = 0.876$? (A numerical answer is required.)



6.17. Make a careful sketch of the peaks (and valleys) of the transmission through a Fabry-Perot cavity as a function of frequency around $\lambda_0 = 6000$ Å with a finesse of 10 and free spectral range of 20 GHz.

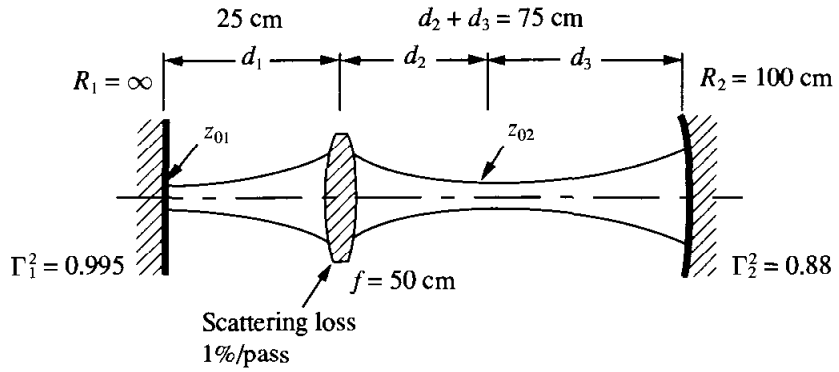
(a) Find a numerical value for FWHM in GHz, Å, and cm^{-1}.

(b) What is the cavity Q?

(c) What is the photon lifetime?

(d) What is the free spectral range in GHz, Å, and cm^{-1}?

6.18. Consider the following laser cavity and assume that the parameters z_{01} and z_{02} are known. Assume also that the cavity is stable.



(a) What is the radius of curvature of the phase at the spherical mirror? (Ans.: 100 cm.)

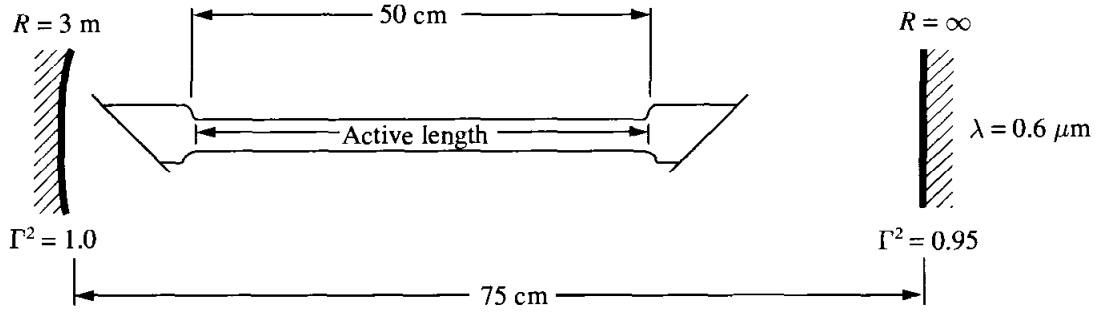
(b) What is the photon lifetime? (Ans.: 47.01 ns)

(c) Derive a formula for the resonant frequency of the $\text{TEM}_{m,p,q}$ mode.

$$\left(\text{Ans.: } \nu = \frac{c}{2(d_1 + d_2 + d_3)}\left\{q + \frac{1 + m + p}{\pi}\left[\tan^{-1}\left(\frac{d_1}{z_{01}}\right) + \tan^{-1}\left(\frac{d_2}{z_{02}}\right) + \tan^{-1}\left(\frac{d_3}{z_{03}}\right)\right]\right\}\right)$$

6.19. Consider the laser shown in the accompanying diagram.

(a) Is this cavity stable?



(b) What would be the frequency difference between the $\text{TEM}_{0,0,q}$ mode and the $\text{TEM}_{1,0,q}$ mode? (Ans.: = 33.3 MHz.)

(c) What should be the bore size of the laser tube so that less than 0.1% of the $\text{TEM}_{0,0,q}$ mode intercepts the tube wall? (Ans.: 2.14 mm diameter.)

(d) What is the minimum gain *coefficient* of the laser tube to sustain oscillation? (Ans.: 5×10^{-4} cm^{-1}.)

6.20. A laser cavity was excited by a 0.1 ns pulse from an external source at $\lambda = 5577$ Å. When the medium was not pumped, the detected transmission was as shown in the part (a) of the diagram. When the medium was irradiated by an intense electron beam, part (b) of the diagram resulted.

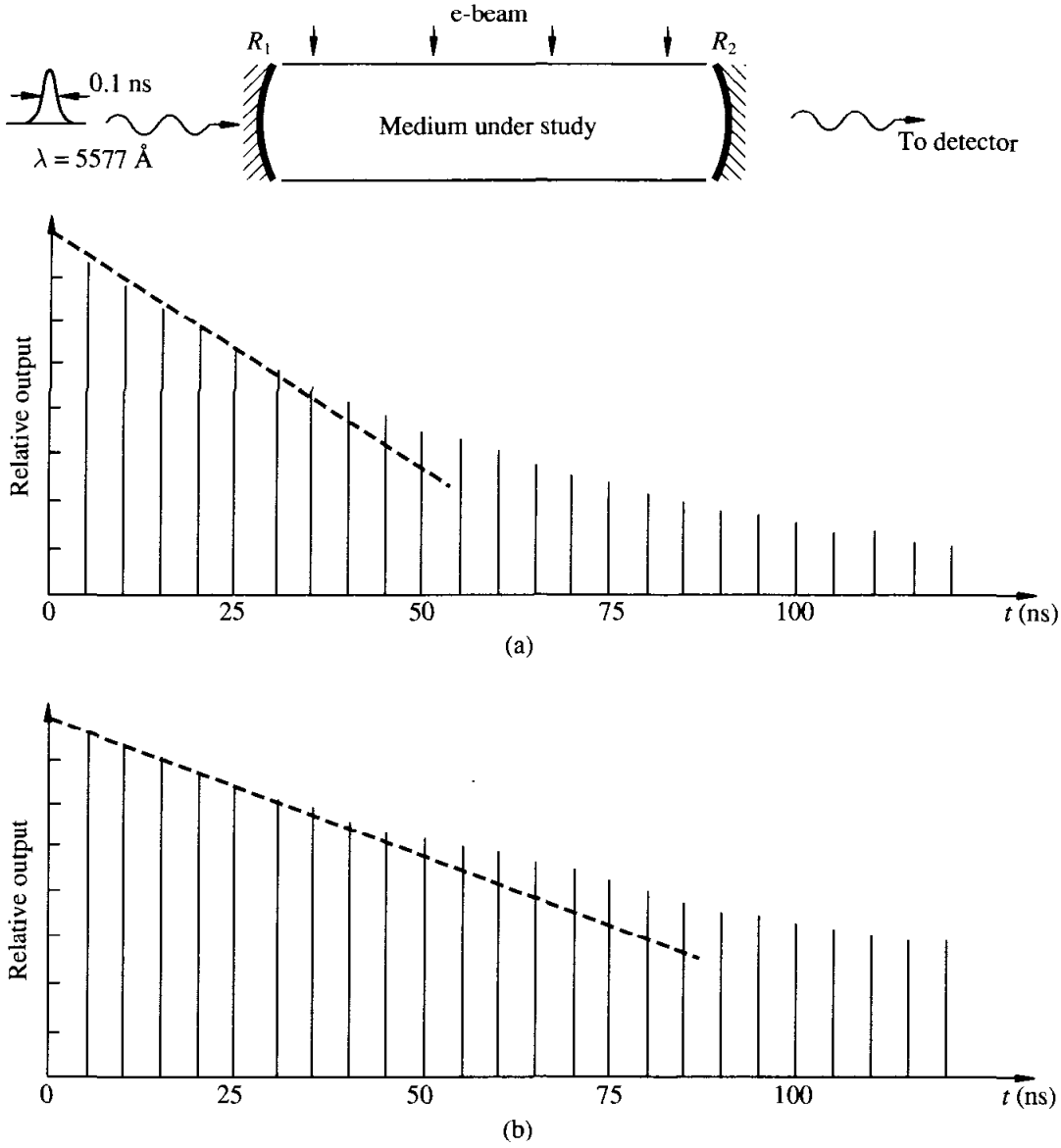
(a) How long is the cavity? (Ans.: 75 cm.)

(b) What is the photon lifetime? (Ans.: From graph, 75 ns.)

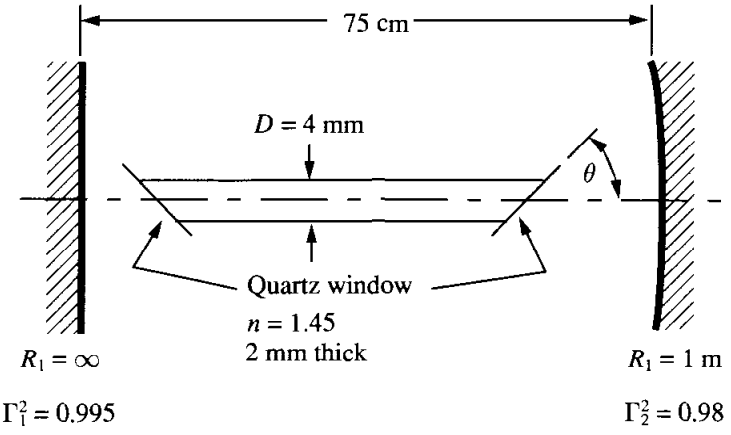
(c) What is the cavity Q? (Ans.: 2.5×10^8.)

(d) What is the single pass gain under e-beam excitation? (Ans.: 1.0142.)

(e) What is the cold-cavity finesse? (Ans.: $F = 94.2$.)



6.21. Consider the accompanying diagram of a cavity designed to be utilized with a helium/neon laser at $\lambda_0 = 632.8$ nm.



(a) Is the cavity stable?
(b) What is the spot size of the beam at the flat mirror?
(c) What is the spot size of the beam at the spherical mirror?
(d) The windows are cemented to the tube at Brewster's angle. What is the angle θ as shown on the sketch?
(e) Assuming that the tube bore is centered with respect to the axis of the $TEM_{0,0}$ mode, compute the loss introduced by the aperturing action of the tube walls. (Zero is *not* an acceptable answer.)
(f) What is the formula for the resonant frequency of the $TEM_{m,p,q}$ mode?

6.22. Consider a single $TEM_{0,0,q}$ mode of the laser shown in Problem 6.21. Because of room vibrations, sound waves, and temperature variations, the distance $d = 75$ cm varies slightly about its nominal value. If the optical frequency of a mode is to be held constant to 1 kHz, what is the maximum allowable variation in d? The answer should disturb you, especially when you consider that atoms are *spaced* about 4 Å apart. Nevertheless, such frequency control is possible.

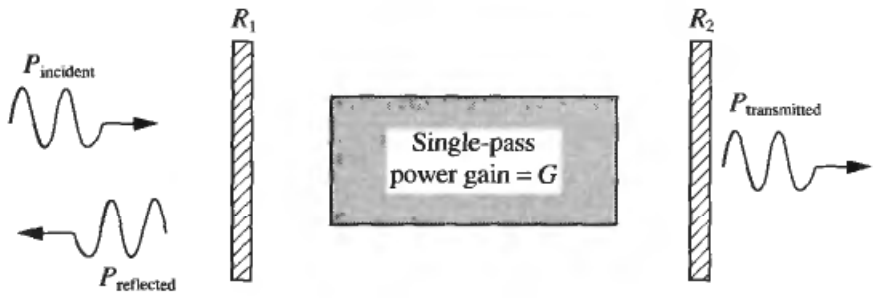
6.23. A variable-frequency, constant-amplitude dye laser irradiates the active Fabry-Perot cavity containing a medium with a single-pass power gain of G. Even though there is gain, it is not sufficient to support oscillation (i.e., $G^2 R_1 R_2 < 1$). Assume $R_1 = R_2 = 0.90$.

(a) Find the power magnification factor of this cavity if the source is tuned to the center of the cavity resonance; that is, find

$$M = \frac{P_{\text{ref}} + P_{\text{trans}}}{P_{\text{inc}}}$$

NOTE: If $G = 1$ (i.e., a passive cavity), then $M = 1$ since this is a lossless system.

(b) Plot the transmission through the cavity as the frequency of dye laser is tuned around resonance (assume $G^2 R_1 R_2 = 0.95$).
(c) Compute and plot the line width of the cavity as the gain is varied from 1 to 99% of its maximum value (i.e., $G^2 = 1/R_1 R_2$).



6.24. One of the means for "tuning" a Fabry-Perot cavity is by changing the index of refraction of the medium between the two mirrors, which is accomplished by varying a gas density in the enclosure. The index of refraction scales as $n = 1 + kp$, where p is the pressure in atmospheres. Suppose the etalon spacing is 0.2 cm, the finesse was

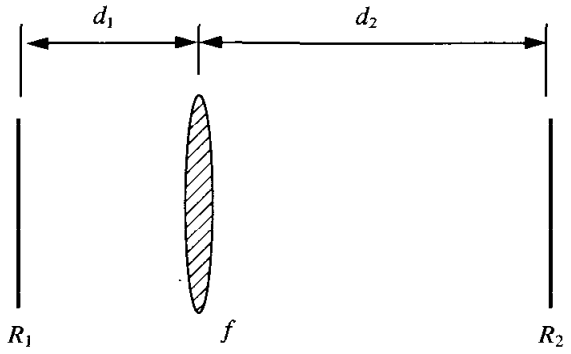
20 at 6328 Å, and the gas constant in the index equation is 8×10^{-4} atmospheres^{-1}. Make a careful sketch of the relative transmission through the etalon as the pressure is varied.

(a) What change in pressure is required to scan the etalon over one free spectral range?

(b) What is the spectral range (in GHz)?

(c) What is the photon lifetime and Q of this cavity?

6.25. For the cavity shown below, the Hermite-Gaussian beam parameters are given by $z_{01} = \pi w_{01}^2/\lambda_0 = 6.45$ cm and $z_{02} = 38.7$ cm with $d_1 = 25$ cm, $d_2 = 50$ cm, $R_1 = 0.98$, $R_2 = 0.93$, and a transmission through the lens of 95%. The wavelength region of interest is 5145 Å.



(a) Find a formula for the resonant frequencies of the $\text{TEM}_{m,p,q}$ modes in terms of the distances $d_{1,2}$ and the parameters z_{01} and z_{02}.

(b) What is the photon lifetime?

(c) Evaluate the passive Q.

(d) If an active medium were incorporated within the cavity and it amplified the intensity by a factor of 1.13 per pass, find the new photon lifetime and line width of the cavity response.

6.26. A GRIN waveguide is constructed from a material described in Problem 5.3. There is a reflection at each end from the air-dielectric mismatch or from intentional coating, and thus a length can be considered as a cavity with Gaussian beams as its characteristic modes.

(a) Use the $ABCD$ law to find the complex beam parameter, q, of a Gaussian beam whose spot size does *not* change along the z axis. Evaluate this spot size for $n_0 = 1.87$, $\lambda_0 = 1\mu m$, and $a = 250\mu m$.

(b) The phase constant of the Gaussian beam found in (a) is given by

$$\beta \approx \left(k - \frac{1}{a}\right)$$

where $k = \omega n_0/c$ as usual. Find an expression for the resonant frequencies of the $\text{TEM}_{0,0,q}$ modes for this cavity.

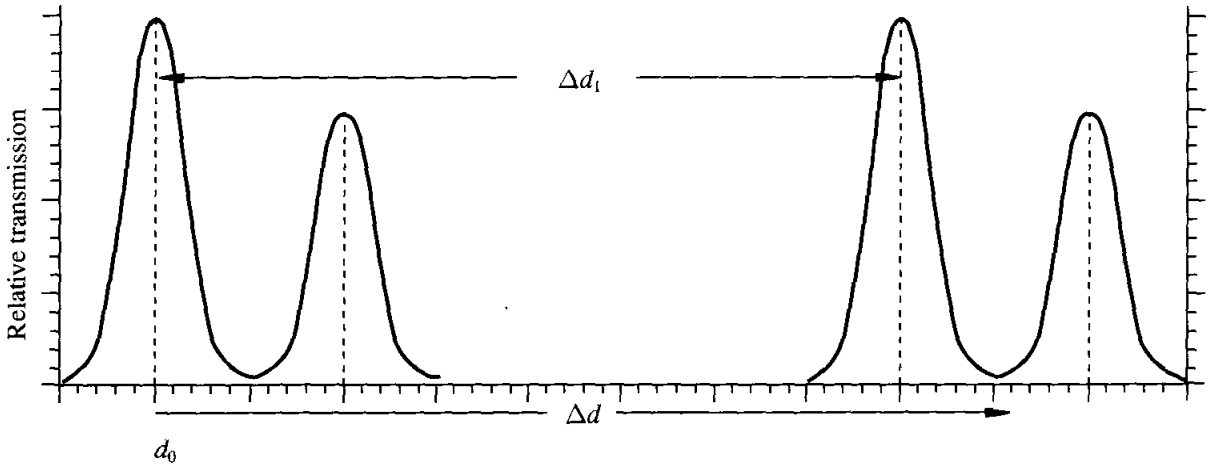
(c) The power reflectivities at the planes $z = 0$ and $z = d$ can be approximated by the usual plane wave or transmission line formula:

$$R = \left[\frac{n_0 - 1}{n_0 + 1}\right]^2$$

Justify the use of this formula, given that the spot size found in (a) is much less than the scale length a.

(d) If the ends were coated such that the reflectivities were 0.5, $d = 10$ cm, $n_0 = 1.87$, and $a = 250\ \mu m$, find the photon lifetime and cavity finesse.

6.27. A particular (semiconductor) laser oscillates at a nominal wavelength of $\lambda_0 = 8800$ Å, but its spectrum consists of two closely spaced modes, $\lambda_1 = 8800.WXYZ$ Å and $\lambda_2 = 8800.IJKL$ Å. To resolve the wavelength difference between the modes, we use a scanning Fabry-Perot cavity in which the mirror spacing is changed from $d_0 = 4$ cm to $d_0 + \Delta d$ with $\Delta d \ll d_0$. The relative transmission through the cavity is shown below (and is drawn to scale) as a function of the change Δd.



(a) What is the distance Δd_1 (in μm) as shown on this diagram?

(b) The major peaks are the resonances associated with λ_1 and the minor ones belong to λ_2. Use the graphical data to estimate $(\lambda_2 - \lambda_1)$.

(c) What is the finesse of the Fabry-Perot cavity?

6.28. The laser shown on the diagram below generates a field given by

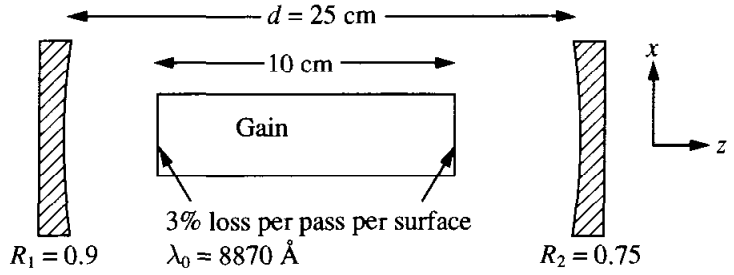
$$\mathbf{E} = \mathbf{a}_y E_{m,p} \frac{y w_0}{w^2(z)} \exp\left[-\frac{(x^2 + y^2)}{w^2(z)}\right] \exp\left[-j\frac{k(x^2 + y^2)}{2R(z)}\right]$$

$$\cdot \exp\left\{j\left[kz - (1 + m + p)\tan^{-1}\left(\frac{z}{z_0}\right)\right]\right\}$$

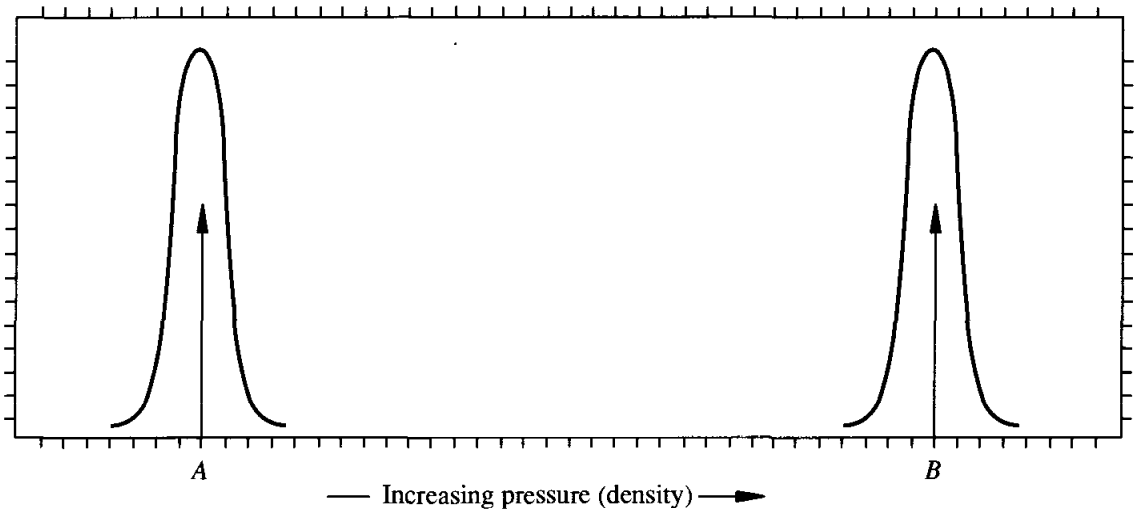
(a) Identify the mode (i.e., $\text{TEM}_{m,p}$; $m, p = ?$).

(b) Find the cavity Q.

(c) Derive an expression for the oscillation frequency given that $z = 0$ occurs at the midpoint of the cavity.



6.29. One method of tuning a Fabry-Perot cavity is to change the gas density in the space between the two mirrors. (Even though the index of most gases is very close to 1, there is a small positive part proportional to density.) The following is a sketch of the transmission through the cavity of 0.6328 μm radiation when the gas pressure is changed. The spacing between the mirrors is fixed at 1 cm.



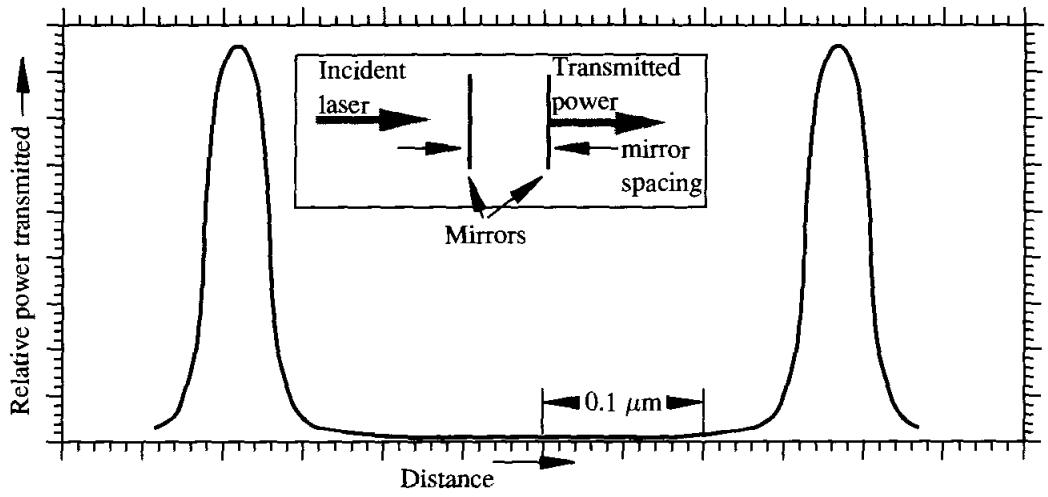
(a) By how much has the index of refraction changed between the points labeled B and A?

(b) What is the photon lifetime of the cavity?

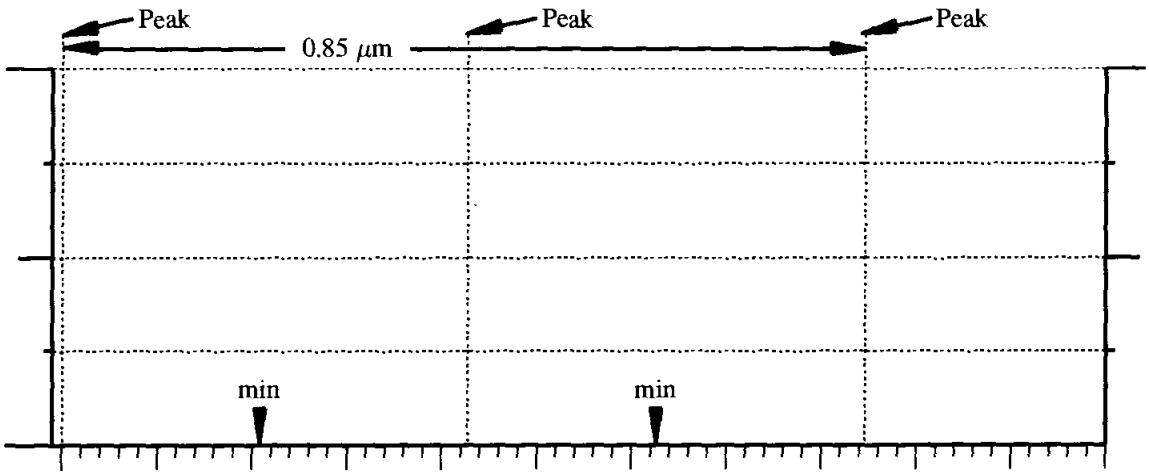
6.30. A laser of unknown wavelength λ_0 excites the cavity in the manner shown in the insert of the graph shown below. The mirror spacing is changed from a nominal value of 0.5 cm to $(0.5 + 6 \times 10^{-5})$ cm and the relative power transmitted through the cavity is recorded on the graph shown below.

(a) What is the wavelength of the laser?

(b) What is the cavity finesse, photon lifetime, and Q. (A graphical solution is desired).



6.31. A laser excites a cavity with the following specifications: F(finesse) $= 10$ with a nominal mirror spacing of $d_0 = 1$ cm and air as the dielectric ($n = 1$). When the mirror spacing is changed from 1 cm to 1 cm+0.85 μm, the transmission exhibits two additional peaks (as indicated below) and decreases to virtually zero at points between the maxima.



(a) What is the wavelength of the laser?

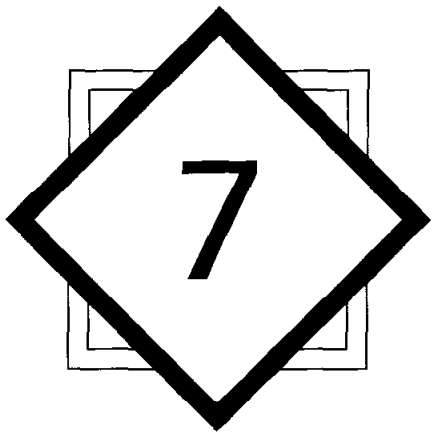
(b) Make a careful sketch of the transmission you would expect through the cavity and *label* the distance δd (in μm) corresponding to the FWHM of the resonance.

(c) What is the FWHM of the cavity resonance expressed in Hz and wavenumber units?

REFERENCES AND SUGGESTED READINGS

It should be obvious to all by now that Chapter 6 is the culmination of the previous five chapters. Much of any "new" material contained in this chapter can be found in any book on elementary electrical circuits, possibly under a different name, but still the same ideas.

1. A. G. Fox and T. Li, "Resonant Modes in a Maser Interferometer," *Bell Syst. Tech. J. 40*, 453-488, Mar. 1961.
2. H. Kogelnik and T. Li, "Laser Beams and Resonators," *Appl. Opt. 5*, 1550-1567, Oct. 1966.
3. A. E. Siegman, "Unstable Optical Resonators for Laser Applications," *Proc. IEEE 53*, 227-287, Mar. 1965.
4. A. E. Siegman, *Lasers* (Mill Valley, Calif.: University Science Books Company, 1986).
5. A. Maitland and M. H. Dunn, *Laser Physics* (Amsterdam: North-Holland, 1969).
6. S. Ramo, J. R. Whinnery, and T. VanDuzer, *Fields and Waves in Communication Electronics* (New York: John Wiley & Sons, 1965).
7. A. Yariv, *Introduction to Optical Electronics* (New York: Holt, Rinehart and Winston, 1971).
8. A. Yariv, *Quantum Electronics*, 2nd ed. (New York: John Wiley & Sons, 1975).
9. H. A. Haus, *Waves and Fields in Optoelectronics* (Englewood Cliffs, N.J.: Prentice Hall, 1984).



Atomic Radiation

7.1 INTRODUCTION AND PRELIMINARY IDEAS

This chapter introduces the elementary concepts leading to gain in an atomic system and thus, with proper feedback, a laser. Since a laser is a quantum device, we first review the necessary concepts from this theory to understand the conditions leading to gain.

These concepts are amazingly simple and are quite palatable, even to those who have not had a formal introduction to quantum mechanics.

1. There are discrete energy levels in an atomic (or molecular, or solid, or semiconductor) system. We represent them by an energy-level diagram as shown in Fig. 7.1.

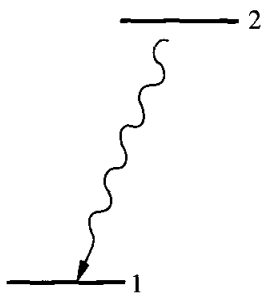


FIGURE 7.1. Energy-level diagram.

2. The system can make a *transition* between these two states by the *emission* of a photon of energy $E = E_2 - E_1 = h\nu$, thereby changing an atom labeled by 2 into one identified with 1. Or to reverse the process, an atom in state 1 can *absorb* a photon of this same energy and be labeled as in state 2.

These two ideas are the only ones we need to "accept"; the rest will follow naturally and with little effort on our part.

In addition to the simple ideas expressed, we need to understand the origin of the correct description of blackbody radiation theory. (At low frequencies, this radiation is referred to as "white" noise, "Johnson's" noise, or "Nyquist" noise, all being equal to Planck's blackbody radiation formula at those limits.) Indeed, it was this problem that inspired Planck to make the quantum hypothesis; namely, that electromagnetic energy at a frequency ν can be present only in discrete multiples of $h\nu$.

Understanding the origin of Planck's development of this formula is essential to appreciate the beauty and *simplicity* of Einstein's description of the role of the atoms in arriving at this relation. The importance of Einstein's approach is that it provides the key to the analysis of systems that are *not* in thermodynamic equilibrium, with the laser being a prime example.

Furthermore, Einstein's approach provides us with a connection between *transition rates* between quantum states and quantities that can be measured experimentally. Consequently, we can approach such transitions in a phenomenological manner by using the published experimental results, thereby bypassing some very complex and complicated quantum calculations.

The foregoing should not be construed to downgrade the importance of a full quantum description of a laser. There are phenomena not predicted by the rate-equation approach and that require a detailed analysis by quantum-theoretical methods. But for the initial understanding, Einstein's rate equations are quite adequate.

We will also have to modify slightly the picture shown in Fig. 7.1 to account for the *uncertainty principle* and "real-life" broadening processes. An atom is never isolated. A gas atom is in helter-skelter motion, it collides with other atoms (with the same kind or with different ones), and it may be subjected to external fields. As a consequence, the energy levels are not perfectly sharp, and thus there is a finite band of frequencies emitted by a collection of atoms and a finite band for amplification.

Once these ideas are in hand, the laser gain equation is a trivial application and oscillation is merely a matter of providing feedback.

7.2 BLACKBODY RADIATION THEORY

The starting point for the quantum era was the derivation, from first principles, of a formula that described the radiation emerging from a small hole in a highly polished "cavity" such as that shown in Fig. 7.2(a). The *experimental facts* known from the late 1800s were as follows:

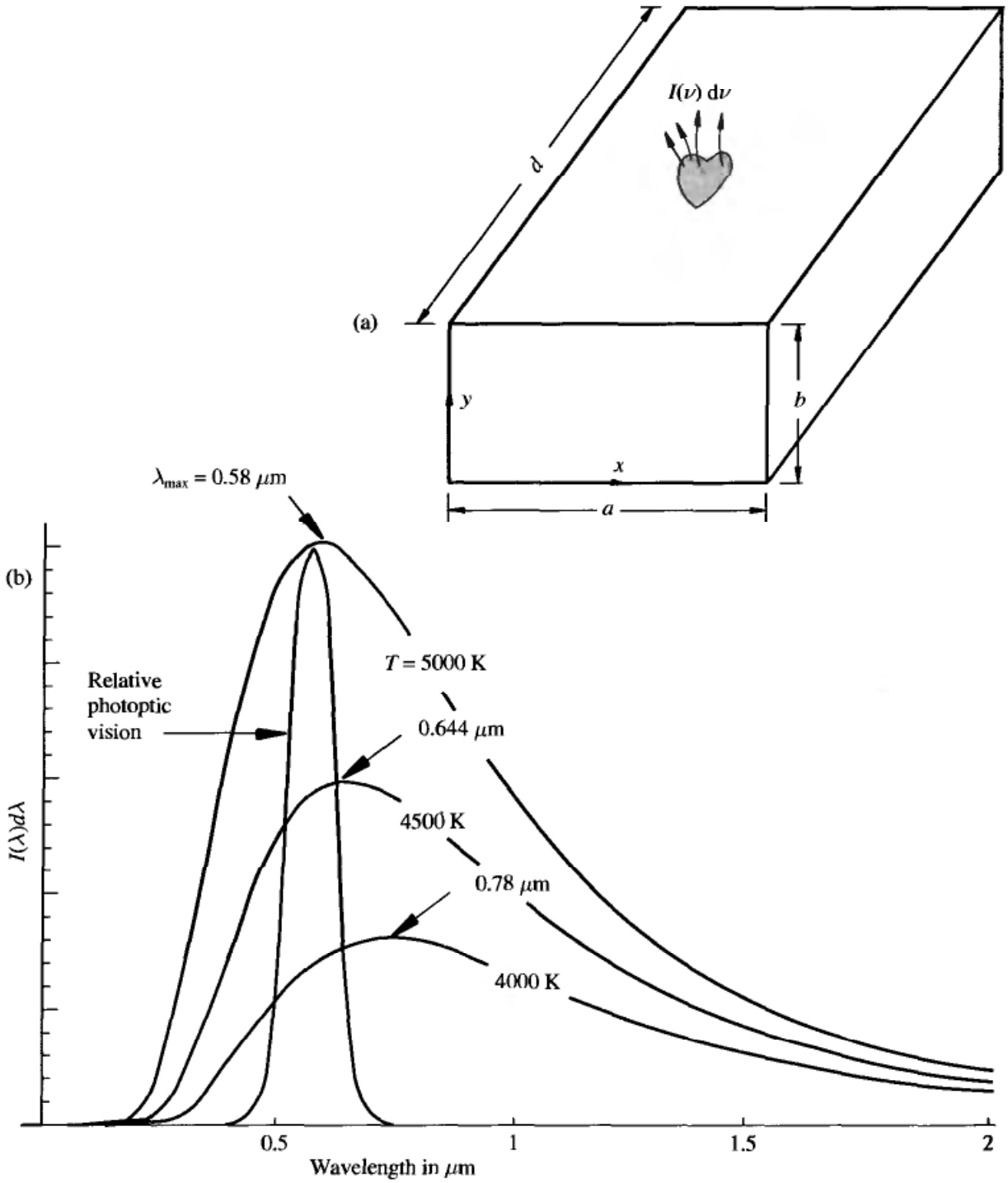


FIGURE 7.2. (a) The cavity used to measure the blackbody spectrum; (b) $I(\lambda)d\lambda$ from experiment. (NOTE: $\lambda_{max}T = 2.898 \times 10^7$ Å-°K.) The bell-shaped curve in (b) illustrates the response of the eye.

1. The hole acted as a nearly perfect blackbody with an emission coefficient* (as a function of frequency) equal to 1 (one).
2. Thus the intensity emitted by this hole is directly proportional to the energy density of the electromagnetic radiation inside the cavity.
3. This energy density, in a fixed-frequency interval dv (determined by the measurement technique), had the experimental value of

$$\rho(v)\,dv = \frac{8\pi h v^3\,dv}{c^3}\,\frac{1}{e^{hv/kT}-1}$$

or

$$\rho(\lambda)\,d\lambda = \frac{8\pi h\,d\lambda}{\lambda^5}\,\frac{1}{e^{hc/(\lambda kT)}-1} \tag{7.2.1}$$

It is important to realize that these facts were known from experiment—even the value of Planck's constant—*before* the advent of quantum mechanics. Even though these facts were known, they were not understood.

One of the triumphs of nineteenth-century physics was the phenomenal success of Maxwell's theory of electromagnetic radiation. Yet here was an electromagnetic problem, apparently coupled to some elementary thermodynamics, which defied explanation. (In fact, Planck, who was ultimately successful, nearly had a nervous breakdown over the dilemma.) To illustrate some of the difficulties, let us follow some of the logic paths used to arrive at the *wrong* answer.

Every electromagnetic phenomenon until that time (and since) had (and has) been explained by Maxwell's theory. Hence it was natural to start there and ask how electromagnetic energy is distributed inside the heated cavity shown in Fig. 7.2(a). (Although a specific geometry is chosen, the results are independent of the shape providing *all* dimensions are large compared to a wavelength in the medium inside cavity.) From Chapter 6 we recognize that only the resonant $TE_{m,p}$ or $TM_{m,p}$ modes of the short-circuited rectangular waveguide can have appreciable energy. Those resonances are determined by the standard formula: RTPS=$q \cdot 2\pi = \beta_{m,p} \cdot 2d$ where $\beta_{m,p}$ is the phase constant of the mode:

$$\beta_{m,p} = \left\{\left(\frac{\omega n}{c}\right)^2 - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{p\pi}{b}\right)^2\right\}^{1/2} \tag{7.2.2a}$$

Equation (7.2.2a) applies for either TE or TM modes. Thus the resonant frequencies of the (m, p, q) mode are given by $\beta_{m,p} \cdot 2d = q \cdot 2\pi$ or

$$\mathbf{k}^2 = \left(\frac{\omega n}{c}\right)^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{p\pi}{b}\right)^2 + \left(\frac{q\pi}{d}\right)^2 \tag{7.2.2b}$$

Equation (7.2.2b) has a nice geometric interpretation: The term on the left is the wave vector $\mathbf{k}$ of a uniform plane wave in the medium; the terms on the right can be considered as the projection of it along the (x, y, z) axes; and, of course, the right-hand side states that the

*Kirchhoff's radiation law states that the emissivity of a body is equal to its absorptivity. Thus the cavity would also absorb all electromagnetic energy entering through the hole and emit as much as possible.

Pythagorean relationship holds. Resonance, then, implies that each component of **k** must be a multiple of a half wavelength. Thus the resonant frequencies are given by

$$\left(\frac{2\pi\nu n}{c}\right)^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{p\pi}{b}\right)^2 + \left(\frac{q\pi}{d}\right)^2 \tag{7.2.2c}$$

To make the formula even simpler, we choose a cube with $b = d = a$ and find that the formula for resonance "looks" similar to a spherical coordinate radius vector with the mode numbers being a measure along the Cartesian axis:

$$\nu = \frac{c}{2na}[m^2 + p^2 + q^2]^{1/2} \tag{7.2.3a}$$

This geometric interpretation is shown in Fig. 7.3 where the "points" are allowed mode numbers scaled by a factor $c/2na$ so that there is one mode in a volume $(dm = 1) \times (dp = 1) \times (dq = 1)$. Thus we can find the number of modes between $\nu = 0$ and ν by finding the volume of one eighth of a sphere with the radius $R = 2na\nu/c$:

$$V_s = \frac{1}{8} \times \frac{4\pi R^3}{3} \tag{7.2.3b}$$

Now each point represents two modes, TE and TM, and thus the total number of resonances between 0 (i.e., DC) and ν, the frequency of interest, is

$$N = 2 \times \frac{1}{8} \times \frac{4\pi}{3}\left(R \to \frac{2na\nu}{c}\right)^3 = \frac{8\pi n^3 \nu^3}{3c^3} a^3 \tag{7.2.4a}$$

We are usually interested in the small frequency interval $d\nu$ around ν and also prefer the answer on a per-unit-volume basis to eliminate the geometry. The mode density $p(\nu)$

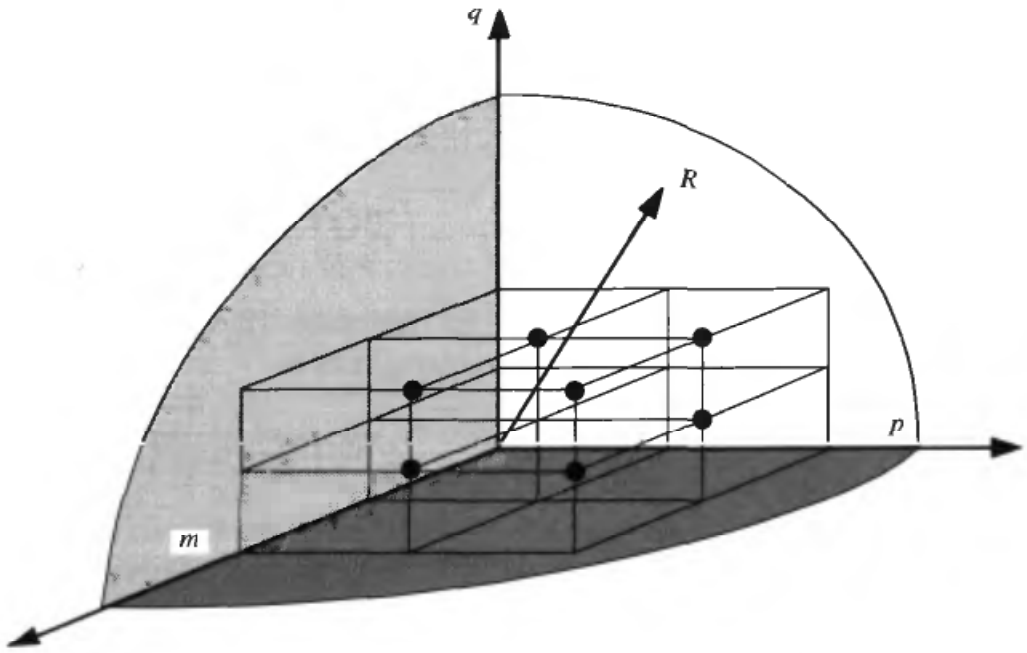


FIGURE 7.3. The mode diagram for cubic cavity.

(volume^{-1} – frequency^{-1}) is given by

$$p(\nu)d\nu = \frac{1}{V}\frac{dN}{d\nu}d\nu = \frac{8\pi n^2 \bar{n}_g}{c^3}\nu^2 d\nu \cong \frac{8\pi n^3}{c^3}\nu^2 d\nu \tag{7.2.4b}$$

where

$$\bar{n}_g = n + \nu\frac{dn}{d\nu} = n - \lambda\frac{dn}{d\lambda} \tag{7.2.4c}$$

The quantity $\bar{n}_g$ is called the group index and plays a role in materials, such as semiconductors and fibers, where the index of refraction has a significant variation with wavelength. However, to make the following relations less cumbersome, we will often use the approximate one, $\bar{n}_g \sim n$.

So much for identifying the electromagnetic modes that couple to that hole in the cavity. Nobody wanted to (or has) changed (7.2.4b), since it is *correct*. The problem always arose in assigning an energy to those modes.

Classical physics gets away with assigning an energy $kT/2$ to each "mode" of motion. For instance, free atoms in a gas can move along x, y, or z, and thus the average energy of N atoms is $N(3kT/2)$, which is in agreement with experiment. If we follow that logic, each electromagnetic mode should have an energy equal to kT since both E and H store energy. Multiplying (7.2.4b) by kT leads to an answer for the energy density in frequency interval $d\nu$, but, unfortunately, it is wrong and does not agree with the experiment.

$$\rho(\nu)\,d\nu = \left(\frac{8\pi n^3 \nu^2 d\nu}{c^3}\right)kT \qquad \text{(Wrong!)} \tag{7.2.5}$$

Equation (7.2.5) is the Rayleigh-Jeans distribution of blackbody radiation, which did *not* agree with experiment. Although it does agree with experiment at long wavelengths (microwaves and lower frequencies), it predicts the absurd conclusion that there is infinite energy in the system when *all* frequencies are concerned (the ultraviolet catastrophe).

It is hard to argue with the line of reasoning leading to (7.2.5): the number of modes times the average energy per mode. Thus the error must be in the computation of one of these quantities. Planck accepted and retained the mode calculation and turned his attention to computing the average energy per mode.

Planck made the initial quantum hypothesis that electromagnetic energy at a frequency ν could only appear as a multiple of the step size $h\nu$. Energies between $h\nu$ and $2\,h\nu$ do *not* occur! This is in stark contrast to classical ideas where all energies are permitted; that is, by allowing the field to have continuously variable amplitudes. Having made this radical departure from accepted concepts, he returned to classical Boltzmann statistics to compute the average energy.

If $\epsilon_1, \epsilon_2, \epsilon_3, \ldots$ are the allowed energies, we must weigh each energy by the relative probability that it can occur and then divide by the sum of all relative probabilities. According to Boltzmann statistics, the relative probability that an energy (ϵ_j) can occur is simply $\exp(-\epsilon_j/kT)$, the Boltzmann factor. Following the recipe above, we obtain

$$\epsilon_n = nh\nu \qquad \text{(the quantum hypothesis)} \tag{7.2.6}$$

$$\langle \epsilon \rangle = \frac{1h\nu e^{-h\nu/kT} + 2h\nu e^{-2h\nu/kT} + 3h\nu e^{-3h\nu/kT} + \cdots}{1 + e^{-h\nu/kT} + e^{-2h\nu/kT} + e^{-3h\nu/kT} + \cdots} \tag{7.2.7}$$

$$= \frac{\sum_{1}^{\infty} nh\nu e^{-nh\nu/kT}}{\sum_{0}^{\infty} e^{-nh\nu/kT}} \tag{7.2.8}$$

where

$e_n = nh\nu$

$n = 0, 1, 2, 3, 4, \ldots$ are the allowed value of the electromagnetic energy

$\exp(-\epsilon_n/kT) =$ relative probability of the energy ϵ_n

$\sum_{0}^{\infty} \exp(-\epsilon_n/kT) =$ the sum of all relative probabilities

We might be tempted to replace the summation by an integral, a procedure that yields an average energy given by (7.2.5), which is wrong. Fortunately, the series given in (7.2.8) can be summed exactly to yield

$$\langle \epsilon \rangle = \frac{h\nu}{e^{h\nu/kT} - 1} \tag{7.2.9}$$

Then the energy density of the electromagnetic field inside the cavity at the center frequency of interest is

$$\rho(\nu) = \left(\frac{8\pi n^2 n_g \nu^2}{c^3} \right) \cdot (h\nu) \cdot \frac{1}{e^{h\nu/kT} - 1} = \frac{8\pi n^2 n_g \nu^2}{c^3} \frac{h\nu}{e^{h\nu/kT} - 1} \tag{7.2.10}$$

Equation (7.2.10) is written as a product of three factors to emphasize the origins: the first is a purely classical electromagnetic result, with the second being the quantum value of the average energy of the classical field. Since the package of energy in the field is a multiple of the quantum of energy $h\nu$, the term $1/[\exp(h\nu/kT) - 1]$ is the number of quanta in a cavity mode.

Let us evaluate this number for a reasonable set of circumstances. Let $T = 1200°\text{K}$ (an "orange" color temperature) and $\lambda = 6000$ Å (an orange color); then $\nu = 5 \times 10^{14}$ and the average number of photons per mode is $\sim 10^{-9}$.

We must interpret that number correctly. If the energy comes in discrete steps, then we cannot have a fraction of a photon in a mode. Rather we have a photon in the particular cavity mode for a small—a very small—fraction of the time. This particular point is important and distinguishes a laser that has at least one photon in a particular mode *all* of the time.

7.3 EINSTEIN'S APPROACH: *A* AND *B* COEFFICIENTS

Einstein was able to arrive at the same functional form for the energy density per frequency interval $\rho(\nu)$ for a thermodynamic environment by a much simpler and much more transparent line of reasoning. Most important is the fact that the reasoning can be applied to a system which is *not* in thermodynamic equilibrium. He focused the attention on the *atoms* in the cavity walls which were responsible for the generation of the electromagnetic energy inside the cavity.

7.3.1 Definition of Radiative Processes

Einstein accepted the quantum hypothesis that the energy came in discrete packages, $E = h\nu$, and as a consequence there must be two energy states,* E_2 and E_1, associated with those atoms separated by that value as shown in Fig. 7.4. He identified three radiative processes that affect the concentrations of atoms in states 2 and 1. (The fact that there are many other very important processes affecting the concentration need not concern us now. We are limiting our view to the blackbody cavity.)

(a) Spontaneous emission (A_{21}). As the name implies, it appeared as if the atoms in state 2 decayed spontaneously to state 1, and, in doing so, they added their excess energy to the cavity field in the form of a photon. If the population density in state 2 was N_2, the decay of this state is given by

$$\left.\frac{dN_2}{dt}\right|_{\substack{\text{spontaneous}\\ \text{emission}}} = -A_{21}N_2 \tag{7.3.1}$$

Note that this equation says something very simple and transparent: If no other process took place, the atomic population would run "downhill" with a time constant $\tau = (A_{21})^{-1}$. Obviously, the bottom of the "hill," N_1, must *increase* just as fast as the top decreases.

(b) Absorption (B_{12}). In this process an atom in state 1 *absorbs* a photon from the field and thus converts the atom into one of those in state 2. The rate at which this process takes place must depend on the concentration of absorbing atoms and the field from which

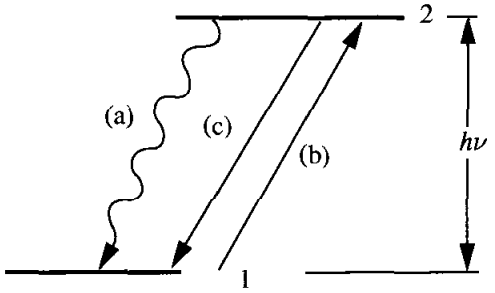


FIGURE 7.4. Radiative processes in a two-level system.

*As a matter of fact, there is a continuum of states, but this does not change the ideas expressed by a single set of states.

they extract the energy. Thus we have

$$\left.\frac{dN_2}{dt}\right|_{\text{absorption}} = +B_{12}N_1\rho(\nu) = -\left.\frac{dN_1}{dt}\right|_{\text{absorption}} \tag{7.3.2}$$

where the string of equalities expresses the obvious idea again that N_1 must decrease if N_2 increases.

(c) Stimulated emission (B_{21}). This process is the reverse of absorption; the atom gives up its excess energy $h\nu$ to the field, adding coherently to the intensity. Thus the added photon is *at the same frequency, at the same phase, in same sense of polarization, and propagates in the same direction* as the wave that *induced* the atom to undergo this type of transition. Obviously, the rate depends on the density of atoms to be stimulated and the strength of the stimulating field.

$$\left.\frac{dN_2}{dt}\right|_{\substack{\text{stimulated}\\\text{emission}}} = -B_{21}N_2\rho(\nu) = -\left.\frac{dN_1}{dt}\right|_{\substack{\text{stimulated}\\\text{emission}}} \tag{7.3.3}$$

These three processes are shown in Fig. 7.5, where some artifacts have been used to emphasize various issues.

Note that "nothing" comes into Fig. 7.5(a), but a photon comes out, most likely in a direction different from the one you predict. This is spontaneous emission; that is, the atom can radiate into any of 4π steradians with any sense of polarization. The absorption

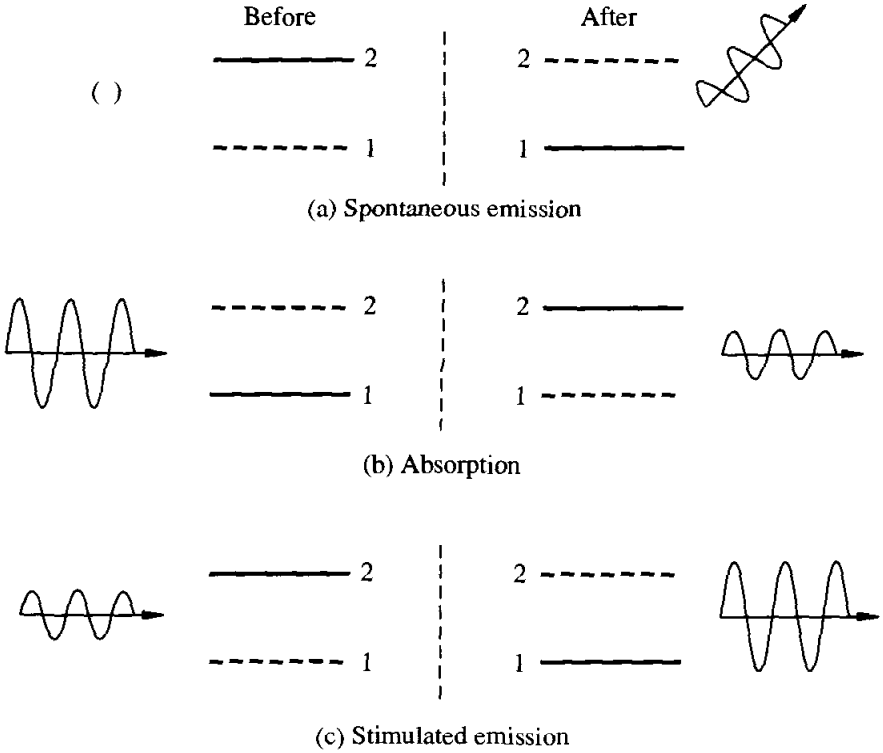


FIGURE 7.5. Effect of radiation on an atom.

should be familiar to most; that is, the wave decreases in amplitude and the atom in state 1 is converted to state 2. Obviously, that part of the wave not absorbed continues along its path. The picture for stimulated emission is just the inverse of absorption. If we accept this inverse relation, much of the "magic" about stimulated emission disappears. We would not question the fact that the transmitted wave in Fig. 7.5(b) is at the same frequency, is in the same direction, and has the same polarization as the incident wave. By identifying Fig. 7.5(c) as the inverse process to Fig. 7.5(b), we attribute those same characteristics to the portion of the wave that *adds* to the stimulating wave.

A very important point to remember from Fig. 7.5 is that stimulated emission adds a photon

1. *At the same frequency* of the stimulating wave
2. *In the same polarization* of the stimulating wave
3. *In the same direction* of travel of the stimulating wave
4. *In the same phase* of the stimulating wave

If these characteristics were not true, then we would obtain various absurdities such as violating most laws of thermodynamics (see [14]). With a more detailed view of the field-atom interaction and much more mathematics, these characteristics become obvious, but, for now, we accept those characteristics on faith.

7.3.2 Relationship Between the Coefficients

By defining these processes, Einstein was able to reproduce the blackbody formula within the framework of thermodynamic equilibrium. The sum of (7.3.1) through (7.3.3) yields the total rate of change of the population density in state 2 (or 1) as a result of the radiative processes.

At thermodynamic equilibrium, each process going "down" must be balanced exactly by that going "up." (This is the reason why we can ignore any other as yet unidentified processes affecting states 2 and 1; the detail balance must occur between the radiative processes considered here.)

$$\left.\frac{dN_2}{dt}\right|_{\text{radiative}} = -A_{21}N_2 + B_{12}N_1\rho(\nu) - B_{21}N_2\rho(\nu) = -\left.\frac{dN_1}{dt}\right|_{\text{radiative}} \tag{7.3.4}$$

At *equilibrium*, the time rate of change must be zero.

$$\frac{N_2}{N_1} = \frac{B_{12}\rho(\nu)}{A_{21} + B_{21}\rho(\nu)} \tag{7.3.5}$$

Einstein invoked classic Boltzmann statistics to provide another equation for the ratio of the two populations in states 2 and 1 and set that value equal to (7.3.5):

$$\frac{N_2}{N_1} = \frac{g_2}{g_1}e^{-h\nu/kT} = \frac{B_{12}\rho(\nu)}{A_{21} + B_{21}\rho(\nu)} \tag{7.3.6}$$

where $g_{2(1)}$ = number of ways that an atom can have the energy $E_{2(1)}$. For a simple atom, this quantity is related to the total angular momentum quantum number $J_{2(1)}$ by

$$g_{2(1)} = 2J_{2(1)} + 1 \tag{7.3.7}$$

It is very easy to verify this formula by using the hydrogen atom for a typical case. The s states have $l = 0$ and are a spherically symmetric and thus all orientations are equivalent. Hence there is only *one* wavefunction associated with the s state energy level (or 2 if we include the doubly degeneracy provided by the spin). For the p states, $l = 1$ and the wavefunction looks like a figure eight, which can be oriented along the x, y or z axis. Hence there are *three* completely different (and orthogonal) wavefunctions that have the same energy (six including spin). We must account for the coupling between L and S to derive J for a multielectron atom, but such a computation is of no concern to us since J will be considered as a given specification of the atomic system. The nuclear spin also enters the picture and contributes an additional multiplicity of $2I + 1$ allowed states. Since we will never be concerned with changes in the nuclear spin, we will often ignore the multiplicative factor of $(2I + 1)$ since it is the ratio of degeneracies that enter into our equations.

The number of states at an energy E will be encountered again in semiconductor lasers but are disguised by the words *density of states*. We need only to solve for $\rho(\nu)$ from (7.3.6), and we manipulate that solution in the following manner:

$$A_{21}\left[\frac{g_2}{g_1}e^{-h\nu/kT}\right] + B_{21}\left[\frac{g_2}{g_1}e^{-h\nu/kT}\right]\rho(\nu) = B_{12}\rho(\nu)$$

or

$$\rho(\nu) = \frac{A_{21}[(g_2/g_1)e^{-h\nu/kT}]}{B_{12} - B_{21}[(g_2/g_1)e^{-h\nu/kT}]} \tag{7.3.8}$$

After dividing by $[(g_2/g_1)\exp(-h\nu/kT)]$ and factoring B_{21} out of the denominator, we obtain

$$\rho(\nu) = \frac{A_{21}}{B_{21}} \cdot \frac{1}{\dfrac{B_{12}g_1}{B_{21}g_2}e^{h\nu/kT} - 1} \tag{7.3.9}$$

This is almost the Planck formula of (7.2.10). To make it so, Einstein forced the fit with identification of various interrelationships between the coefficients:

$$\frac{B_{12}g_1}{B_{21}g_2} = 1 \quad \text{or} \quad \boxed{g_2B_{21} = g_1B_{12}} \tag{7.3.10a}$$

and

$$\boxed{\frac{A_{21}}{B_{21}} = \frac{8\pi n^2 n_g h\nu^3}{c^3}} \tag{7.3.10b}$$

With these identifications, (7.3.9) is identical to (7.2.10).

Equations (7.3.10a) and (7.3.10b) are very important because they show a connection between three different radiative processes: spontaneous emission, absorption, and stimulated emission. If one is known, all are known. Although a particular experiment may emphasize one or another coefficient, the result may be applied to a completely different one. For instance, an absorption experiment yields vital information on the stimulated emission coefficient.

It is most important to realize that these coefficients are characteristic of the *atom*. The atom, per se, does not know (or care) whether it is in a thermodynamic equilibrium environment of a heated cavity or in the presence of an intense field (a laser) generated by other atoms. It responds according to the rates indicated by (7.3.1), (7.3.2), and (7.3.3) for electromagnetic radiation.

However, radiation is not the only thing that can affect an excited atom. The atoms can undergo a collision with another atom, an electron, or a lattice vibration (a phonon), which can also cause transitions to take place. Einstein's approach places radiation on an equal footing with these other processes and incorporates the interchange between the states in a natural and straightforward manner. The power of this approach is shown in the next section.

7.4 LINE SHAPE

Let us leave the blackbody environment and its prominent application* and go to the case where the densities of the atomic states 1 and 2 may be transient in nature (time varying) and/or be far removed from thermodynamic equilibrium. Thus the ratio N_2/N_1 will *not* be given by the Boltzmann factor, nor will the radiative processes sum to zero as was assumed in the previous section, and $\rho(\nu)$ is not given by the blackbody formula. Indeed much of our attention will be focused on systems that emit a "line" or a narrow band of frequencies whose width $\Delta\nu$ is much less than the central frequency ν, but not a continuum as is a blackbody spectrum. While we could say that the line is very sharp or that the atomic Q is very high, it is not infinite. The width $\Delta\nu$ in Hz of even a very narrow line may be comparable to many of the communication channels in use today, a fact that immediately attracted the attention of communication engineers. Thus even the simple statement that $h\nu = E_2 - E_1$ has to be viewed with a bit of skepticism. The uncertainty principle precludes an absolutely precise energy state (unless we accept complete ignorance as to when the atom was in a particular state) and thus a single frequency from a transition should also be viewed with skepticism. If we consider more than just one atom (and a single atom is not a very practical system!), then it becomes imperative to recognize that the energies are not perfectly defined.

Thus we give a bit on the specification of the energy levels states 1 and 2 in the manner shown in Fig. 7.6. We say that there is a distribution of energies around the nominal values labeled by E_0, E_1, and E_2. Thus, there can be spontaneous transitions from any parts of the manifold labeled 2 to those parts labeled 1 at the various frequencies shown there. The

*The incandescent bulb is the most common application of a system that is almost in thermodynamic equilibrium with Fig. 7.2 illustrating that our eyes respond to a very small fraction of the emission from the hot filament.

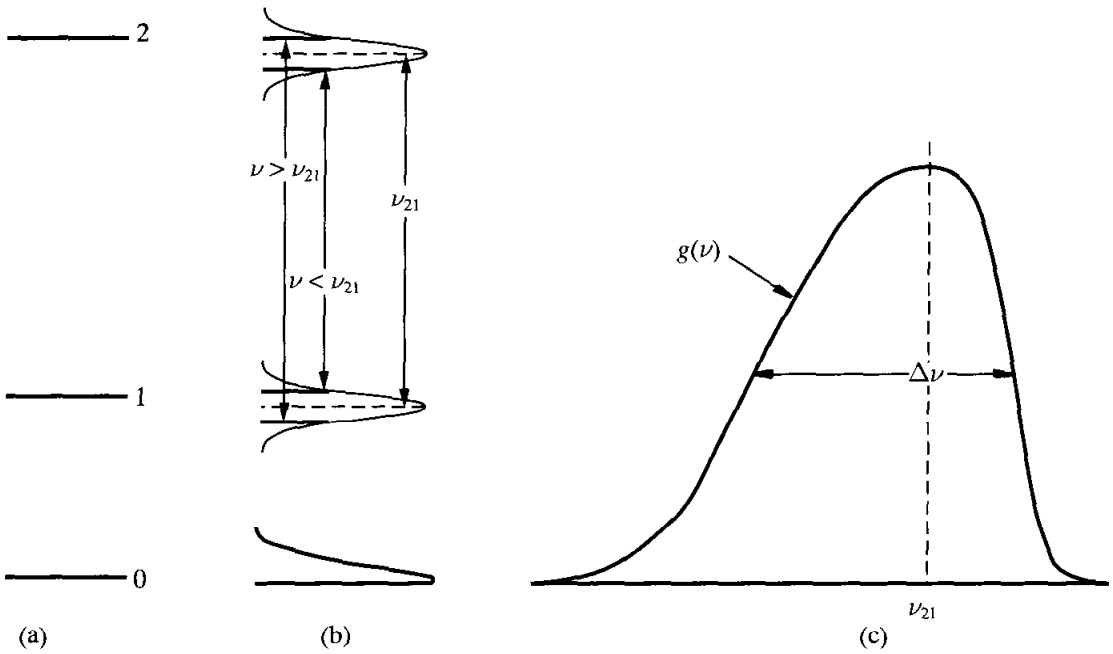


FIGURE 7.6. The evolution of the energy level diagram in (a) emitting a zero width line to broadened levels in (b) yielding the spectral line shape shown in (c).

broadening of each state can be different and may or may not be symmetrical. The important issue is that there is a distribution of photon frequencies that can be emitted spontaneously. This relative distribution is called the line-shape function with $g(\nu)d\nu$ being the fraction of that spontaneous emission being emitted into the interval $d\nu$.

Definition 1 of $g(\nu)$.

$$g(\nu')d\nu' = \text{probability that a spontaneously emitted photon will appear at a frequency between } \nu' \text{ and } \nu' + d\nu'. \tag{7.4.1a}$$

Obviously, a photon has to have some frequency and thus the integral of $g(\nu')$ over all frequencies is 1.

$$\boxed{\int_0^\infty g(\nu')\, d\nu' = 1} \tag{7.4.2}$$

The dimensions of $g(\nu)$ are (1 ÷ frequency) or seconds. A quick estimation for its peak value is that $g(\nu_0) \sim 1/\Delta\nu$, where $\nu_0 = \nu_{21}$ is the frequency of the peak and $\Delta\nu$ is the FWHM of the emission [see Fig. 7.6(c)]. The actual shape of the function $g(\nu)$ will be discussed in detail later in Sec 7.6, but it should be clear that both the broadening of the upper and lower states contribute to $g(\nu)$.

Since the actual broadening of each state in Fig. 7.6 is very small compared to its central value, the spontaneous emission *rate* from each $d\nu'$ is given by the same A_{21} coefficient, and the spectral distribution of emitted spontaneous power (into all directions) is

$$\{I(\nu)d\nu\}\,[\text{surface area}] = \{h\nu \cdot A_{21}N_2 g(\nu)d\nu\}\,[\text{volume of surface}] \tag{7.4.3}$$

Thus $g(\nu)d\nu$ can be measured if one has the instruments with sufficiently narrow bandpass or resolving power. (One of the motivations for the invention of the Fabry-Perot cavity of Chapter 6 was to examine the spectral shape of "line" emissions by using the narrow bandpass of the transmission such as that shown in Fig. 6.3.)

While this discussion emphasized the role of spontaneous emission in defining the line shape function, an equally valid and even more important definition is based upon absorption and stimulated emission.

Definition 2 of $g(\nu)$.

$g(\nu')d\nu' =$ relative strength of absorption of radiation in the interval ν' to $\nu' + d\nu'$ by atoms in state 1

Definition 3 of $g(\nu)$.

$g(\nu')d\nu' =$ relative strength of the stimulation by radiation in the interval ν' to $\nu' + d\nu'$ in inducing the atoms in state 2 to give up their internal energy

where the electromagnetic energy density in the interval ν' to $\nu'+d\nu'$ (joules/L^3) is $\rho(\nu')d\nu'$. The important point is that same line shape function $g(\nu)$ applies to all three of the radiative processes as discussed in the previous section and as defined by Einstein. (The proof that the same $g(\nu)$ applies to all three processes follows the line of reasoning used in Sec. 7.3.2. Assume different line shapes for each process, assume detailed balancing and a Boltzmann distribution of the populations, and evaluate the thermodynamic equilibrium limit for $\rho(\nu)$. The only way to reproduce the universal Planck function, which was known from experiment, is for all line shapes to be equal.)

The spectral width of radiant energy density $\rho(\nu)$ may be much wider than $g(\nu)$, as it is for the blackbody continuum, or may be much smaller than $\Delta\nu$ if $\rho(\nu)$ were provided by an external single frequency laser that might have all of its energy at one frequency and nothing at others. To account for either or any possibility, the radiative part of the rate equation for N_2 should be written as

$$\left.\frac{dN_2}{dt}\right|_{\text{radiative}} = -A_{21}N_2\left[\int_0^\infty g(\nu')\,d\nu' = 1\right] + B_{12}N_1\int_0^\infty \rho(\nu')g(\nu')\,d\nu' - B_{21}N_2\int_0^\infty \rho(\nu')g(\nu')\,d\nu' \tag{7.4.4}$$

Note that if $\rho(\nu')$ is very broad compared to $g(\nu')$, we can evaluate it at $\nu' = \nu$ and pull it outside of the integral leaving $\int g(\nu')\,d\nu' = 1$. Then (7.4.4) reproduces the original formulation of Einstein [i.e., (7.3.4)].

Most of the remainder of the book will consider the case at the opposite extreme where the spectral width of $\rho(\nu)$ is very small compared to $g(\nu)$, and we can consider all of the photons to have single frequency or $\rho(\nu)$ can be approximated by a δ function.

For $\rho(\nu') \approx \rho_\nu\delta(\nu' - \nu)$ (i.e., only one frequency)

then $$\left.\frac{dN_2}{dt}\right|_{\text{radiative}} = -A_{21}N_2 - B_{21}N_2\rho_\nu g(\nu) + B_{12}N_1\rho_\nu g(\nu) \tag{7.4.5}$$

This equation is always converted to another format by using the following substitutions:

$$B_{12} = \frac{g_2}{g_1} B_{21} \qquad \text{Recall 7.3.10a}$$

$$B_{21} = \frac{c^2}{8\pi n^2 \nu^2} \frac{c}{n_g} \frac{1}{h\nu} A_{21} = \frac{c}{n_g} \frac{\lambda_0^2}{8\pi n^2} \frac{A_{21}}{h\nu} \qquad \text{Recall 7.3.10b}$$

and

$$\rho_\nu \left(\frac{\text{joules}}{\text{volume}} \right) = \frac{I_\nu(\text{watts/area} = \text{joules} \cdot \text{time}^{-1} \cdot L^{-2})}{\text{photon velocity} = L/T}$$

or

$$\rho_\nu = \frac{I_\nu}{\text{group velocity} = c/n_g} \qquad \text{(elementary EM theory)}$$

Equation (7.4.5) assumes the following format:

$$\left.\frac{dN_2}{dt}\right|_{\text{radiative}} = -A_{21}N_2 - \left\{ A_{21} \frac{\lambda^2}{8\pi n^2} g(\nu) \right\} \frac{I_\nu}{h\nu} \left[N_2 - \frac{g_2}{g_1} N_1 \right]$$

and we abbreviate the quantity in the braces by $\sigma(\nu)$

or

$$\boxed{\left.\frac{dN_2}{dt}\right|_{\text{radiative}} = -A_{21}N_2 - \frac{\sigma(\nu) I_\nu}{h\nu} \left[N_2 - \frac{g_2}{g_1} N_1 \right]} \qquad (7.4.6)$$

A quick check indicates that the quantity σ has the dimensions of "area" (L^2) and is referred to as the stimulated emission cross section ($A_{21} \to T^{-1}$; $g(\nu) \to T$; $\lambda^2 \to L^2$; all other factors are dimensionless).

$$\boxed{\text{stimulated emission cross section} = \sigma(\nu) = A_{21} \frac{\lambda^2}{8\pi n^2} g(\nu)} \qquad (7.4.7)$$

As will become obvious in the next section, the cross section σ can be interpreted as the *cross sectional area* of the atom to the photon flux $I_\nu/h\nu$. A typical value for it is on the order of the area of an atom (10^{-16} cm^2), but can be as much as 10^{-12} cm^2 or as small as 10^{-20} cm^2. It is a very convenient number to know about a laser transition.

[NOTE: The serious student should verify the conversion of (7.4.5) into (7.4.6). The latter is the format to be used throughout the book with few exceptions.]

Thus the line shape function $g(\nu)$ is a most important parameter describing the interaction of the electromagnetic waves with the atoms. We are now poised at a decision point with two alternative paths for our progress.

1. We can proceed to the next section (Sec. 7.5) to derive the gain coefficient of the laser, which is the central equation of quantum electronics. Unfortunately, this requires a

bit of faith about the existence of the line shape function since only the rudimentary explanation as to the causes for it have been given here.

2. Thus we could jump to Sec. 7.6 where various processes are described and then return to Sec. 7.5. This provides the functional forms for $g(\nu)$, but experience indicates that the subject tends to bore the student.

In any case, the next section is so important that I recommend that it be covered twice, immediately following this one and then repeated after Sec. 7.6.

7.5 AMPLIFICATION BY AN ATOMIC SYSTEM

We are now in the position to describe the process of amplification (or attenuation) of electromagnetic energy by its interaction with the atoms. We take Einstein's view of this interaction, although we will *not* assume that the population densities (m^{-3}) of the various states N_2 and N_1 are in thermodynamic equilibrium. We will assume that these densities are being created by an external pumping process that is exactly balanced by a loss process, the details of which need not concern us at present.

In Fig. 7.7, we imagine a "Gedanken" experiment in which a slab of these atoms, Δz long, is being irradiated by a polarized electromagnetic wave of intensity $I_\nu(W/m^2)$, which, after amplification (or attenuation), is received by our detector. We must recognize, however, that a detector system *cannot* distinguish between the various physical processes discussed in this chapter: A photon radiated spontaneously by this slab and reaching the detector causes the same response as does one that has been added by stimulated emission. Obviously, a spontaneous emission from this slab contributes "noise" in our experiment.

However, if we choose the source carefully, the noise can be minimized: The optical bandwidth of the detector system is limited by some sort of a filter with a passband $\Delta\nu$ around the frequency ν of the source, a polarizer is used to reject one half of the spontaneous power

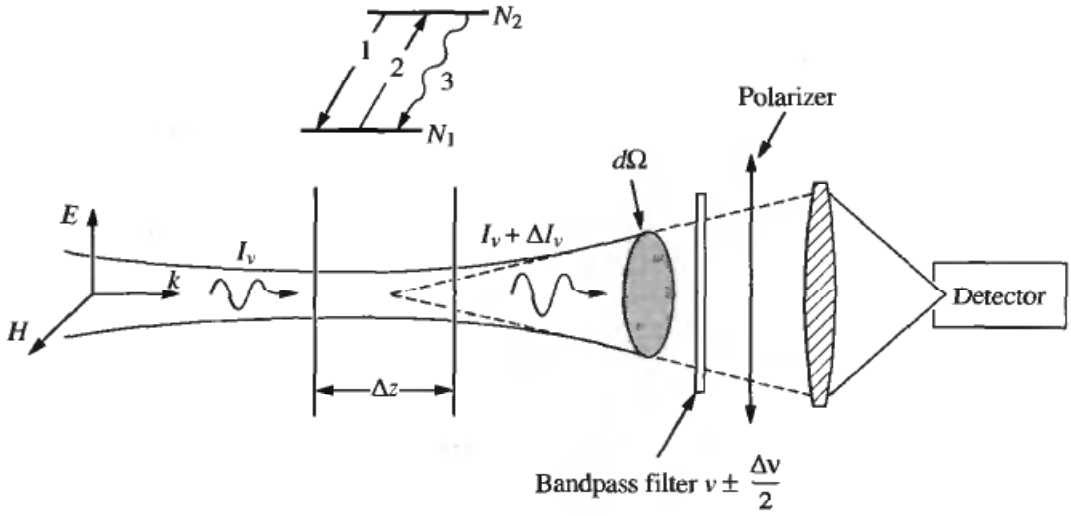


FIGURE 7.7. Measurement of the gain of an atomic system.

orthogonal to the source, and the field of view (FOV) is limited to match the incoming beam. Thus the output consists of the input intensity *plus* that added by the following processes:

1. *Stimulated emission*: the amount of radiation stimulated by the incoming wave. Since this is stimulated emission, the frequency, phase, and direction of the added signal are the same as the incoming wave (this is indicated by process 1 on the energy-level diagram)

minus:

2. *Absorption*: the amount of radiation absorbed by the atoms in state 1

plus:

3. *Spontaneous emission*: the amount of radiation emitted spontaneously by the atoms in state 2 in the direction of the input wave and in the same frequency within bandwidth $\Delta\nu$ of our detector (this is indicated by the wavy line going from state 2 to state 1).

Now the easiest way to do the bookkeeping on these processes is to arrange the physical factors into the vertical columns. We recognize that each transition caused by the foregoing processes contributes (or subtracts) a package of energy $h\nu$ (column 1), which, when multiplied by the rate per atom, is the power contributed by each atom (column 2), but we also recognize that this contribution (or interaction) scales according to the line shape (column 3). Column 4 is the probability that the photon involved has the proper polarization, and column 5 is the probability that it is within the solid angle specified by the experimental setup (which is heavily weighted favoring the first two processes).

However, some of the spontaneous power reaches the detector. The radiation from the slab is spread uniformly into 4π steradians, or equivalently into the various electromagnetic modes defined by the beam size and the length Δz. Hence, the detector, with an acceptance cone of $d\Omega$, will only collect a fraction, $d\Omega/4\pi$, of the power radiated by these atoms, and only one half of that fraction has the proper polarization. Finally, column 6 is the number of atoms involved in the interaction. The arrangement is given in (7.5.1).

$$
\begin{array}{cccccccccccc}
 & 1 & & 2 & & 3 & & 4 & & 5 & & 6 \\
\Delta I_\nu = & +h\nu & \times & B_{21}\dfrac{I_\nu}{(c/n_g)} & \times & g(\nu) & \times & 1 & \times & 1 & \times & N_2\Delta z \\
 & -h\nu & \times & B_{12}\dfrac{I_\nu}{(c/n_g)} & \times & g(\nu) & \times & 1 & \times & 1 & \times & N_1\Delta z \\
 & +h\nu & \times & A_{21}\Delta\nu & \times & g(\nu) & \times & \dfrac{1}{2} & \times & \dfrac{d\Omega}{4\pi} & \times & N_2\Delta z
\end{array}
\tag{7.5.1a}
$$

Further manipulation yields

$$
\frac{\Delta I_\nu}{\Delta z} \to \frac{dI_\nu}{dz} = \left[\frac{h\nu}{(c/n_g)}(B_{21}N_2 - B_{12}N_1)g(\nu)\right] I_\nu + \frac{1}{2}\left[h\nu A_{21}N_2 g(\nu)\Delta\nu\frac{d\Omega}{4\pi}\right] \tag{7.5.1b}
$$

We can readily appreciate that this last term can be called "noise," since this signal is present at the detector without any input I_ν to the slab. Even though it is essential to the laser (to initiate the oscillation), its presence is not required at this time and we neglect it for now.

It is convenient to use some of the relations found in Sec. 7.3 to change the appearance of (7.5.1b). Recall that $B_{12}/B_{21} = g_2/g_1$ and $A_{21}/B_{21} = 8\pi n^2 n_g h\nu^3/c^3$. Thus, the basic equation becomes

$$\boxed{\frac{dI_\nu}{dz} = \left\{\left[A_{21}\frac{\lambda^2}{8\pi n^2}g(\nu)\right]\left[N_2 - \frac{g_2}{g_1}N_1\right]I_\nu\right\} \triangleq \gamma(\nu)I_\nu} \tag{7.5.2}$$

The intensity increases with distance if all terms on the right-hand side of (7.5.2) are positive and thus the coefficient $\gamma(\nu)$ is called the gain coefficient and (7.5.2) is the central equation of laser theory and thus treated with due respect. (***Memorize it!***). If the intensity does not change the population density of states 2 and 1, then it is a "small signal," and a subscript is attached to $\gamma(\nu) \rightarrow \gamma_0(\nu)$ to indicate that fact. All lasers start with a small signal value, but eventually the intensity does affect the populations, and thus N_2 and N_1 are implicit functions of I_ν. That situation will be addressed in the next chapter.

All terms on the right are always positive other than the difference in populations, and thus

$$\boxed{N_2 > \frac{g_2}{g_1}N_1 \qquad \text{for gain}} \tag{7.5.3}$$

This is an inversion of the normal state of affairs such as that given by the Boltzmann relationship* and points out a fundamental requirement for amplification. If the population is "normal" and $N_2 < (g_2/g_1)N_1$, then dI_ν/dz is negative and (7.5.2) predicts the absorption coefficient.† Much of our immediate surroundings are in a quasinormal state of affairs, and thus absorption is the rule. Thus, (7.5.2) applies to all cases in which photons are added to or subtracted from any electromagnetic wave at any frequency.‡

The first bracket in the brace is the stimulated emission cross section, as defined by (7.4.7) and is a collection of atomic constants describing the atom (through λ and A_{21}) and its environment through $g(\nu)$. It is a most convenient specification of a fundamental parameter of the system, since one merely multiplies this "area" by the effective inversion density $N_2 - (g_2/g_1)N_1$ to obtain the gain coefficient. If, for example, the peak value of $\sigma(\nu_0)$ were 10^{-16} cm^2 (typical), then it requires a net inversion of 10^{14} cm^{-3} to obtain a gain coefficient of 0.01 cm^{-1} (or 1%/cm). Sometimes the absorption cross section is specified,

*If we use the Boltzmann factor to describe $N_2/N_1 = (g_2/g_1)\exp[-h\nu/kT]$, then the temperature T must be negative to obtain gain.

†(In the older literature, the absorption coefficient is called the "extinction" coefficient and is abbreviated by $k(\nu)$. A formula such as (7.5.2) is in books printed in 1934. See Ref. [1].)

‡The case of the "scattering" photons out of a beam does not quite fit the details of (7.5.2). However, the generic equation still applies; namely, $dI/dz = (N\sigma_{\text{scat}}) \cdot I$ where σ_{scat} would be the scattering cross section.)

which is related to the stimulated emission value by

$$\sigma_{abs}(\nu) = (g_2/g_1)\sigma_{stim}(\nu) \tag{7.5.4}$$

Unless the subscript on σ is indicated explicitly, we will always specify the stimulated value.

Having found the differential form of the gain, we can integrate (7.5.2) to obtain

$$\begin{aligned} I_\nu(z) &= I_\nu(0)\exp[\gamma_0(\nu)z] \\ &= G_0(\nu)I_\nu(0) \\ G_0 &= \exp[\gamma_0(\nu)d] \end{aligned} \tag{7.5.5}$$

where G_0 is the small signal (power) gain of an amplifier of length d.

It is most important to remember that the gain coefficient is frequency dependent; consequently, the gain G_0 is even more so, since $\gamma_0(\nu)$ appears in the exponent. Thus we have a narrow bandpass amplifier. In Chapter 8 we put this narrow band amplifier into a feedback loop to obtain oscillation.

There is another logic path that can be used to arrive at (7.5.2). It requires us to translate the differential equation definition of the gain coefficient into simple words. The gain coefficient is defined by

$$\boxed{\gamma(\nu) \triangleq \frac{1}{I_\nu}\left(\frac{dI_\nu}{dz}\right)} \tag{7.5.6}$$

which translates to

$$\gamma(\nu) \triangleq \frac{[dI_\nu/dz] = \text{net power emitted per unit of volume}}{I_\nu = \text{power per unit of area traversing that volume}}$$

Note the net power emitted per unit of volume is equal to $h\nu$ times the number of transitions, per unit of volume, per unit of time by stimulated emission minus that by absorption. The last term in (7.4.6) is that difference, hence

$$\begin{aligned} \gamma(\nu) &= \frac{1}{I_\nu}\cdot h\nu\left\{\left.\frac{dN_2}{dt}\right|_{(\text{stim}-\text{abs})} = \frac{\sigma(\nu)I_\nu}{h\nu}\left[N_2 - (g_2/g_1)N_1\right]\right\} \\ &= \sigma(\nu)\left[N_2 - \frac{g_2}{g_1}N_1\right] \end{aligned}$$

which is identical to (7.5.2). There is not any real difference in the two derivations, but it is convenient to have the different pictures in mind. In the first, we should have the mental picture provided by Fig. 7.7 where the optical beam is highlighted. In the second approach denoted by (7.5.6), we are observing the atoms' response to the optical beam. In either case, we are merely conserving total energy, the sum of that carried by the beam and that stored in the atoms.

7.6 BROADENING OF SPECTRAL LINES

There are many factors that affect the shape and width of a line and thus affect the gain coefficient. The purpose of this section is to identify the various classifications of broadening and, where possible, provide an analytic expression for $g(\nu)$. Sometimes, however, we must wave the white flag and give up on an analytic approach, because of incomplete knowledge or because of the excessive complexity of the mathematics. Then we use a *measurement* of the spectral shape of the spontaneous emission, absorption, or gain coefficient and apply the normalization condition (7.4.2) to evaluate its peak value.

All transitions have a finite width if, for no other reason, to ensure compliance with the uncertainty principle. There are two generic classifications of processes that contribute to the width of a spectral line. They are

1. Homogeneous broadening: a mechanism that applies to all atoms.
2. Inhomogeneous broadening: one that is caused by some identifiable difference between groups of atoms.

In any practical case, we usually have a mixture of both types and the combination is almost always intractable or not worth the mathematical effort for the minimal gain in precision. In such cases, we choose the dominant one and proceed.

7.6.1 Homogeneous broadening mechanisms

Lifetime broadening. We have already argued in Sec. 7.4 that a perfectly sharp energy level is impossible, and thus there is a distribution of transitions from (occupied) states near E_2 to (empty) states* near E_1 and, thus, there is a spectral width associated with the state $2 \rightarrow 1$ "line" emission. Let us consider an important special case of symmetrical level broadening that illustrates the procedure to be followed in any case.

Let $P_2(E)\,dE$ be the probability of a state existing in the energy interval E to $E + dE$ near the energy E_2 and thus $N_2 P_2(E)\,dE$ represents the number of atoms that have an energy between E and $E + dE$. A similar definition is used for the states around E_1. To make the arithmetic tractable, we assume a Lorentzian distribution of energy levels with different widths $\Delta E_{2,1}$.

$$P_{2,1}(E) = \frac{(\Delta E_{2,1})}{2\pi\left[(E - E_{2,1})^2 + (\Delta E_{2,1}/2)^2\right]} \tag{7.6.1}$$

where $\Delta E_{2,1}$ = FWHM of the energy spread around $E_{2,1}$ assuming that $\Delta E_{2,1} \ll (E_2 - E_1)$ as implied in Fig. 7.6.

The spontaneous emission of a photon of energy $h\nu$ requires that there be a (filled) band dE near E_2 *and* an (empty) band dE at exactly $h\nu$ below the first. For any given value of $h\nu$, we must sum over the various possibilities that yield the same value of $h\nu$, and thus

*For an atomic system, the words *occupied* or *empty* are superfluous since the atom is either in states 2, 1, or elsewhere. However, those words will play an important role in transitions in semiconductors (see Sec. 11.4).

the line shape reflects the "joint probability" of the parts of the two bands with intervals dE being separated by $h\nu$. Let

for band 1: $$E = x + E_1 \quad \text{and} \quad a = \Delta E_1/2 \tag{7.6.2a}$$

for band 2: $$E = x + E_1 + h\nu \quad \text{and} \quad b = \Delta E_2/2 \tag{7.6.2b}$$

and define $$\delta = h\nu - (E_2 - E_1) \tag{7.6.3}$$

Then

$$g(h\nu) = \int_{-\infty}^{+\infty} \left\{ \frac{\Delta E_1}{2\pi} \frac{1}{[x^2 + a^2]} \right\} \left\{ \frac{\Delta E_2}{2\pi} \frac{1}{[(x+\delta)^2 + b^2]} \right\} dx \tag{7.6.4}$$

where x is the energy integration variable and the two braces represent the joint probability that the two dE exist. This integral can be evaluated by the following steps:

1. Expand the integrand into partial fractions:

$$\frac{1}{x^2 + a^2} \cdot \frac{1}{(x+\delta)^2 + b^2} = \frac{Ax + B}{x^2 + a^2} + \frac{C(x+\delta) + D}{(x+\delta)^2 + b^2}$$

to find

$$B = \frac{\delta^2 + (b^2 - a^2)}{[\delta^2 + (b-a)^2][\delta^2 + (b+a)^2]} \qquad D = \frac{\delta^2 - (b^2 - a^2)}{[\delta^2 + (b-a)^2][\delta^2 + (b+a)^2]}$$

2. The terms involving A and C are antisymmetric and vanish in the integration process.
3. The integration of B and D terms yield $B\pi/a$ and $D\pi/b$, respectively. After combining these terms, the factor $\delta^2 + (b-a)^2$ cancels yielding

$$g(h\nu) = \frac{1}{2\pi} \cdot \frac{\Delta E_2 + \Delta E_1}{[h\nu - (E_2 - E_1)]^2 + [(\Delta E_2 + \Delta E_1)/2]^2} \tag{7.6.5}$$

Equation (7.6.5) demonstrates the fact that the broadening of both levels contributes to the line shape function. In terms of frequency units, we have

$$g(\nu) = \frac{\Delta\nu}{2\pi[(\nu_0 - \nu)^2 + (\Delta\nu/2)^2]} \tag{7.6.6}$$

where $$\Delta\nu = \frac{1}{2\pi}\left\{\frac{1}{\tau_2} + \frac{1}{\tau_1}\right\} \tag{7.6.7}$$

and the energy widths of the levels are related to the lifetimes by $(\Delta E_{1,2}) \cdot (\tau_{1,2}) = 1$, which is consistent with assuming a Lorentzian probability distribution for the broadened levels.

There are many causes of the finite lifetime of the various states. For instance, state 2 can decay to lower states and can transfer its stored energy to a different type of atom or

can be lost by "quenching" collisions*

$$\frac{dN_2}{dt} = \left[\sum_{j<2} A_{2j}\right] N_2 - k_2 N_2 \triangleq -\frac{N_2}{\tau_2}$$

with a similar equation applying to state 1. The sum of the A coefficients is the inverse of the radiative lifetime of state 2:

$$\frac{1}{\tau_{\text{rad}}} = \sum_{j<2} A_{2j} \tag{7.6.8a}$$

The branching ratio is defined by

$$\phi_{21} = \frac{\text{rate of decay from states 2 to 1}}{\text{sum of all decay rates from state 2}} = \frac{A_{21}}{\sum A_{2j} + k_2} \tag{7.6.8b}$$

We can never do any better than eliminate processes that quench the upper state, and hence the radiative branching ratio is usually specified with $k_2 = 0$.

Since the radiative processes are always present, there is a minimum width of any spectral line caused by the radiative rates of the two states, and that width has the unfortunate name of the *natural width* (implying by default that there is something un-natural about other widths). Equation (7.6.7) becomes

$$\Delta\nu_n = \frac{1}{2\pi}\{A_2 + A_1\} \tag{7.6.9}$$

Most practical lasers use environments leading to transitions whose width is many times that specified by (7.6.9), and, hence, natural broadening is seldom important.

Collision broadening (also known as *pressure broadening* in a gas). Probably the most important homogeneous broadening mechanism is that described by the following reaction equation:

$$[N_{2,1}] + [M] \rightarrow [N_{2,1}] + [M] \tag{7.6.10}$$

Yes, it does appear that nothing is happening since the left-hand side is the same as the right, and, no, it is not a misprint! Equation (7.6.10) describes an "elastic" collision of the atomic states (2,1) with something whose density or concentration is $[M]$. During the "collision," the energy levels $E_{2,1}$ can be considered as functions of time with some of the potential energy being converted to a kinetic variety due to the attractive or repulsive forces between the states and $[M]$. But since it is an elastic collision, the atom emerges with a final energy equal to initial value. Now the time scale for this collision is extremely short, roughly (diameter of the atoms) $\approx$ 1.0Å$\div$ [thermal velocity of the atoms $\approx 3 \times 10^4$ cm/sec] $\sim 3 \times 10^{-13}$ sec—which is much shorter than the radiative lifetime. This will be discussed in greater detail in Chapter 14. But for now, a simple picture is that such collisions interrupt

*Quenching expresses the known fact of the loss of population that proceeds at the rate $k[N]$, but the details as to where the atom goes or why may be unknown.

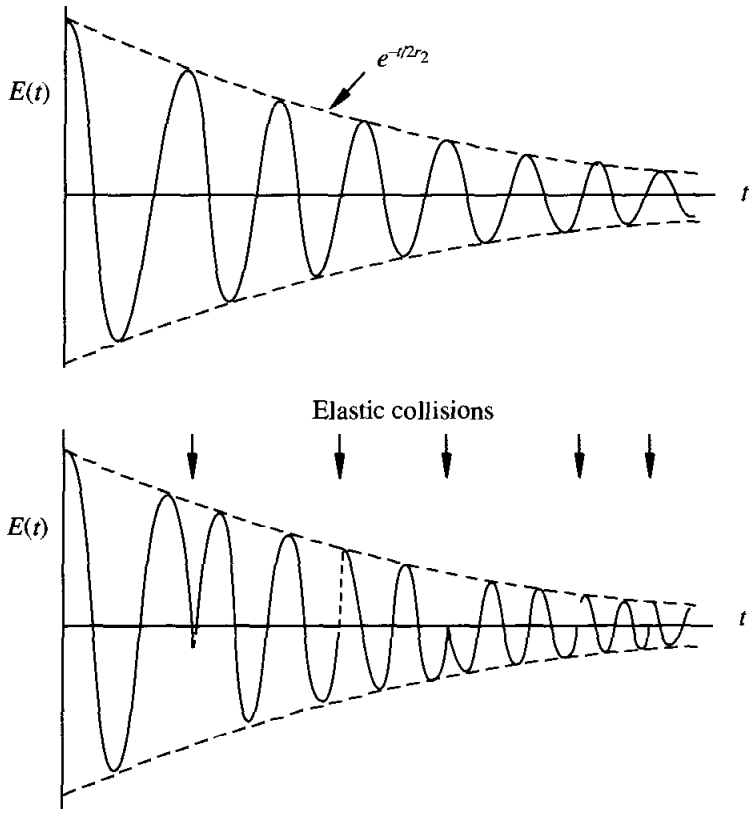


FIGURE 7.8. Classical picture of the effect of (b) elastic collisions on the emission radiation shown in (a).

the phase of the wavefunctions that, in turn affect the classical field being radiated by the group of atoms in the manner shown in Fig. 7.8.

A classical picture of this process is that shown in Fig. 7.8 (with great exaggeration). In Fig. 7.8(a), the classical field associated with the radiation from a large number of noninteracting atoms is shown. Fig. 7.8(b) is for atoms undergoing elastic collisions. The field decays at the same rate, indicative of the finite lifetime of the state; however, the phase of the radiation takes discontinuous jumps. This, in turn, shows up as a broadening in the frequency domain.

It is a rather unpleasant task in mathematics to account for the random nature of the collision rate around some mean rate ν_{col}, and then to Fourier-analyze its effect on the spectrum, but the result is somewhat transparent to a physical interpretation. Inasmuch as there are *two* quantum states involved in the transition $2 \rightarrow 1$, each atomic wave function is interrupted ν_{col} times per second; hence, the additional broadening is $2\nu_{\text{col}}$. The full width of the transition is given by

$$\Delta\nu = \frac{1}{2\pi}\left[(A_2 + k_2) + (A_1 + k_1) + 2\nu_{\text{col}}\right] \tag{7.6.11}$$

where the grouping $(A + k)$ indicates the decay rates of the states.

Whether the factor of 2 belongs in (7.6.11) is somewhat academic, since we must depend on experimental measurements to infer the pressure-broadened line width of a transition. As an example, E. T. Gerry and D. A. Leonard presented the data of Fig. 7.9 on the *absorption* coefficient of 10.6 μm radiation by CO_2 at various pressures. Without becoming unnecessarily complex about the energy levels of CO_2, it should be clear to all that the number of absorbing molecules increases as the pressure is increased. If we followed our

intuition, we would therefore expect the absorption to increase linearly with the number of molecules. The absorption does increase linearly with pressure at low pressures, but beyond about 10 to 20 torr becomes independent of pressure.

As shown by (7.5.2), the absorption coefficient is proportional to the number of absorbers (as one would expect) *and* the line-shape function $g(\nu_0)$. For a pressure-broadened line, $g(\nu_0)$ is inversely proportional to the collision rate. Hence, the product becomes independent of pressure, as shown.

The prediction of the broadening from first principles is a significant chore and beyond the scope of this book, but we should be able to use experimental results such as that given in Fig. 7.9. Usually, this contribution is specified in the form of "MHz of broadening per unit of pressure or density of gas," where a typical value might be 20 MHz/torr. Alternatively, a collision cross section might be given that expresses the size (area) of one atom, the target, to the other, the projectile (M). The collision rate is proportional to the density of projectiles, N_m, the collision cross section, and the relative velocity between the active atom and the gas broadening the transitions.

For a Maxwellian distribution specified by a temperature T, with a cross section independent of velocity, the collision rate becomes

$$\nu_{\text{col}} = N_m \sigma \left[\frac{8kT}{\pi} \left(\frac{1}{M_m} + \frac{1}{M_n} \right) \right]^{1/2} \tag{7.6.12}$$

where the M are the masses of the colliding atoms of type m and n.

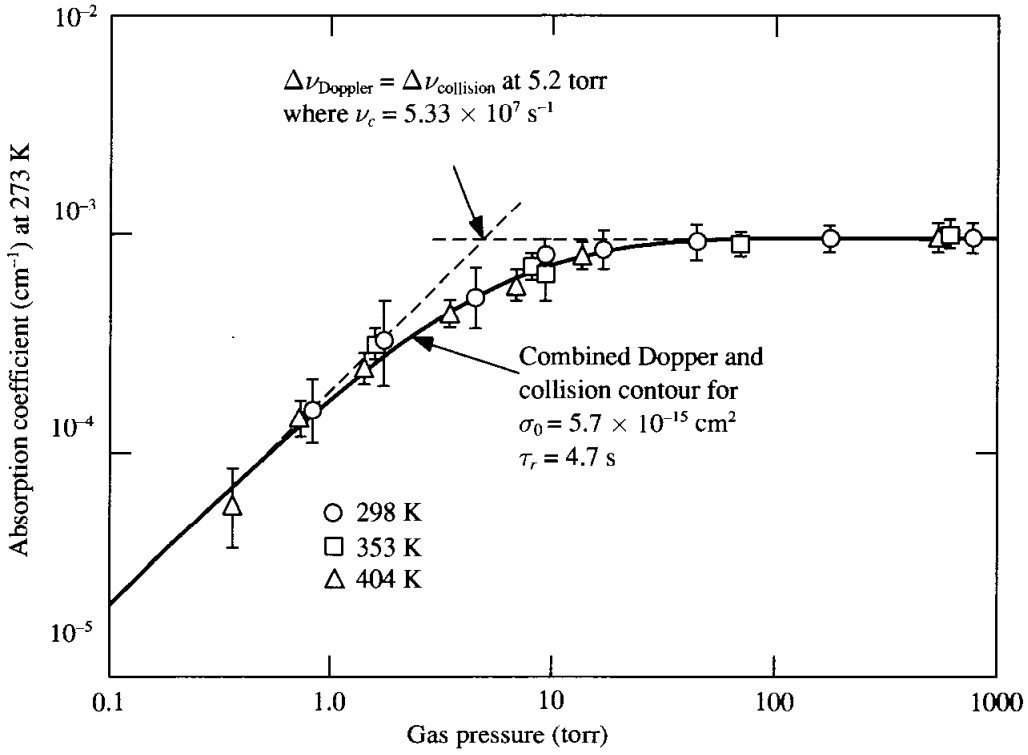


FIGURE 7.9. Absorption coefficient in CO_2 at 10.6 μm as a function of CO_2 pressure. (After E. T. Gerry and D. A. Leonard, *Appl. Phys. Lett.* 8, 227, 1966.)

Note that every atom has been treated on equal footing in this subsection. We have assumed that every atom is more or less the same as the other one, or that there was no distinguishing feature about any one group. This is characteristic of *homogeneous* broadening, and thus we add the subscript h to $\Delta\nu$.

In all the cases studied, the line shape $g(\nu)$ has the same functional form: the Lorentzian.

$$g_h(\nu) = \frac{\Delta\nu_h}{2\pi\left[(\nu_0 - \nu)^2 + (\Delta\nu_h/2)^2\right]} \tag{7.6.6}$$

where

$$\Delta\nu_h = \frac{1}{2\pi}\left[(A_2 + k_2) + (A_1 + k_1) + 2\nu_{\text{col}}\right] \tag{7.6.11}$$

In most practical cases, the last term of (7.6.11) dominates, and the width of the homogeneously broadened line becomes

$$\Delta\nu_h \simeq \frac{\Delta\nu_{\text{col}}}{\pi} = \frac{1}{\pi T_2} \tag{7.6.12}$$

where T_2 is the mean time between phase interrupting collisions. If we can distinguish between different groups of atoms under special circumstances, a different functional form of the line shape results.

7.6.2 Inhomogeneous Broadening

Although all atoms are broadened by the homogeneous process just indicated, there are "shifts" in the central frequency of same type of atoms (ie, neon) but with different characteristics (i.e., mass) that contribute to the overall line shape. These shifts are usually caused by some feature that distinguishes one group of atoms from another, such as: isotope effect, atoms of the same type but with different masses; Doppler effect, atoms of the same type and mass but different velocities; nuclear spin, coupling to the angular momentum (hyperfine splitting); Zeeman splitting, owing to magnetic fields; and Stark splitting of the lines, owing to the local crystalline electric field in a solid. More often than not, more than one mechanism is operative in addition to the homogeneous processes for a given transition. A very common example is the 2537 Å transition in Hg whose radiation excites the phosphor in the common fluorescent lamp: there are 7 isotopes, which lead to 10 lines owing to the combination of isotope effect and nuclear spin, and each of those are Doppler broadened by the random velocity of the atoms and also broadened by the homogeneous processes of the previous section. With all of these complicated processes, the 2537 Å transition is still only $\sim$ 15 GHz wide (or 0.032 Å).

Probably the simplest practical case of importance for the gas laser that emphasizes the fact that these shifts lead to an effective broadening is illustrated in Fig. 7.10. Naturally occurring neon has two isotopes with any significant abundance: 80% will have a mass of 20 AMU (atomic mass units = a multiple of $1.67252 \cdot 10^{-27}$ kg) and most of the remainder (20%) has a mass of 22. Because of the isotope effect, the center of the common "red" He:Ne laser at 6328 Å ($3s_2 \rightarrow 2p_4$) in neon changes slightly depending upon which isotope is

considered. Each contribution to the line shape scales according to the relative abundance and are symmetrical about the respective center frequencies. The composite line shape, the quantity of interest to us, is the sum of the two and is asymmetrical with its peak shifted from either center. It is obvious that the line width $\Delta\nu$ of the composite is larger than that for either one and illustrates the fact that some line shapes might be asymmetrical and a terrible arithmetic mess to describe. Nevertheless, we must learn to tolerate such a situation since the effective line-shape function appears in the laser gain equation.*

A case of considerable practical importance is due to the Doppler effect from the thermal velocity of the atoms in a gas. If a select group of atoms emit a frequency ν_0 in their rest frame which is moving towards an observer in the laboratory with a velocity v_z, then the observer would measure the Doppler-shifted frequency

$$\nu_0' = \nu_0\left(1 + \frac{v_z}{c}\right)$$

Thus the homogeneous line width radiated by that particular group identified by their velocity is

$$g(v_z, \nu) = \frac{\Delta\nu_h}{2\pi\left[(\nu - \nu_0 - \nu_0 v_z/c)^2 + (\Delta\nu_h/2)^2\right]} \tag{7.6.13}$$

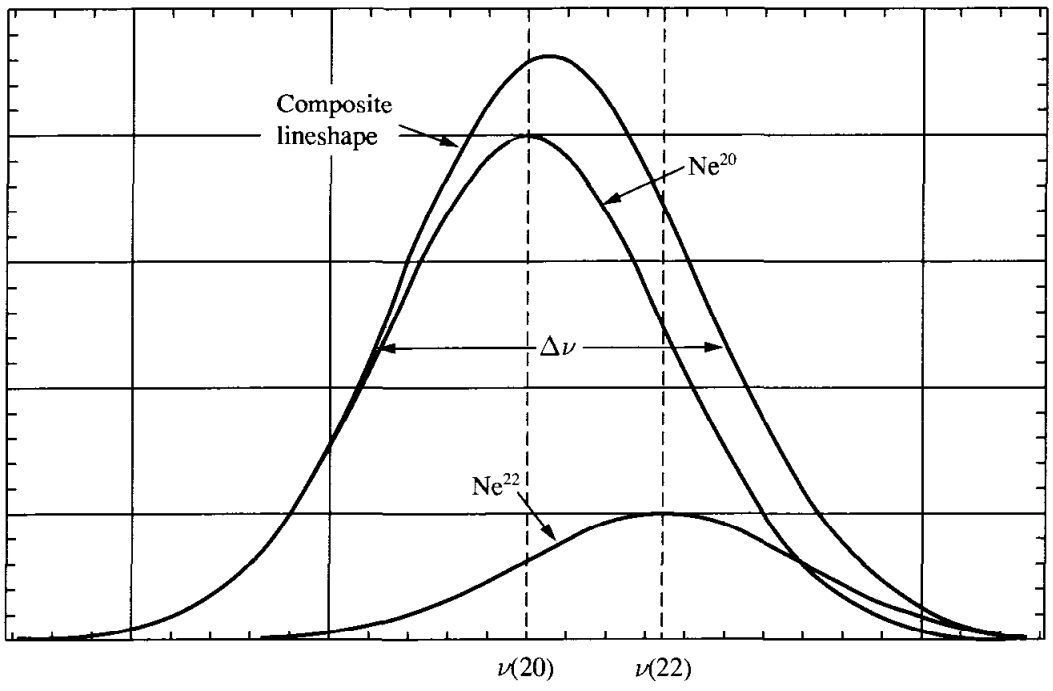


FIGURE 7.10. Inhomogeneous broadening in neon owing to the isotope effect. Each component line is symmetrically broadened owing to the Doppler effect and other homogeneous causes, but the composite line shape is slightly asymmetrical.

*(Since the gain coefficient at line center is proportional to $1 \div$ (line width), we would obtain the highest gain with a single isotope.)

Now we must multiply (7.6.13) by the probability of occurrence of a group of atoms traveling away from us at a velocity of v_z (within dv_z). This is given by the Maxwell-Boltzmann distribution function.

fraction of the atoms with velocity in the z direction with velocity between v_z and $v_z + dv_z$:

$$\frac{dN}{N} = \left(\frac{M}{2\pi kT}\right)^{1/2} \exp\left(-\frac{Mv_z^2}{2kT}\right) dv_z \tag{7.6.14}$$

By multiplying (7.6.13) by (7.6.14) and integrating over all velocities, we obtain the more realistic line-shape function for these atoms.

$$g(\nu) = \left(\frac{M}{2\pi kT}\right)^{1/2} \int_{-\infty}^{+\infty} \left\{ \frac{\Delta\nu_h}{2\pi\left[(\nu - \nu_0 - \nu_0 v_z/c)^2 + (\Delta\nu_h/2)^2\right]} \right\} \times \exp\left(-\frac{Mv_z^2}{2kT}\right) dv_z \tag{7.6.15}$$

This function has been tabulated and is called the Voigt function. We usually resort to the approximation of replacing the Lorentzian, the quantity in the braces, by a delta function. (We will see later when this approximation is valid.) That is,

$$L(x - x') = \frac{\Delta x}{2\pi[(x - x')^2 + (\Delta x/2)^2]} \to \delta(x - x')$$

Thus (7.5.15) can be integrated easily:

$$g(\nu) = \left(\frac{M}{2\pi kT}\right)^{1/2} \int_{-\infty}^{+\infty} \delta\left(\nu - \nu_0 - \frac{\nu_0 v_z}{c}\right) \exp\left(-\frac{Mv_z^2}{2kT}\right) d\left(\frac{v_z \nu_0}{c}\right) \frac{c}{\nu_0}$$

or

$$g(\nu) = \frac{1}{\nu_0}\left(\frac{Mc^2}{2\pi kT}\right)^{1/2} \exp\left[\left(-\frac{Mc^2}{2kT}\right)\left(\frac{\nu - \nu_0}{\nu_0}\right)^2\right] \tag{7.6.16}$$

This is the "pure" Doppler line shape. The peak value occurs at ν_0 (naturally) and falls to one half of that at frequencies given by

$$\frac{Mc^2}{2kT}\left(\frac{\nu_{+,-} - \nu_0}{\nu_0}\right)^2 = \ln 2$$

Thus the full width of the transition is given by

$$(\nu_+ - \nu_-) = \Delta\nu_D = \left(\frac{8kT \ln 2}{Mc^2}\right)^{1/2} \nu_0 \tag{7.6.17}$$

and (7.5.16) can be reexpressed in the following manner:

$$g(\nu) = \left(\frac{4 \ln 2}{\pi}\right)^{1/2} \frac{1}{\Delta\nu_D} \exp\left[-4 \ln 2\left(\frac{\nu - \nu_0}{\Delta\nu_D}\right)^2\right] \tag{7.6.18}$$

We can now evaluate the validity of our approximation of replacing the Lorentzian by a delta function. If the Doppler width (7.6.17) is much larger than the homogeneous width, (7.6.18) is valid. If the reverse is true, we should replace the exponential term in (7.6.15) by a delta function centered at $v_z = 0$ and thus recover the homogeneous line shape.

An electronic (i.e., visible) transition in a low-pressure gas tends to be dominated by inhomogeneous or Doppler broadening. If, however, the pressure is high enough, or the center frequency ν_0 is low enough, homogeneous or pressure broadening will dominate.

The Stark effect is very important for solid state lasers in which the active atoms, say a rare earth atom such as neodymium (Nd) or erbium (Er), is doped into a solid, say crystalline YAG or silica (SiO_2 = glass). Usually, the local electric field experienced by the active atoms is so strong that the levels split into distinct and separate parts. For instance, the ground state of Nd in YAG has the name $^4I_{9/2}$ and, due to the crystalline field, splits into five distinct levels with the group called the $^4I_{9/2}$ manifold.* This is shown in Fig. 7.11. This major effect is not a broadening; rather, it is a (Stark) splitting. We will assume that some one has measured this shift, and thus we have the correct value for the center of each level.

Because of the lattice vibrations (phonons) and/or the slight inhomogeneities in the composition of the host, the local electric field experienced by each atom may not be the same from site to site, and thus the center of each level will reflect the statistical distribution of the local electric field. Since the phonons affect all atoms, it is a homogeneous broadening mechanism, whereas the differences between different sites is an inhomogeneous one.

In most cases, we must resort to experiment to ascertain the line width and, more often than not, use a Lorentzian to describe $g(\nu)$ even though the dominant broadening process may be inhomogeneous.

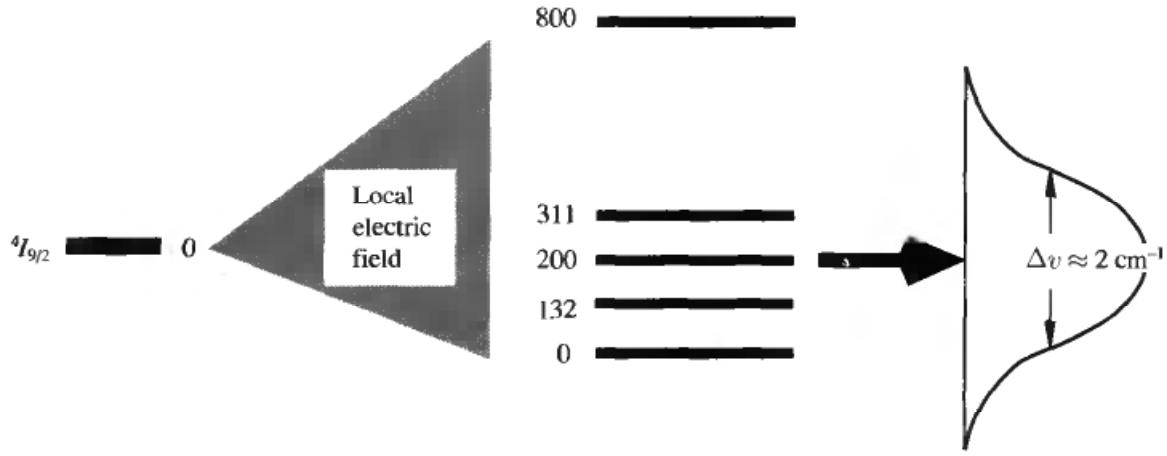


FIGURE 7.11. The Stark splitting of the $^4I_{9/2}$ level of neodymium in YAG (After Kaminski [17]). The numerical values refer to the energies of the levels in cm^{-1}. More detail will be discussed in Chapter 10.

*A short primer on the meaning of the spectroscopic names, such as $^4I_{9/2}$, is given in Chapter 15, with more detail on the Nd system in Chapter 10. For now, these are names and additional information is not needed.

7.6.3 General Comments on the Line Shape

The reader should be cautioned against assuming a direct relationship between the amount of mathematics expended here on a broadening mechanism and its relative importance. Although Doppler and pressure-broadening mechanisms are important, they do not overwhelm all other types (indeed, they do not even apply in a solid). In fact, only the *central portion* of some transitions in a gas is adequately described by the theory presented here.

However, the idea of a line shape is most important, quite general, and independent of the maze of mathematics surrounding its development. The *line-shape function*, $g(\nu)\,d\nu$, is the relative probability that

1. A photon emitted by a *spontaneous* transition will appear between ν and $\nu + d\nu$.
2. Radiation in the frequency interval ν to $\nu + d\nu$ can be *absorbed* by atoms in state 1.
3. Radiation in this interval will *stimulate* atoms in state 2 to give up their internal energy.

Obviously, the first applies to spontaneous emission, the second to absorption, and the third to stimulated emission. However, the same line-shape function applies to all three processes.

Many of the real-life line-shape functions are asymmetric and mathematically intractable. However, the atoms have no knowledge of and no trouble with *our* arithmetic. In response, we must be prepared to tolerate and use a real-life line-shape function about which we have imperfect information.

7.7 REVIEW

Before we proceed to laser oscillation, it is appropriate to stand back and review the material in this chapter. Although a lot of arithmetic was used, most of it was used to convert one equation into another form. Consequently, definitions are an important part of this chapter. If we remember the physical meaning of the quantities defined, the few lines of arithmetic follow most naturally.

For instance, the Einstein A and B coefficients were introduced to describe the *rate* at which the atoms respond to the electromagnetic field. In covering this, we are naturally led to the physical interpretation of A and B and then to the idea of *line shape*. This was the probability of a group of atoms interacting with a wave by stimulated emission or absorption, or in producing the wave by spontaneous emission. We also observed that there are different types of mechanism leading to different types of line shapes, or broadened transitions. If there is no distinguishing feature about the atoms, we have homogeneous broadening, whereas if two (or more) groups can be identified, we have inhomogeneous broadening.

All of these concepts will play a major role in laser oscillation.

PROBLEMS

7.1. The derivation of the formula for the electromagnetic mode density assume that the cavity dimensions are very large and $N(\nu)$ was found to be $N(\nu)$ = number of electromagnetic modes between 0 and ν (per unit of volume of a cavity)

$$N(\nu) = \frac{8\pi\nu^3}{3(c/n)^3} \text{ where } n = \text{index of refraction}$$

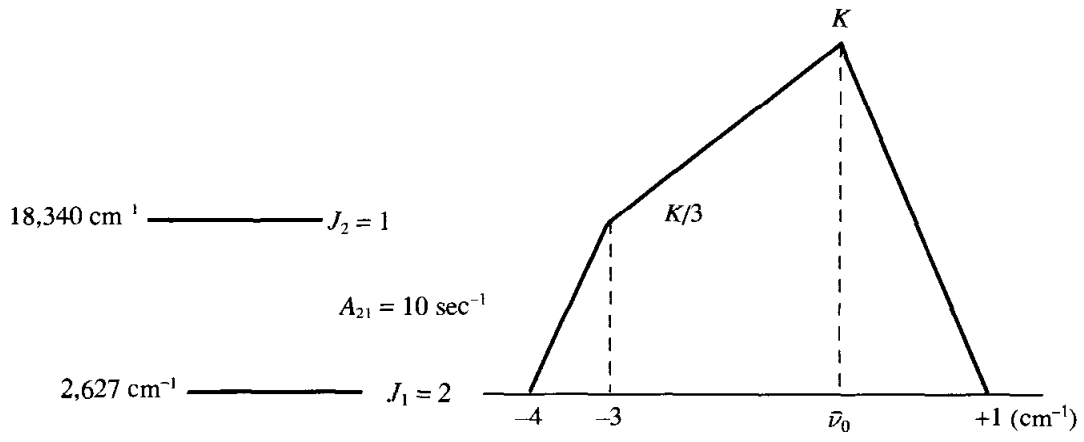
Hence, the mode density (per unit of volume per unit of frequency) was found to be

$$n(\nu) = \frac{dN}{d\nu} = \frac{8\pi n^2 n_g \nu^2}{c^3}$$

When the dimensions of the cavity become comparable to a wavelength, the approximation is *not* accurate and one must count the modes. This problem is intended to give you confidence in the formula above and also to indicate the exact procedure. Plot the number of modes between 0 and 10 GHz as a function of frequency for a rectangular cavity of dimensions $2 \times 5 \times 6$ cm using the approximate formula given above and by actually counting the allowed modes and plotting the resulting stair-step function. (CAUTION: Only the m or p index of the $TE_{m,p,q}$ mode may be zero, but not both, and only the q index may be zero for the TM mode.)

7.2. Show that the ratio of stimulated emission to spontaneous emission is equal to the number of photons per mode.

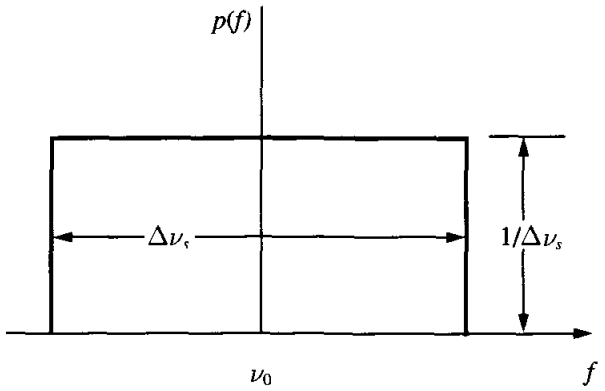
7.3. The spontaneous emission profile from a certain transition can be approximated by the shape shown below.



(a) What is the stimulated emission cross section?

(b) What is the absorption cross section?

7.4. Suppose the distribution of center frequencies is a "square" function for those values between $\nu_0 - (\Delta\nu_s)/2 < f < \nu_0 + (\Delta\nu_s)/2$ and zero, otherwise in the manner shown on the following diagram.



Each group of atoms, $p(f)\,df$, is characterized by a Lorentzian with a width $\Delta\nu_h$ (i.e., a homogeneous width).

(a) Reformulate Eq. 7.6.15 for this case, and derive an analytic expression for the combined line shape.

(b) Plot the lineshape as a function of frequency for a ratio $\Delta\nu_h/\Delta\nu_s = 0.01, 0.1$ and 0.5.

7.5. Consider a transition of 5000 Å with a width of 1 Å, a cavity of 2 cm^3 in volume and let $n = 1$.

(a) Convert this wavelength interval (1 Å) to frequency units (i.e., GHz and cm^{-1}).

(b) How many electromagnetic modes exist in this frequency band for this cavity?

(c) Suppose that the cavity were in the form of a cylinder with a cross-sectional area of 0.1 cm^2 (and thus is 20 cm long). How many $TEM_{0,0,q}$ cavity modes would fit within the frequency band specified by this 1 Å? (Do not forget the two polarizations.)

(d) Combine the results of (b) and (c) to estimate the probability of a spontaneous photon appearing in one of the polarized $TEM_{0,0,q}$ modes.

(e) If the A coefficient for this transition is 10^7 sec^{-1}, what is the stimulated emission cross section?

7.6. Evaluate the Doppler widths of the following helium-neon transition: (a) $3s_2 - 2p_4$ at $\lambda_0 = 6328$ Å; (b) $2s_2 - 2p_4$ at 1.1523 μ m; and (c) $3s_2 - 3p_4$ at $\lambda_0 = 3.39$ μ m. Express your answers in GHz and in Å units.

7.7. Note that many of the helium-neon laser transitions share a common upper level (the first-named state) or a common lower level (the second) in Problem 7.6. Use the specifications of the wavelengths to construct a partial energy-level diagram. (NOTE: The answer is given in Chapter 10.)

7.8. Reformulate (7.6.15) using the following substitutions:

$$\omega = \frac{2(\nu - \nu_0)}{\Delta\nu_D}(\ln 2)^{1/2} \qquad a = \frac{\Delta\nu_h}{\Delta\nu_D}(\ln 2)^{1/2} \qquad y = \frac{2v_z}{c}\nu_0\frac{(\ln 2)^{1/2}}{\Delta\nu_D}$$

(a) Show that (7.6.15) becomes

$$g(\omega) = \left[\left(\frac{4\ln 2}{\pi}\right)^{1/2} \frac{1}{\Delta\nu_D}\right] \tilde{g}(\omega) \text{ where } \tilde{g}(\omega) = \frac{a}{\pi}\int_{-\infty}^{\infty} \frac{e^{-y^2}}{a^2 + (\omega - y)^2}\, dy \quad \text{(P7.8)}$$

(b) Evaluate this integral numerically for $a = 1$:

Ans:

ω	0	1	2	3	4	6	8	10
$\tilde{g}(\omega)$	0.428	0.305	0.140	0.066	0.037	0.016	0.009	0.005

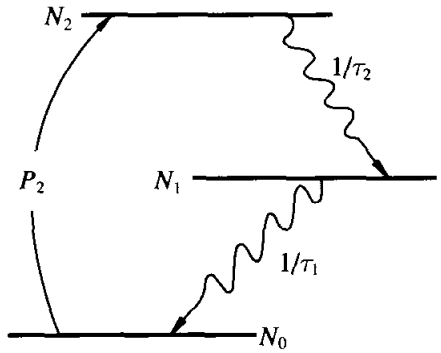
(c) Find an analytic expression for $\tilde{g}(\omega)$ for small values of a:

$$\left[\text{Ans.:}\quad \tilde{g}(\omega) = \exp(-\omega^2) - \frac{2a}{\pi^{1/2}}\left(1 - 2\omega e^{-\omega^2}\int_0^{\omega} e^{x^2}\, dx\right)\right]$$

(d) Show that (P7.8) reduces to a Lorentzian for large values of a.

7.9. On a single sheet of three-cycle semilog graph paper, plot the combined Doppler and homogeneous line shape for the numerical values given in the previous problem (i.e., $a = 1$). The linear frequency scale should be normalized to the Doppler width; that is, $(\nu - \nu_0)/\Delta\nu_D$ and the line shape plotted on the logarithmic axis. On this same graph, plot the "pure" Doppler line shape (i.e., $a = 0$). What is the combined width (expressed as a numerical factor times the Doppler width)?

7.10. In the diagram shown below, the pump P_2 (i.e., electrons, flash lamp, another laser, etc.) excites atoms from state 0 to state 2, nothing to state 1. To make the problem simple and tractable, assume state 0 is not depleted to any significant extent for any time (i.e., $dN_0/dt = 0$); use the simple decay route indicated; neglect stimulated emission; assume $\tau_2 = 1\ \mu s$ and $\tau_1 = 2\ \mu s$; and let $P_2 = 10^{20}\ \text{cm}^{-3}\ \text{s}^{-1}$. Use symbols for (a) and (b) and find numerical values for (c) and (d).

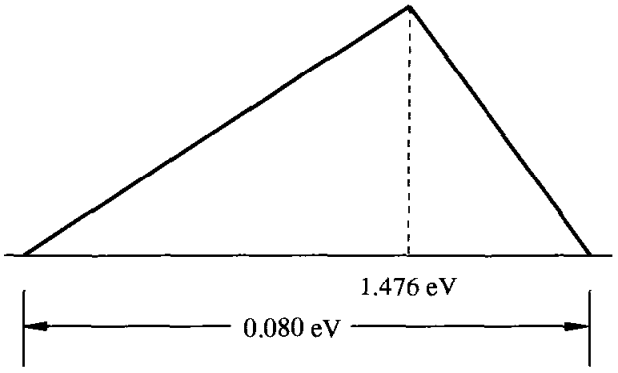


(a) What are the rate equations for states 2 and 1?
(b) Give an expression for the densities $N_{2,1}$ as a function of time.
(c) Over what time interval, δt, is the population difference $N_2 - N_1 > 0$?
(d) What are the steady-state populations in 2 and 1?

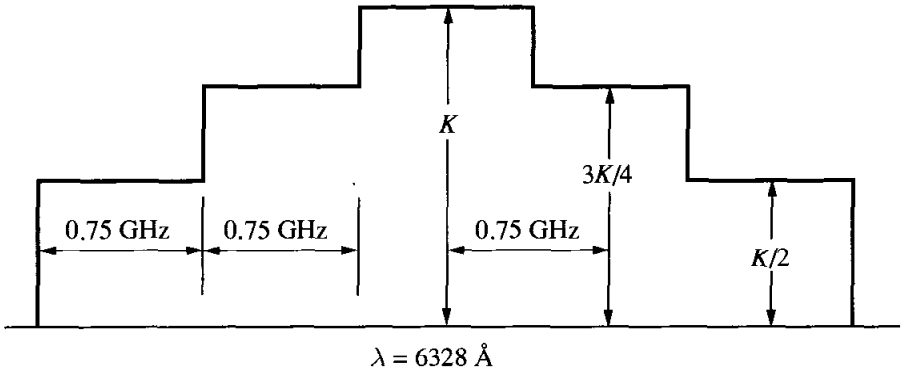
7.11. The spontaneous emission profile of a certain laser can be approximated by the triangular shape shown below. If the spontaneous lifetime were 5 nsec and the gain coefficient were 10 cm^{-1}, find

(a) The value of the line shape (in sec) at $h\nu/e = 1.476$ eV

(b) The inversion necessary to obtain that gain coefficient



7.12. The following is a tractable representation of the line shape for the helium/neon transition at $\lambda_0 = 6328$ Å ($3s_2 - 2p_4$), which has an A coefficient of 6.56×10^6 sec^{-1}.



(a) What is the value of $g(\nu_0)$?

(b) What is the stimulated emission cross section?

(c) Give a short word description of the physical significance of $g(\nu)$ as it applies to spontaneous emission, absorption, and stimulated emission.

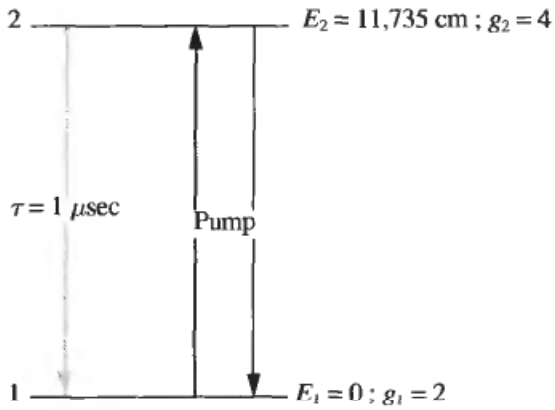
7.13. Consider the atomic system shown below being irradiated by an external wave tuned to the center of the $2 \rightarrow 1$ transition with 1 being the ground state. The wave pumps the atoms from 1 to 2 and also stimulates the atoms back to 1 from 2. In addition, the atoms in state 2 decay back to 1 by spontaneous emission and/or by other processes with a rate given by $(\tau)^{-1}$. The total density of atoms is $[N]$. Assume $\sigma = 10^{-14}$ cm^2

(a) Formulate the rate equations for the two states in terms of the intensity of the external wave, the stimulated emission cross section, the frequency $h\nu = E_2 - E_1$, τ, and the degeneracies of the states (g_2, g_1).

(b) What would be the population *ratio* N_2/N_1 if the intensity of the external wave were infinite?

(c) What must be the intensity to make the population ratio N_2/N_1 equal to 1/2?

(d) If the ambient temperature were such that $kT = 208\ \text{cm}^{-1}$ and the intensity were zero, what is the steady state population ratio N_2/N_1?



7.14. Let us take another "wrong" approach to the derivation of the blackbody formula. Assume $g_1 = g_2, n = 1$.

(a) Neglect stimulated emission in (7.3.4) and solve for the energy density $\rho(\nu)$ using the Boltzmann relation (7.3.6) for $N_2/N_1 = \exp[-h\nu/kT]$ (known to be true from experiment).

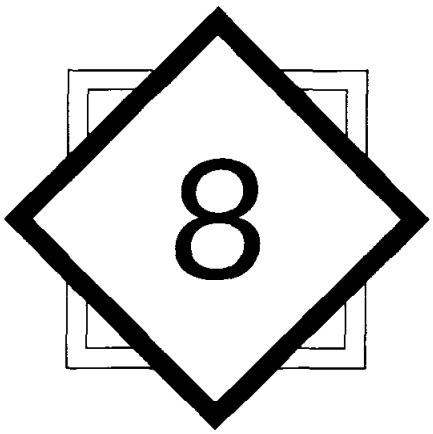
(b) The energy density $\rho(\nu)$ (7.2.1) and even Planck's constant were known before Planck devised the theory. Hence we are free to match the theory from part (a) for $\rho(\nu)$ to the experiment as expressed by (7.2.1). Assume $h\nu/kT \gg 1$, evaluate the ratio A_{21}/B_{21}, and use this result to obtain Wein's distribution analogous to the Rayleigh–Jeans one (7.2.5).

(c) Compare the Rayleigh–Jeans and Wein formulas for $\rho(\nu)$ to the correct Planck one. Indicate the region of $h\nu/kT$ where the two incorrect ones are "asymptotic" to the Planck formula.

REFERENCES AND SUGGESTED READINGS

1. A. C. G. Mitchell and M. W. Zemansky, *Resonance Radiation and Excited Atoms* (New York: Cambridge University Press, 1971). Chapter 3 is especially germane to the material of this chapter and is highly recommended. Please note: This book was first printed in 1934 and yet the material is pertinent to modern quantum electronic devices. Indeed this book is one of the most quoted references in gas laser theory. The laser gain equation is not new.
2. G. Herzberg, *Atomic Spectra and Atomic Structure*, (Englewood Cliffs, N.J.: Prentice Hall, 1937; New York: Dover, 1944). An excellent introduction to atomic spectra.
3. G. Herzberg, *Molecular Spectra and Molecular Structure*, Vol. 1: *Spectra of Diatomic Molecules* (Princeton, N.J.: D. Van Nostrand, 1967).
4. A. E. Siegman, *Introduction to Lasers and Masers* (New York: McGraw–Hill, 1971), Chap. 3.

5. A. Yariv, *Introduction to Optical Electronics*, 2nd ed. (New York: Holt, Rinehart and Winston, 1976).
6. R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynmann Lectures on Physics*, Vol. 3 (Reading Mass.: Addison–Wesley, 1965).
7. W. S. C. Chang, *Principles of Quantum Electronics* (Reading, Mass.: Addison–Wesley, 1969), Chap. 5.
8. A. Maitland and M. H. Dunn, *Laser Physics* (Amsterdam: North–Holland, 1969), Chaps. 2 and 3.
9. R. M. Eisberg, *Fundamentals of Modern Physics* (New York: John Wiley & Sons, 1961), Chaps. 2 and 13, 468-471.
10. A. E. Siegman, *Lasers* (Mill Valley, Calif.: University Science Books, 1986), Chaps. 5-6.
11. A. Yariv, *Quantum Electronics* (New York: John Wiley & Sons, 1975).
12. G. H. B. Thompson, *Physics of Semiconductor Laser Devices* (New York: John Wiley & Sons, 1980).
13. See the Historical Paper "On the Quantum Theory of Radiation," A. Einstein, reprinted from *The Old Quantum Theory* (New York: Pergamon Press, 1967) in *Laser Theory*, Ed. Frank Barnes (New York: IEEE Press, 1972). There are many other papers of interest in this collection.
14. L. Oster, "Some Applications of Detailed Balancing," *Am. J. Phys. 38*, 754–761, 1970.
15. J. H. Van Vleck and D. L. Huber, "Absorption, Emission and Line Breadths: A Semi-Historical Perspective," *Rev. Mod. Phys. 49*, 939–959, 1977.
16. R. G. Breene, Jr., "Line Shape," *Rev. Mod. Phys. 29*, 94–143, 1957.
17. A. A. Kaminiski, *Laser Crystals, Springer Ser. in Opt. Sci.* (New York: Springer–Verlag, 1981).



Laser Oscillation and Amplification

8.1 INTRODUCTION: THRESHOLD CONDITION FOR OSCILLATION

It should be obvious from the laser-gain equation

$$\gamma_0(\nu) = A_{21} \frac{\lambda_0^2}{8\pi n^2} g(\nu) \left(N_2 - \frac{g_2}{g_1} N_1 \right) \tag{7.5.2}$$

that it is necessary to have $N_2 > (g_2/g_1)N_1$ in order to have gain. Inasmuch as this is an "abnormal" state of affairs in nature, the population densities are said to be inverted. We will not be concerned here with the specific details as to how this condition is created but with the consequences.

It is obvious that we can construct a narrow-band amplifier. If the length of the medium is l_g, the small-signal power gain is

$$G_0 = \exp\left[\gamma_0(\nu) l_g\right] \tag{8.1.1}$$

To make this into an oscillator, we provide sufficient feedback in the manner illustrated in Fig. 8.1. Threshold for oscillation is determined by the requirement that the round-trip gain exceed 1. If one considers only mirror losses, this implies that the gain coefficient $\gamma_0(\nu)$ must be sufficiently large, so that

$$R_1 R_2 e^{2\gamma_0(\nu) l_g} \geq 1$$

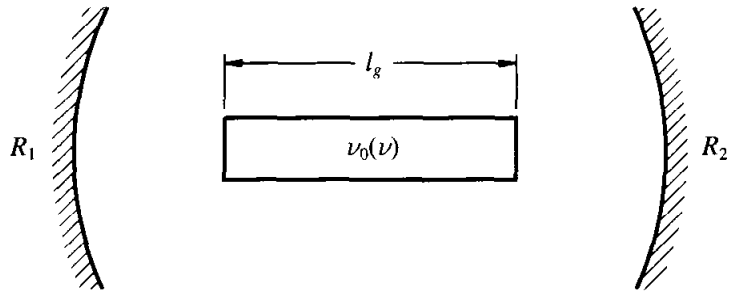


FIGURE 8.1. A simple laser.

or

$$\gamma_0(\nu) \geq \frac{1}{2l_g} \ln\left(\frac{1}{R_1 R_2}\right) = \alpha \tag{8.1.2}$$

In other words, the gain per unit of length, $\gamma_0(\nu)$, must exceed the loss when that loss is prorated on a per unit of length basis.*

There is a considerable amount of physics buried in such a simple equation, so much so that it is appropriate to consider a graphical solution, as shown in Fig. 8.2.

It should be obvious from Fig. 8.2 that "threshold" refers to a situation where the gain coefficient exceeds the loss over a very small band of frequencies. If the gain is made much larger than threshold, there is a considerable band of frequencies over which the inequality (8.1.2) is satisfied.

Which frequency oscillates? What limits the amplitude of oscillation? What physical mechanism starts the laser oscillating?

The remaining sections of this chapter address these questions together with others, as they naturally arise.

8.2 LASER OSCILLATION AND AMPLIFICATION IN A HOMOGENEOUS BROADENED TRANSITION

Figure 8.2 implies that there is a considerable band of frequencies over which the gain exceeds the loss. This is a very practical—maybe even desirable—situation. For instance, suppose that we are dealing with a homogeneously broadened transition with a width of 1 GHz and that the small-signal gain coefficient at line center ν_0 is N times the loss.

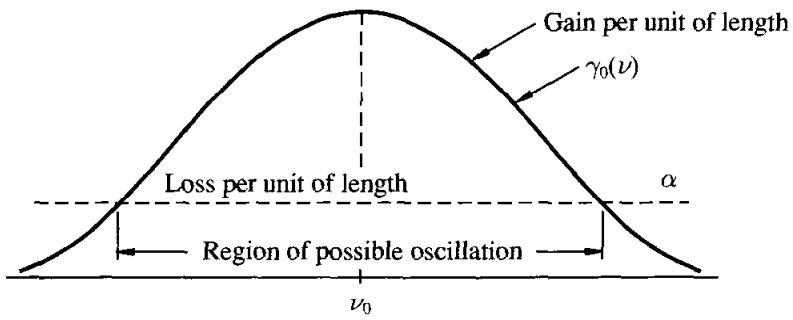


FIGURE 8.2. Graphical solution of the threshold equation.

*The symbol α will be used to denote a loss per unit of length.

Now the frequency dependence of $\gamma_0(\nu)$ is contained in the line shape, and for a Lorentzian, we have

$$\gamma_0(\nu) = \gamma_0(\nu_0) \frac{(\Delta\nu/2)^2}{(\nu_0 - \nu)^2 + (\Delta\nu/2)^2} \tag{8.2.1}$$

For $\gamma_0(\nu_0)/\alpha = N$, the gain exceeds the losses over the frequency interval:

$$2|\nu_0 - \nu| \leq (N - 1)^{1/2}\Delta\nu \tag{8.2.2}$$

For $\gamma_0(\nu_0)/\alpha = 4$, for instance, and $\Delta\nu = 1$ GHz, there is a band 1.7 GHz wide where laser oscillation *can* take place. However, laser oscillation *will* occur at a discrete frequency* dictated by the cavity mode that has the highest gain-to-loss ratio.

To appreciate the logic of this last statement, recall the basic equation associated with stimulated emission:

$$\left.\frac{dN_2}{dt}\right|_{\substack{\text{stimulated}\\ \text{emission}}} = -B_{21}N_2\rho_\nu g(\nu) = -\frac{\sigma(\nu)I_\nu}{h\nu}N_2 \tag{7.4.5}$$

Recall that the coefficient B_{21} represents an integral part of the atom. Just because we have decided to build a laser, that atom is not going to change its characteristics. The line-shape function $g(\nu)$ expresses how each atom, on the average, responds to an electromagnetic wave at various frequencies. Although an intense field can affect the line shape, only the characteristic broadening mechanisms affect its shape before oscillation starts. Thus although $g(\nu)$ expresses a preference for frequencies close to line center, it is a mild one, changing only by a factor of 2 for $|\nu - \nu_0| = \Delta\nu/2$. Although we would prefer N_2 to be large, its maximum value is fixed by the external pumping mechanism.

Is there anything in (7.4.5) that makes a sharp distinction as to which frequency or narrow band of frequencies stimulates the atom at the greatest rate? By the process of elimination, we are left with I_ν. Is there anything we can do to make any one frequency more favorable than another? If you recall the discussion on resonance from Chapter 6, you know the answer is "yes."

Recall that the field at a cavity resonance is much larger than those at, say, antiresonance, (see Fig. 6.3). Thus, we can construct the following scenario for the start-up of laser oscillation. We assume that the pumping agent has created the population inversion $N_2 - (g_2/g_1N_1)$. Even though an inversion exists, spontaneous emission still occurs, spewing out electromagnetic energy into any one of the $(8\pi n^3\nu^2\Delta\nu/c^3)V$ modes that are present in the volume of the active medium.

A few numerics are in order here. The number of modes that receive this spontaneous emission is just huge. For instance, suppose that the center wavelength of the transition is 5000 Å ($\nu_0 = 6 \times 10^{14}$ Hz). The width is, as before, 1 GHz; the volume of the active medium is 10 cm^3; and $n = 1$. Then there are $(8\pi\nu^2\Delta\nu/c^3)V = 3.35 \times 10^{10}$ different modes to which an atom can give up its internal energy to this electromagnetic field.

But most of these modes represent waves that are going in the wrong direction and not toward the mirrors but out the side of the laser cell. *But* that part of the spontaneous

*We postpone until later issues such as spatial hole burning and the spectral width of the oscillation.

emission that is in the proper frequency interval to coincide with a cavity resonance (and, of course, is along the axis of the laser) is bounced back and forth between the mirrors, greatly enhancing the standing-wave field.

Thus the initial frequency dependence of the electromagnetic energy density is governed by the frequency dependence of the spontaneous emission [i.e., $g(\nu)$] and by the cavity response. This situation is shown in Fig. 8.3(a), where the fields in the cavity modes are just beginning to be formed by spontaneous emission. Once the energy is present, transitions caused by stimulated emission can take place, adding energy in the proper phase, at the proper frequency, in the proper direction, and in the proper polarization so as to add coherently to the field that stimulated the atoms. Consequently, those resonant fields close to line center are amplified, and, in a few, cavity transit times are much greater than their initial values. Those modes in the wings $(0, 0, q - 2)$ and $(0, 0, q + 3)$ are amplified, but much less so.

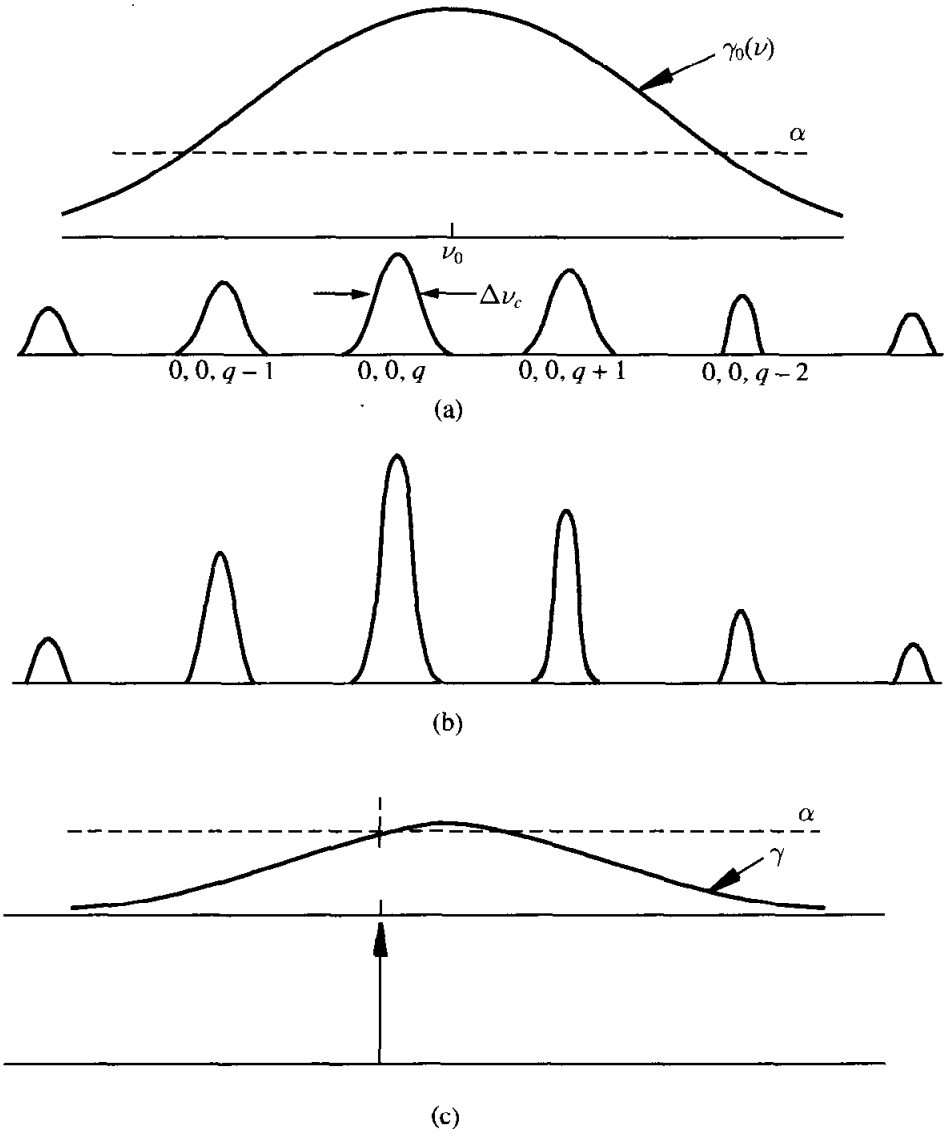


FIGURE 8.3. Evolution of laser oscillation from spontaneous emission: (a) initial; (b) intermediate; and (c) final.

Now a few round trips through the cavity are sufficient to make the field quite large. For instance, suppose that the peak intensity of the (0, 0, q) mode in Fig. 8.3(a) were 1 μW/cm^2 and that the net gain $(R_1 R_2)^{1/2} \exp[\gamma_0(\nu_q)l] = 4$. After just five round trips, this intensity will have grown by a factor of $4^{10} \sim 1.05 \times 10^6$ provided that the gain coefficient stays constant during this process. The other modes of Fig. 8.3 grow but not nearly as fast. If, for instance, $(R_1 R_2)^{1/2} \exp[+\gamma(\nu_{q+1})l] = 2$ and its initial intensity were $0.5\mu W/\text{cm}^2$, its value would be 0.5 mW/cm^2 after five round trips through the cavity. Clearly, the (0, 0, q) mode is much larger than the rest, but, even so, the 0.5 mW/cm^2 intensity of the (0, 0, $q+1$) mode is significant.

It should also be clear that something has to give because the intensity cannot keep growing indefinitely through more and more cavity transit times. Every time one more photon is added to the field inside the cavity, the population inversion must have decreased by 2. (Why not by 1?)

When the stimulating field is so large as to cause the atoms to give up their energy as fast as they are being pumped up the energy scale, we have reached an equilibrium. Thus the gain of the system must change to a lower value until the rate of *production* of the excess inverted population is balanced by the use rate by stimulated emission. (This is called gain saturation and will be analyzed later.)

A moment's consideration of the issues raised in the previous two paragraphs leads to the acceptance of Fig. 8.3(c) for the representation of the final state of the laser gain and spectral content. There are three items to be emphasized and noted.

1. The laser gain coefficient has decreased (or saturated*) to the loss coefficient *at the frequency of laser oscillation.* The relative shape of the gain curve is similar to its initial one, although the peak value at line center is considerably reduced.
2. The spectral shape has changed dramatically from Fig. 8.3(a) or (b), with the central mode much larger than any of the other modes. Indeed, all other modes are now *below* threshold and are invisible in comparison to the laser amplitude. For instance, the gain on the $(q+1)$ mode at ν_{q+1} is less than the loss at that frequency, and in spite of the initial phase of rapid growth of that mode, dies down to a level sustained only by spontaneous emission.
3. Laser oscillation occurs at the center of the cavity mode with the highest net gain.

The fact that the laser oscillates on only one cavity mode is a consequence of the assumption of *homogeneous broadening.* Recall that the definition of homogeneous broadening is that *all* atoms behave in the same manner. Thus, if any one atom gives up its energy (as a photon) to any field at any frequency, that atom can no longer contribute to the gain at another frequency. Consequently, the gain profile sags while maintaining proportionality over the spectrum.

*The gain will actually saturate at a value slightly lower than the loss line, with the very slight difference being made up by the spontaneous emission. This leads to a finite spectral width of oscillation. For many, if not most applications, this fine point can be neglected.

However, the spectral shape within a cavity-mode resonance changes dramatically. Just as one cavity mode wins the foot race for the energy stored in the population inversion at the expense of other cavity modes, so does the frequency at the peak of the cavity resonance overwhelm the other nearby ones. There is a nonzero width to the oscillation spectrum because of the quantum nature of the generation process, but usually the mechanical, acoustical, or thermal fluctuations in the cavity length contribute a spectral width that is many times that allowed by quantum effects.

8.3 GAIN SATURATION IN A HOMOGENEOUS BROADENED TRANSITION

To describe the final state of the laser, we need a mathematical description of gain saturation. Toward this end, we construct a generalized model of the two atomic states involved in the laser, as shown in Fig. 8.4. In the figure, R represents the rate of producing the appropriate state as a result of all causes other than those indicated on the diagram. For instance, R_1 includes direct excitation from the ground state to state 1 and also any indirect routes such as excitation to a higher state followed by spontaneous emission from the higher state back to 1. It does *not* include the spontaneous decay of state 2 into 1 nor the stimulated processes. These will be considered separately. State 2 is pumped directly from ground at a rate R_2 that includes indirect paths to higher levels followed by a decay to 2, or it may represent a transfer from a different gas into the one of interest.

Once the atoms are in the atomic levels, they suffer a variety of fates. For instance, state 2 can radiate a photon of energy $h\nu_{12}$ spontaneously, converting an atom from 2 into an atom named by 1, or some internal collision can effect this conversion, but in either case the rate of *decay* of state 2 is proportional to the density in 2 times a rate constant. Atoms in state 1 suffer similar fates: They can radiate spontaneously to another level, be deactivated by a collision, or be simply swept out of the volume of interest by mass motion (as is done in a flowing dye laser). The stimulated emission rates shown in Fig. 8.4 will be in addition to the above "natural decay" processes. To avoid excessive arithmetic, we assume that the

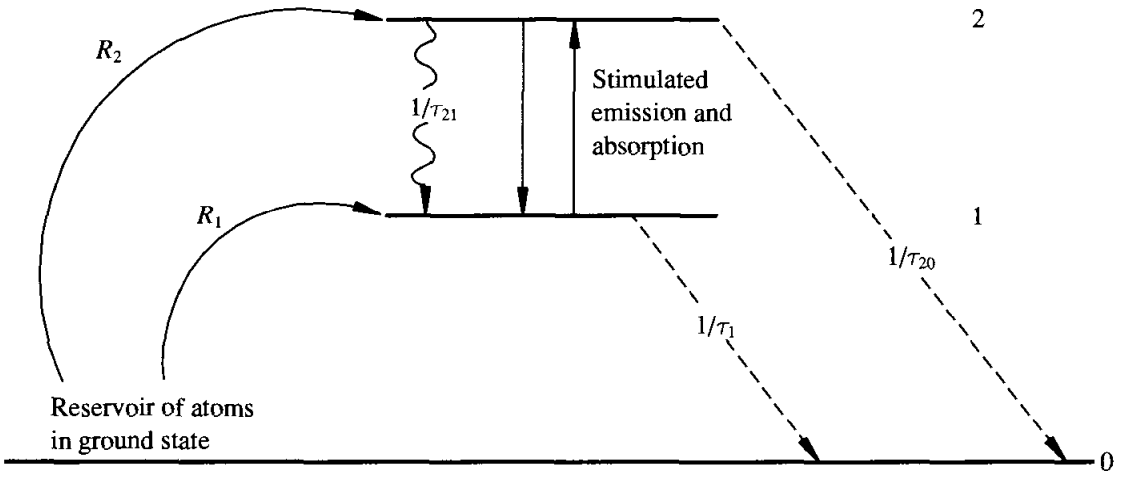


FIGURE 8.4. Generalized pumping scheme of a laser.

populations in 2 or 1 are very small compared to that in 0 so that we need not worry about the conservation of mass.*

The differential equations describing the dynamics of the populations will be written in terms of the *lifetime* for certain processes to take place. For instance:

1. The sum of the spontaneous emission rate and any other collision process which decreases the population in 2 and simultaneously increases that in 1 is denoted by $1/\tau_{21}$.
2. The rate of loss of state 2 that does not result in an atom in state 1 is denoted by $1/\tau_{20}$.
3. The total rate of decay of state 2 is the sum of the above rates and defines the "lifetime" of state 2 by $1/\tau_2 = 1/\tau_{21} + 1/\tau_{20}$.
4. The decay rate of state 1 caused by any and all effects is denoted by $1/\tau_1$.

The stimulated emission and absorption rates can be expressed in terms of the Einstein coefficients by $B_{21}g(\nu)\rho_\nu$ and $B_{12}g(\nu)\rho_\nu$, respectively, but it is more convenient to use the relationships between them and express the final result in terms of the stimulated emission cross section and the intensity of the radiation. (We assume equal degeneracies, $g_2 = g_1$, so as to allow any serious student of lasers to repeat the following to obtain a more general analysis, which should be done.) Recall that

$$B_{21}\rho_\nu g(\nu) = \frac{c/n_g}{h\nu} \cdot \left\{ A_{21} \frac{\lambda_0^2}{8\pi n^2} g(\nu) \right\} \cdot \frac{I_\nu}{c/n_g} = \frac{\sigma(\nu) I_\nu}{h\nu} \tag{8.3.1}$$

Thus the dynamics of the populations involved with the lasing process is described by the following coupled differential equations:

$$\boxed{\frac{dN_2}{dt} = R_2(t) - \frac{N_2}{\tau_2} - \frac{\sigma(\nu) I_\nu}{h\nu}[N_2 - N_1]} \tag{8.3.2a}$$

$$\boxed{\frac{dN_1}{dt} = R_1(t) + \frac{N_2}{\tau_{21}} + \frac{\sigma(\nu) I_\nu}{h\nu}[N_2 - N_1] - \frac{N_1}{\tau_1}} \tag{8.3.2b}$$

For most of the book, we will be considering cases where the frequency of the stimulating wave is close to line center and thus we use the peak value of σ rather than the explicit form $\sigma(\nu)$ unless the frequency dependence is crucial to the problem.

This model is so general with the equations being so fundamental and leading to so many important conclusions that it is appropriate to take a few special cases so as to appreciate the implications. Our goal is to determine the general pumping requirements to establish a population inversion and to model its use by the stimulating wave.

*In other words, $N_1 + N_2 \ll N_0$ under all circumstances, and thus N_0 is independent of the pumping rates. This is an approximation that will have to be changed for heavily pumped lasers or ones whose lower level is state 0 (see the discussion of ruby lasers in Chapter 10).

Case 1. Assume $I_\nu = 0$ (no stimulated emission), pumping to state 2 only (i.e., $R_1 = 0$), and R_2 is in the form of a step function. $R_2(t) = R_{20}u(t)$ where $u(t)$ = Heaviside step function (= 0 for $t < 0$ and = 1 for $t > 0$).

Equation 8.3.2(a) becomes

$$\frac{dN_2}{dt} + N_2/\tau_2 = R_{20}$$

which has a homogeneous solution of the form $A \exp[-t/\tau_2]$ and a particular one* equal to $R_{20}\tau_2$. The application of the initial condition $N_2(0) = 0$ leads to

$$N_2(t) = R_{20}\tau_2(1 - \exp[-t/\tau_2]) \tag{8.3.3}$$

The dynamics of state 1 is a bit more complicated since the source term N_2/τ_{21} from 2 feeds it, and thus there is a finite population in state 1 even though there is no direct pumping. Rearranging (8.3.2b) and inserting the above results leads to

$$\frac{dN_1}{dt} + \frac{N_1}{\tau_1} = \frac{N_2(t)}{\tau_{21}} = \phi_{21}R_{20}\left(1 - \exp[-t/\tau_2]\right) \tag{8.3.4a}$$

where $\phi_{21} = \tau_2/\tau_{21}$ = branching ratio.

This equation can be solved by many methods with the choice being somewhat arbitrary. The integrating factor approach leads to a nice physical interpretation and will be followed here. We multiply both sides by a factor, $\exp[+t/\tau_1]$ for this problem, which makes the left-hand side of (8.3.4b) a perfect differential:

$$\frac{d}{dt}\left\{N_1(t)\exp[+t/\tau_1]\right\} = \frac{N_2(t)\exp[+t/\tau_1]}{\tau_{21}}$$

and thus

$$N_1(t) = \frac{\exp[-t/\tau_1]}{\tau_{21}}\int_0^t N_2(t')\exp[+t'/\tau_1]\,dt' \tag{8.3.4b}$$

For times $t \ll \tau_1$, $\exp[\pm t'/\tau_1] = 1$, and thus $N_1(t) \propto \int N_2(t)\,dt$, and that fact was the reason for the choice of the solution technique. For $t \gtrsim \tau_1$, we must evaluate the integral with $N_2(t)$ as given by (8.3.3).

$$N_1(t) = \phi_{21}R_{20}\tau_1\left\{1 + \frac{\tau_1/\tau_2}{1 - \tau_1/\tau_2}\exp[-t/\tau_1] - \frac{1}{1 - \tau_1/\tau_2}\exp[-t/\tau_2]\right\} \tag{8.3.4c}$$

As $t \to \infty$, a steady state is reached with $N_2(\infty) = R_{20}\tau_2$ and $N_1(\infty) = \phi_{21}R_{20}\tau_1$. A sketch of these solutions is shown in Fig. 8.5 for two different choices of the lifetime time ratio τ_2/τ_1. If $\tau_2/\tau_1 > 1$ (Fig. 8.5, Case 1a), then the density in state 2 is always greater than that of state 1 and the system exhibits gain for all values of t. This is called a favorable lifetime ratio.

If the reverse is true, $\tau_2/\tau_1 \leq 1$ (Fig. 8.5, case 1b), called an unfavorable lifetime ratio, then gain is possible for only the short initial interval of roughly τ_2. If such a system

*The homogeneous solution always has the arbitrary constant multiplier and is found by setting the right-hand side equal to zero. The particular solution is always of the form of the forcing function *plus* all of its derivatives.

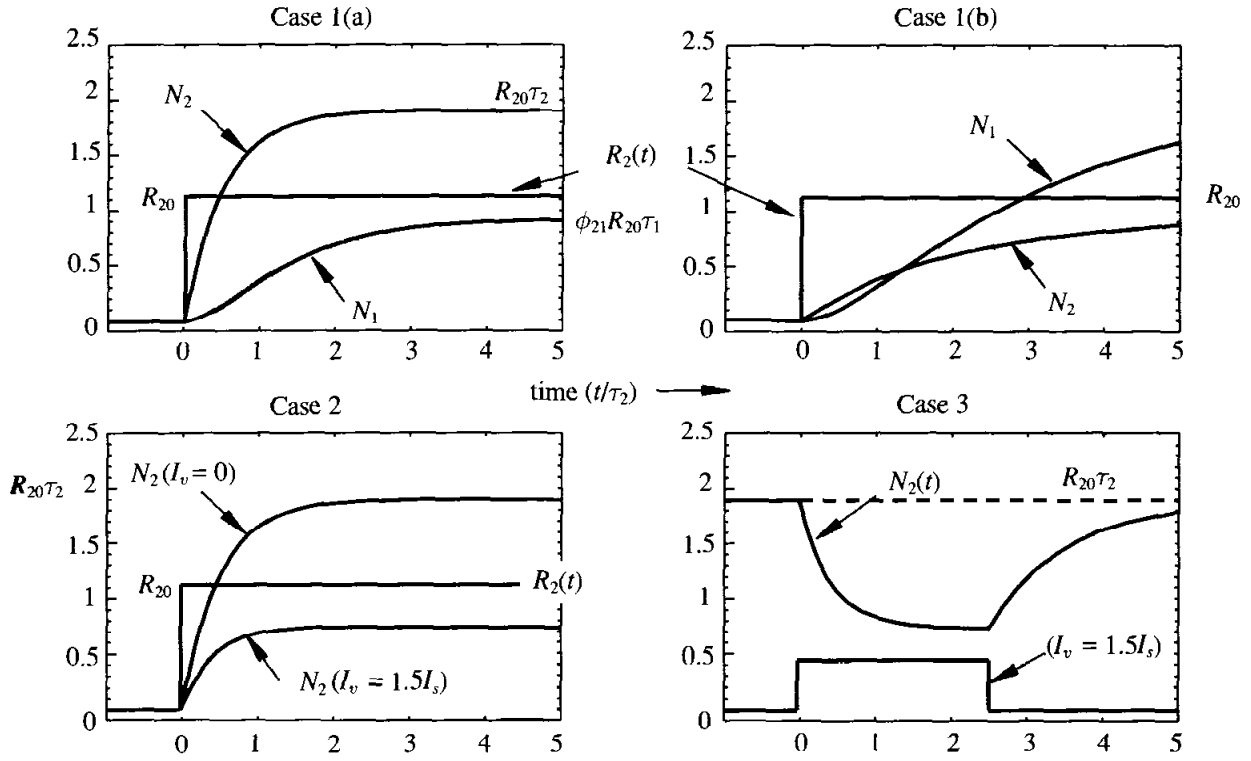


FIGURE 8.5. The variation of the populations with time (t/τ_2) for the three time dependent examples. For case 1a, the lifetime ratio $\tau_2/\tau_1 = 2$, whereas the ratio was 0.5 for 1b, and $\tau_1 = 0$ for cases 2 and 3.

were to be used for a laser, the excitation must be in the form of a fast rising pulse, since the time interval beyond the gain interval is just a waste of power. The molecular nitrogen laser at $\lambda_0 = 337$ nm is an example of one with a very unfavorable lifetime ratio ($\tau_1 \sim 10\ \mu$s, $\tau_2 \approx 10$ ns), whereas the Nd:YAG system has a very favorable lifetime ratio with $\tau_2 = 255\ \mu$s and $\tau_1 \approx 30$ ns.

The spontaneous emission emitted on $2 \rightarrow 1$ and $1 \rightarrow 0$ has the same time dependence as $N_2(t)$ and $N_1(t)$, respectively.

Case 2. Now we include a finite value of the intensity in the rate equations, but assume $\tau_1 = 0$ (just to simplify the mathematics). Since N_1 decays infinitely fast, the population in state 1 must be zero. Such a situation corresponds to an ideal laser system and is a very good model for the rare gas-halogen (excimer) lasers and for the important solid-state Nd:YAG system.

$$\frac{dN_2}{dt} + \frac{1}{\tau_2}\left[1 + \frac{\sigma\tau_2}{h\nu} I_\nu\right] N_2 = \frac{dN_2}{dt} + \frac{1}{\tau_2}\left[1 + \frac{I_\nu}{I_s}\right] N_2 = R_{20}u(t) \tag{8.3.5}$$

which has the solution given by

$$N_2(t) = \frac{R_{20}\tau_2}{(1 + I_\nu/I_s)} \left\{1 - \exp\left[-\frac{t}{\tau_2}\left(1 + \frac{I_\nu}{I_s}\right)\right]\right\} \tag{8.3.6}$$

where $$\boxed{I_s = h\nu/\sigma\tau_2 = \text{saturation intensity}} \tag{8.3.7}$$

Equation (8.3.6) may appear to be more complicated than (8.3.3), but it really has the same generic mathematical format with a characteristic amplitude multiplier and an exponential approach to equilibrium. Stimulated emission by the intensity I_ν affects both of these quantities as illustrated in Fig. 8.5 (case 2).

The time dependence of the population N_2 as predicted in case 1 (with $I_\nu = 0$) is shown along with that given by (8.3.6) for $I_\nu/I_s = 1.5$. The quantity $h\nu/[\sigma(\nu)\tau_2]$ is an important collection of atomic constants with the dimension of watts/area (i.e., $J \cdot T^{-1} \cdot L^{-2}$) and is called *the saturation intensity*. The ratio of the intensity, I_ν, to this benchmark or characteristic value indicates whether stimulated emission is or is not important. For example,

1. If $I_\nu \ll I_s$, the population N_2 is barely affected by the stimulating wave.
2. If $I_\nu = I_s$, the steady state peak amplitude of N_2 is one-half of the value given by (8.3.3). Furthermore, the time constant controlling the approach to equilibrium is reduced, (i.e., $\tau_2' = \tau_2 \div [1 + I_\nu/I_s]$.
3. If $I_\nu \gg I_s$, the peak population in N_2 is greatly depressed compared to case 1, and the effective time constant is also reduced.

Since the population in state 2 is reduced, the spontaneous emission from it is also reduced.

Case 3. Assume the steady-state conditions found in case 1, assume $\tau_1 = 0$ to simplify the arithmetic, but now let the intensity have the form of a pulse of duration $T \gg \tau_2$. The dynamics are described by (8.3.5), but the initial value of N_2 at $t = 0$ is $R_{20}\tau_2$. The integrating factor that makes the left-hand side a perfect differential is given by

$$\exp\left[+\frac{1}{\tau_2}\int_0^t \left(1 + \frac{I_\nu(t)}{I_s}\,dt\right)\right] \rightarrow \exp\left[+\frac{t}{\tau_2'}\right] \qquad \text{with} \qquad \tau_2' = \frac{\tau_2}{1 + I_0/I_s}$$

for $I_\nu(t) = I_0[u(t) - u(t - T)]$.

$$\therefore N_2(t) = A\exp\left[-\frac{t}{\tau_2}\left(1 + \frac{I_\nu}{I_s}\right)\right] + \frac{R_{20}\tau_2}{1 + I_\nu I_s}$$

$$N_2(0) = R_{20}\tau_2 = A + \frac{R_{20}\tau_2}{1 + I_\nu I_s} \qquad \text{or} \qquad A = R_{20}\tau_2 \cdot \frac{I_\nu/I_s}{1 + I_\nu/I_s}$$

Hence,

$$N_2(t) = \frac{R_{20}\tau_2}{1 + I_\nu/I_s}\left\{(I_\nu/I_s)\exp\left[-\frac{t}{\tau_2'}\right] + 1\right\} \tag{8.3.8}$$

This is sketched in Fig. 8.5 and shows the dynamic behavior of the population N_2 in response to the stimulating wave. Note that N_2 approaches the saturated value with a time constant

of τ_2', but relaxes back to the equilibrium value at τ_2. The spontaneous emission from N_2 would have the same temporal behavior, and thus the presence of the pulse of intensity depresses the side fluorescence as well. The physical reason for the depression of the side fluorescence is that the stimulating wave forces the atoms to give up their stored energy to it rather than permitting the atoms to radiate in any of the 4π directions.

If $\tau_1 \neq 0$ but still less than τ_2, then the density N_1 would be enhanced by the stimulated emission from 2, and thus the spontaneous emission out of 1 would increase during the pulse.

When the intensity is turned off, the population in state 2 reapproaches the initial equilibrium with a time constant $\tau_2 > \tau_2'$. The decrease in the lifetime of state 2, the depression of the upper state density (and thus side fluorescence), and enhancement of the lower state population are easily measured responses of the atoms to the stimulation. These effects are large or small depending upon the ratio of the intensity I_ν to this collection of atomic parameters called the saturation intensity. For start-up behavior, the intensity is always small since the laser starts from noise. Hence

Stimulated emission can always be neglected for threshold calculations.

For a laser system above threshold, the internal intensity grows with time with the stimulated emission depressing the population of N_2 (and thus increasing N_1). A steady state or equilibrium in the laser intensity is established when the excess population inversion is used just as fast as it is produced making the *saturated* gain equal to the inverse of the survival factor of the passive cavity. Thus, I_ν will always be comparable to I_s and

Stimulated emission can never be neglected in the dynamics of the laser.

Thus, if we were to imagine increasing the pumping rate from zero, say, by twisting some knob to the right, the upper state density and thus the spontaneous power out of it would increase until threshold is reached. Any further increase results in a dramatic increase in intensity along the axis of the laser, but *no increase* in the other 4π directions.

This behavior is shown very clearly by the data from Paoli using a semiconductor laser and is presented in Fig. 8.6. For injected currents below threshold, a laser is a light emitting diode (LED) with the photons going in all directions, including to the side and that emission increases with current. Once threshold is exceeded, the emission to the side is *clamped* at the value it approached from below. This is a general characteristic of all CW lasers where the time scale for any transients is slow compared to atomic lifetimes or the photon lifetime of the cavity. It is not true for fast phenomena, as we shall see.

Let us now address another practical case of a steady state system but make no assumptions about the lifetime ratios.

Case 4. $\partial/\partial t = 0$. Assuming all time derivatives equal to zero reduces the differential equations to a simple algebraic set:

$$\frac{dN_2}{dt} = 0 = R_2 - \frac{N_2}{\tau_2} - \frac{\sigma I_\nu}{h\nu}(N_2 - N_1) \tag{8.3.9a}$$

$$\frac{dN_1}{dt} = 0 = R_1 + \frac{N_2}{\tau_{21}} - \frac{N_1}{\tau_1} + \frac{\sigma I_\nu}{h\nu}(N_2 - N_1) \tag{8.3.9b}$$

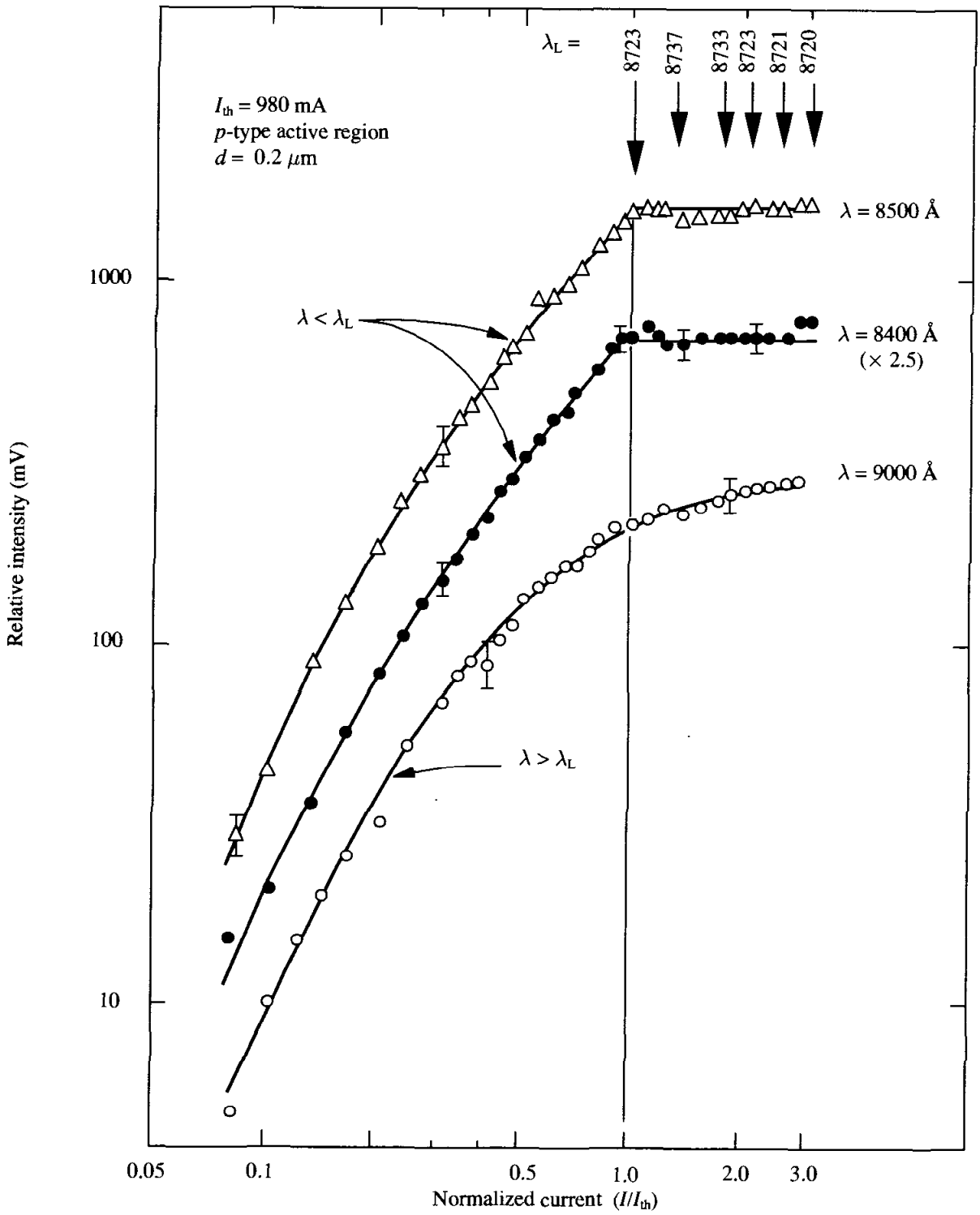


FIGURE 8.6. Variation of the spontaneous emission from the side of a semiconductor diode as the pumping current is increased. Once the diode starts lasing, stimulated emission uses the carriers as fast as they are injected into the junction. Hence the inversion is clamped at threshold. (Data from T. Paoli. *IEEE J. Quant. Electr. QE-9*, 267, 1973.)

Rewriting these equations in matrix form emphasizes the simplicity (but drudgery) involved with the determination of $N_{2,1}$ as a function of $R_{2,1}$ and I_ν.

$$\begin{bmatrix} \left(\frac{1}{\tau_2} + \frac{\sigma I_\nu}{h\nu}\right) & -\frac{\sigma I_\nu}{h\nu} \\ -\left(\frac{1}{\tau_{21}} + \frac{\sigma I_\nu}{h\nu}\right) & \left(\frac{1}{\tau_1} + \frac{\sigma I_\nu}{h\nu}\right) \end{bmatrix} \begin{bmatrix} N_2 \\ N_1 \end{bmatrix} = \begin{bmatrix} R_2 \\ R_1 \end{bmatrix} \tag{8.3.9c}$$

Now

$$\Delta = \frac{1}{\tau_1\tau_2} + \left[\frac{1}{\tau_1} + \frac{1}{\tau_2} - \frac{1}{\tau_{21}}\right]\left(\frac{\sigma I_\nu}{h\nu}\right)$$

or

$$\Delta = \frac{1}{\tau_1\tau_2}\left[1 + \left(\tau_1 + \tau_2 - \frac{\tau_1\tau_2}{\tau_{21}}\right)\left(\frac{\sigma I_\nu}{h\nu}\right)\right] \tag{8.3.10}$$

Using Cramer's rule, we find the populations $N_{2,1}$:

$$N_2 = \left\{R_2\left(\frac{1}{\tau_1} + \frac{\sigma I_\nu}{h\nu}\right) + R_1\left(\frac{\sigma I_\nu}{h\nu}\right)\right\}/\Delta \tag{8.3.11a}$$

$$N_1 = \left\{R_1\left(\frac{1}{\tau_2} + \frac{\sigma I_\nu}{h\nu}\right) + R_2\left(\frac{1}{\tau_{21}} + \frac{\sigma I_\nu}{h\nu}\right)\right\}/\Delta \tag{8.3.11b}$$

Now the pertinent parameter insofar as a laser is concerned is the difference in populations, $N_2 - N_1$. Hence after some obvious manipulation, we can express that difference as

$$N_2 - N_1 = \frac{R_2\tau_2\left(1 - \frac{\tau_1}{\tau_{21}}\right) - R_1\tau_1}{1 + \left(\tau_1 + \tau_2 - \frac{\tau_1\tau_2}{\tau_{21}}\right)\left(\frac{\sigma I_\nu}{h\nu}\right)} \tag{8.3.11c}$$

Now if (8.3.11c) is multiplied by the stimulated emission cross section, we have an expression for the gain coefficient for any value of the intensity. If the intensity is small enough such that the denominator is approximately 1, then the numerator is an expression for the small-signal gain coefficient in terms of the pumping rates and the lifetimes:

$$\gamma_0(\nu) = \sigma(\nu)\left[R_2\tau_2\left(1 - \frac{\tau_1}{\tau_{21}}\right) - R_1\tau_1\right] \tag{8.3.12a}$$

The rate of increase of intensity with distance divided by the intensity is, by definition, the gain coefficient (for any intensity) and is given by

$$\frac{1}{I_\nu}\frac{dI_\nu}{dz} \triangleq \frac{\gamma_0(\nu)}{1 + I_\nu/I_s(\nu)} \tag{8.3.12b}$$

where I_s is now given by

$$I_s = \frac{h\nu}{\sigma(\nu)\tau_2} \cdot \frac{1}{1 + \frac{\tau_1}{\tau_2}\left(1 - \frac{\tau_2}{\tau_{21}}\right)} \tag{8.3.12c}$$

Let us focus on (8.3.12c) for a moment: if $\tau_1 \ll \tau_2$, which is referred to as a favorable lifetime ratio, or the quantity $\tau_2/\tau_{21} \sim 1$, which is called a favorable branching ratio, or $g_2/g_1 < 1$, which is called a favorable degeneracy ratio, or any combination of these three "ifs" (which is quite common), then the denominator of (8.3.12c) is approximately 1 and the saturation intensity is identical to that found previously from the temporal analysis (8.3.7).

Thus there are two equivalent definitions of the saturation intensity:

$$I_s = \begin{cases} \text{that intensity, which shortens the lifetime of state 2 by a factor of 2 (8.3.7)} \\ \\ \text{that intensity, which reduces the gain } \textit{coefficient} \text{ by a factor of 2 (8.3.12c)} \end{cases} \tag{8.3.13}$$

It must be emphasized that the saturation intensity is just a collection of constants which have the dimensions of intensity and indicate when a stimulating wave is strong or weak. As we shall see, all lasers operate with an intensity which is more or less equal to the saturation intensity, and, hence, this is the first-order estimation for a laser amplitude inside the cavity. It should also be noted that we have buried the frequency dependence of the stimulated emission cross section under the symbol σ. If we retrace the steps made above, the line shape favor $\bar{g}(\nu)$, normalized to be unity at line center, should multiply σ everywhere it appears. For instance, if the homogeneous line shape were a Lorentzian, then $\bar{g}(\nu)$ is given by

$$\bar{g}(\nu) = \frac{(\Delta\nu/2)^2}{(\nu - \nu_0)^2 + (\Delta\nu/2)^2} \tag{8.3.14}$$

and the saturation law should be written as

$$\frac{1}{I_\nu}\frac{dI_\nu}{dz} = \frac{\gamma_0(\nu)}{1 + \bar{g}(\nu)\frac{I_\nu}{I_s}} \tag{8.3.15}$$

For the most part, we are concerned with radiation close to line center where $\bar{g}(\nu) \sim 1$, and most of the following discussion assumes that approximation.

This analysis is most important, and the serious laser student would do well to practice arriving at the above conclusions by as many different routes as possible. It is much more than a dry mathematical description of the formation, decay, and use of atoms by stimulated emission; it also indicates the general physical requirements for achieving a population inversion and gain.

Now let us focus the discussion on (8.3.11c). Obviously, we want the first term in the numerator to be as large as possible, keeping the second as small as possible. Thus we want the pumping rate of the upper state to be high (R_2 large), keeping R_1 small. Not only should the pumping rates be in the appropriate direction, but the lifetime of state 2 should be long, whereas that of state 1 should be short. However, the limiting lifetime of state 2 is, of course, τ_{21}. For, if τ_2 were infinite, τ_{21} must also be infinite; then A_{21} would be zero and

the stimulated emission cross section is also zero. The lifetime of state 1 should be as short as possible so as to deplete the population N_1 as fast as possible. Indeed, as we will see later, the *depletion* of the lower state and the pumping of the upper state are the rate-limiting processes for a good laser.

Now let us return to (8.3.15), which describes the situation shown in Fig. 8.7, where a signal I_1 from an external laser is injected into an amplifier. The intensity of the injected signal is $I_\nu(z = 0)$, and the problem is to determine the output $I_\nu(z = l_g) = I_2$. It must be emphasized that the solution to (8.3.12) is *not* given by the simple exponential law with an intensity in the exponent.

$$I_2 \neq I_1 \exp\left[\frac{\gamma_0(\nu) l_g}{1 + \bar{g}(\nu)(I_\nu/I_s)}\right] \qquad \text{(No!)}$$

We must solve the differential equation as given by (8.3.15). Bringing all factors involving I_ν to the left side and dz to the right leads to

$$\int_{I_1}^{I_2} dI_\nu \left[\frac{1}{I_\nu} + \frac{\bar{g}(\nu)}{I_s}\right] = \int_{z=0}^{z=l_g} \gamma_0(\nu)\, dz \tag{8.3.16}$$

thus

$$\ln \frac{I_2}{I_1} + \frac{\bar{g}(\nu)}{I_s}[I_2 - I_1] = \gamma_0(\nu) l_g \tag{8.3.17a}$$

or

$$\ln G + \bar{g}(\nu)\frac{I_1}{I_s}(G - 1) = \gamma_0(\nu) l_g \tag{8.3.17b}$$

with

$$G \triangleq I_2/I_1 \tag{8.3.17c}$$

Note that if the output I_2 [and, of course I_1] is much smaller than I_s, then we can ignore the second term on the left and recover the simple amplification law:

$$\left.\frac{I_2}{I_1}\right|_{\text{small signal}} = \exp[\gamma_0(\nu) l_g] = G_0(\nu) \qquad \text{(power gain)} \tag{8.3.18}$$

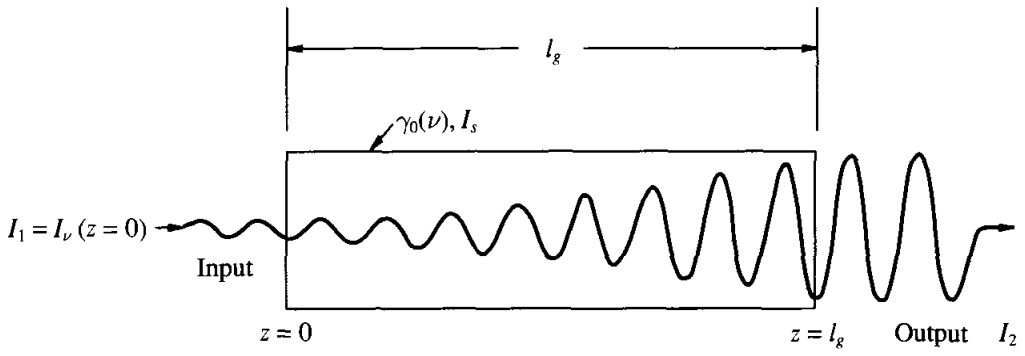


FIGURE 8.7. An optical amplifier.

If the input is comparable to the saturation intensity, the rate of increase of intensity is smaller; consequently, the net power gain, G, is smaller. Although (8.3.17) is fairly simple, it is a transcendental equation that must be solved numerically for the output in terms of the input. The general trend of a solution is shown in Fig. 8.8, where the power gain (in dB) is plotted as a function of ratio $I_{\text{in.}}/I_{\text{sat.}}$ (at line center, where $\bar{g} = 1$). It is instructive and quite revealing to go to the extreme limit of $I_1 \gg I_s$ to find the output intensity directly from (8.3.15).

$$I_2 = I_1 + \left[\frac{\gamma_0(\nu)I_s}{\bar{g}(\nu)}\right] l_g \qquad (\text{for} \quad I_1 \gg I_s) \tag{8.3.19}$$

Note that $\gamma_0(\nu)$ contains the same frequency dependence as $\bar{g}(\nu)$, and thus we can only *add* a certain amount of intensity $\gamma_0(\nu_0)I_s$ per unit of length of our amplifier independent of the detuning from line center. Let us reinsert the parameters of $\gamma_0(\nu_0)$ and I_s from (8.3.12a) and (8.3.12c) to show the obvious logic of the result.

$$\gamma_0(\nu) = \left[R_2\tau_2\left(1 - \frac{\tau_1}{\tau_2}\right) - R_1\tau_1\right]\sigma(\nu_0)\bar{g}(\nu) \tag{8.3.12a}$$

$$I_s = \frac{h\nu}{\sigma(\nu_0)\tau_2\left[1 + \dfrac{\tau_1}{\tau_2}\left(1 - \dfrac{\tau_2}{\tau_{21}}\right)\right]} \tag{8.3.12c}$$

Thus the power added (per unit of length) of this saturated amplifier is

$$\frac{\Delta P}{\text{area} \times \text{length}} = \frac{\gamma_0(\nu)I_s}{\bar{g}(\nu)} = h\nu\left[\frac{R_2\tau_2(1 - \tau_1/\tau_{21}) - R_1\tau_1}{\tau_1 + \tau_2 - \tau_1\tau_2/\tau_{21}}\right] \tag{8.3.20}$$

This is the *best* we can do. Note that the detailed picture of the photon interacting with the atom has completely disappeared, leaving only lifetimes and pumping rates for our consideration. Indeed, the quantity in the brackets can be considered as an effective pumping rate, each effective pumping event contributing one package of energy, $h\nu$, to the incoming

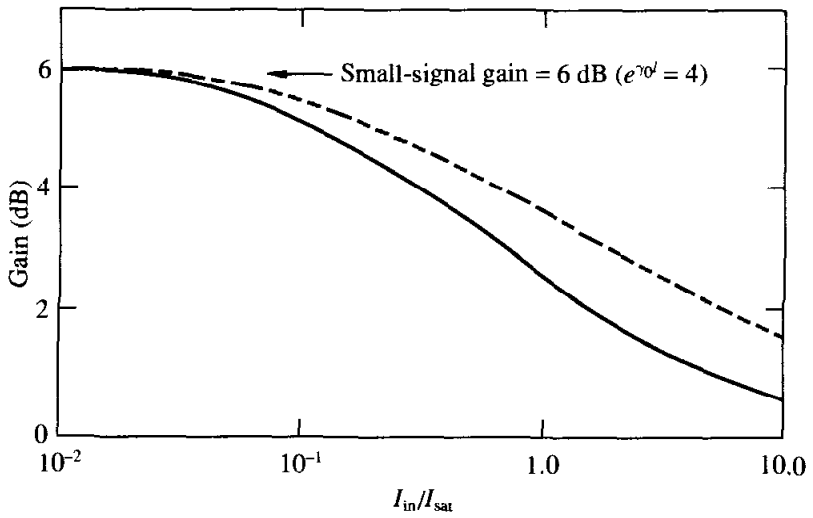


FIGURE 8.8. Saturation of the gain of an optical amplifier. The solid curve is for a homogeneously broadened system and the dashed curve is for an inhomogeneous one. (See Problem 8.6.)

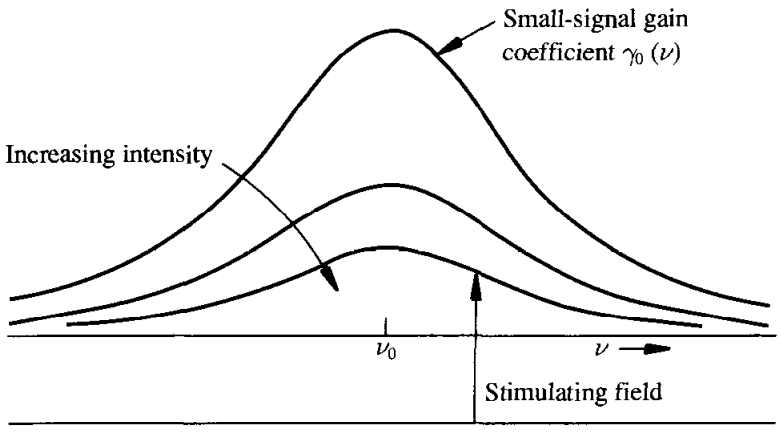


FIGURE 8.9. Saturation of a homogeneously broadened transition.

wave.

$$R_{\text{eff}} = \frac{R_2\tau_2(1 - \tau_1/\tau_2) - R_1\tau_2}{\tau_1 + \tau_2 - \tau_1\tau_2/\tau_{21}} \tag{8.3.21}$$

If we again consider the ideal laser system—weak pumping of the lower state (R_1 small), lower state lifetime small compared to the upper one ($\tau_1 \ll \tau_2$) and unity branching ratio ($\tau_2 \sim \tau_{21}$)—then R_{eff} reduces to R_2, the pumping rate of the upper state. This means that we can only extract what we put in, a most logical result! Thus the primary job in laser research is to devise a means of pumping the upper state. For that reason, then, we will devote a considerable amount of effort in later chapters to the physical processes that lead to the formation of the upper state.

Before this section is closed, it is well to reexamine the assumptions implicit in the derivation so as to appreciate the limitations of the result. The major assumption lies in the cavalier manner of writing the rate equations (8.3.2a) and (8.3.2b). We assumed that the radiation at ν interacted with the atoms according to the line shape $g(\nu)$ and implicitly assumed that the line shape for all atoms was the same. Thus, if an atom in state 2 gave up its internal energy to the stimulating radiation field and became part of the population in state 1, the gain coefficient throughout the entire line profile would be reduced. This is shown in Fig. 8.9.

8.4 LASER OSCILLATION IN AN INHOMOGENEOUS SYSTEM

Many (if not most) lasers use a system that is inhomogeneously broadened by one or more of the many mechanisms noted in Sec. 7.6. For instance, the common He:Ne laser at 6328 Å oscillates on a line that is composite of two isotopes with both being Doppler broadened so that the width $\Delta\nu_D$ ($\sim$ 1.5 GHz) is many times the homogeneous width. One of the consequences of operation on an inhomogeneously broadened line is that multi-longitudinal mode oscillation is the rule and oscillation on the mode nearest to line center does not necessarily depress the gain farther out on the wings as was indicated in Fig. 8.9 for a homogeneous system.

Probably the easiest way to appreciate these statements would be to apply the logic of Sec. 8.2 to the case of a two-isotope system assuming that the same transitions are distinct.

For example, consider the ($n = 3 \rightarrow n = 2$) line in the Balmer series in hydrogen (1 AMU) and deuterium (2 AMU). The wave number (i.e., $1/\lambda_0$) of each is found in every elementary physics book.

$$\bar{\nu}_{H \text{ or } D} = \left[R_{H \text{ or } D}\right]\left(1/2^2 - 1/3^2\right) = 15{,}233\text{cm}^{-1} \Rightarrow 6564 \text{ Å } (H_\alpha)$$

where $R_{H \text{ or } D} = \dfrac{M_{(H,D)}}{m_e + M_{(H,D)}} R_\infty$ and $R_\infty = 109{,}737.318 \text{ cm}^{-1}$, the Rydberg constant.

Thus, the shift between the center of the lines is given by $\bar{\nu}_D - \bar{\nu}_H = 4.1473 \text{ cm}^{-1}$ or $\nu_D - \nu_H = 124.4$ Ghz, $\lambda_H - \lambda_D = 1.79$ Å. Such a separation is small compared to the mean wavelength but huge compared to a free spectral range (FSR) of ~ 200 MHz of a typical gas laser cavity. Hence, *if* atomic hydrogen would lase on the $n = 3 \rightarrow n = 2$ transition then a 50:50 mixture of H_2 and D_2 would lase on at least two modes separated by 1.79 Å.

This example is, of course, an extreme and unrealistic case, but it does illustrate that multimode oscillation is possible with the parts contributing gain according to individual line shapes making up the inhomogeneous profile. Let us consider a more realistic (but still hypothetical) case of an inhomogeneously broadened line being composed of 17 different parts with a common homogeneous line width $\Delta\nu_h$. These different parts have an occurrence probability of (1:2:4:8:12:16:18:19:20:19:18:16:12:8:4:2:1) with each "line" being shifted by a half width $\Delta\nu_h/2$ from the preceding one. Let $\nu_9 = \nu_0$ be the central or most probable group that corresponds to the part with weight of 20. Thus the small signal gain coefficient is just the weighted sum of the various parts:

$$\begin{aligned}\gamma_0(\nu) = A_{21}\frac{\lambda^2}{8\pi n^2}\frac{2}{\pi\Delta\nu_h}\Bigg\{ &\frac{1}{180}\frac{1}{(\nu_1 - \nu)^2 + (\Delta\nu_h/2)^2} \\ &+ \frac{2}{180}\frac{1}{(\nu_2 - \nu)^2 + (\Delta\nu_h/2)^2} + \cdots \\ &+ \frac{19}{180}\cdot\frac{1}{(\nu_8 - \nu)^2 + (\Delta\nu_h/2)^2} + \cdots \\ &+ \frac{1}{180}\frac{1}{(\nu_{17} - \nu)^2 + (\Delta\nu_h/2)^2}\Bigg\}\left[N_2 - \frac{g_2}{g_1}N_1\right]\end{aligned}$$

where the numerical factor of 180 is just the sum of the weights, the frequencies $\nu_1, \nu_2, \cdots \nu_{17}$ are the centers of each line, and the terms in the sum are the homogeneous (Lorentzian) line shapes for each component.

The brace is plotted in Fig. 8.10 and illustrates that the composite laboratory line width is considerably larger than the homogeneous width of each of its components.* A few of the individual lines are also shown in this diagram to emphasize that this inhomogeneous line shape is the sum of its various parts. Aside from the minor "bumps" (owing to the finite number of distinct terms), the line shape "looks" like another Lorentzian but with a larger width.

*(The interested student will measure the FWHM of Fig. 8.10 to find its width is ~ 9.6 normalized units or $9.6 \times \Delta\nu_h/2 = 4.8\Delta\nu_h$ frequency units.)

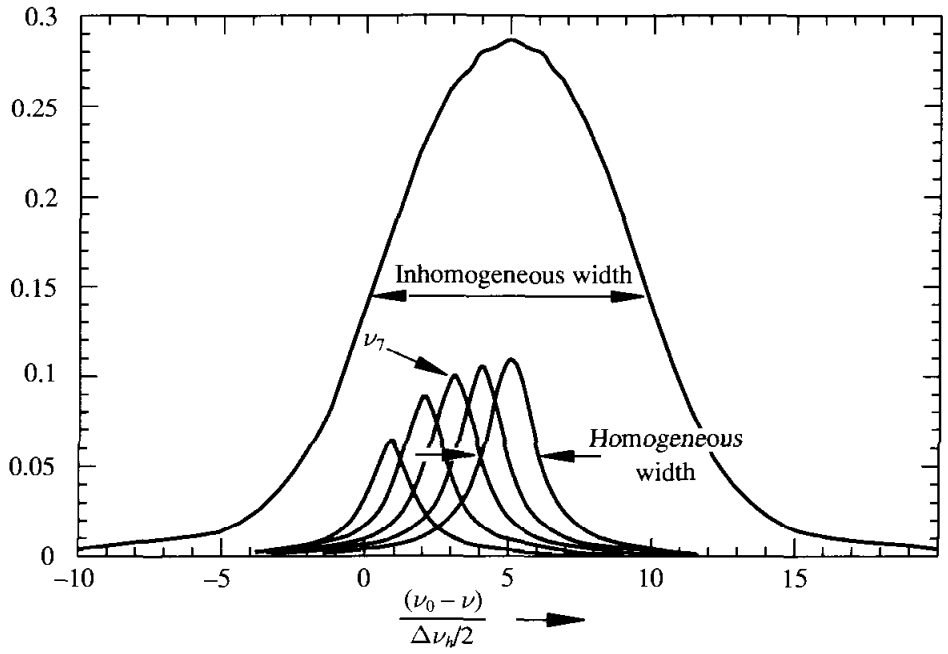


FIGURE 8.10. The weighted sum of 17 homogeneous lines to form the inhomogeneous line shape.

When we consider the use of the population inversion by stimulated emission (i.e., saturation), we must be precise and very careful by applying the analysis of Sec. 8.3 to each component part that saturates according to its homogeneous law. Thus the large signal gain coefficient becomes:

$$\frac{1}{I_\nu}\frac{dI_\nu}{dz} = \sum_m \frac{\left[N_2^m - (g_2/g_1)N_1^m\right]\sigma_h(\nu)}{1 + \dfrac{\sigma_h(\nu)\tau_2}{h\nu}\cdot I_\nu} \tag{8.4.1}$$

where the sum is over the m different groups and the superscript m indicates the fraction of the density of atoms in (2,1) that belong to the group m in the absence of stimulated emission. Equation (8.4.1) might appear to be similar to (8.3.12), but the summation can make a significant difference in the use of the total inversion depending on how the various groups interact.

If the various groups are closely coupled by some fast rate so as to maintain the populations with the same relative weights, then the result of Sec. 8.3 are applicable. If this rate is faster than the stimulated emission rate for any one group, then the use of the gain is according to the homogeneous law and the inhomogeneous line shape is maintained. The inhomogeneous line shape is used to evaluate the peak gain coefficient and the saturation intensity. We will see an example of this when the YAG laser is discussed in Chapter 10.

If, however, the groups are not closely coupled, the saturation behavior is considerably different: One group can be used by stimulated emission whereas the wave may ignore another. Figure 8.11 is a plot of (8.4.1) for the 17 group example chosen previously for a stimulating wave of $I/I_s = 2$ at $\nu = \nu_0 + \Delta\nu_h/2$. Notice that the gain profile is not depressed uniformly. Rather, a "hole" is being burned into the small signal line shape.

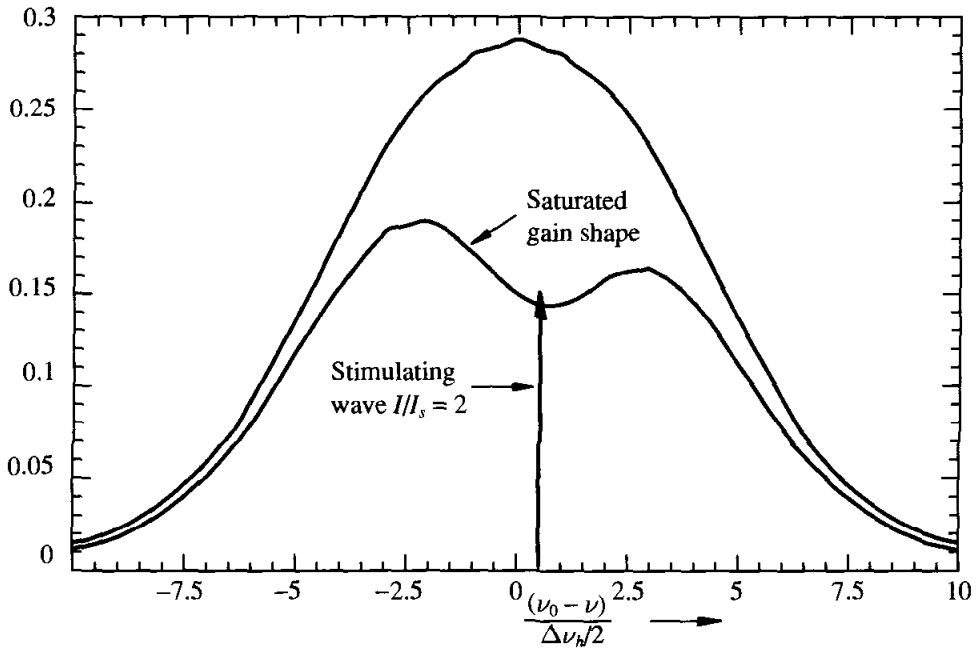


FIGURE 8.11. The saturation of the gain profile presuming weak coupling between the various groups.

It uses the atoms close to the stimulating frequency but leaves those for the large detunings unaffected. The line shape of the common helium/neon gas laser is an example of inhomogeneous broadening in which "hole" burning is readily observed.

While the stimulation always results in a photon of the same frequency as that of the inducing wave, the relative strength of its interaction is proportional to the homogeneous line shape. The ratio I/I_s still determines whether I is strong or weak, but we must use the stimulated emission cross section with the homogeneous line shape to evaluate $I_s = h\nu/\sigma_h\tau_2$.

This hole burning phenomenon has some important consequences for lasers operating in an inhomogeneously broadened line. Let us talk our way through a specific case of a low-pressure Doppler-broadened system, such as that found in the He:Ne laser, to emphasize the contrast between it and that described in Sec. 8.2.

Consider Fig. 8.12, which describes the various phases of the buildup of coherent optical power within the laser cavity. Aside from minor differences in the shape of the gain curve, Figs. 8.12(a) and 8.12(b) are identical to Figs. 8.3(a) and 8.3(b). The spontaneous emission goes into the cavity modes more or less like the line shape, and each cavity mode grows according to the gain minus the loss formulation, as before.

The difference becomes startling when saturation occurs. [Compare Figs. 8.12(c) with 8.3(c).] The fact that *we* know that the laboratory line shape is inhomogeneously broadened is quite irrelevant—nobody informed the atoms or the field of this fact. The field interacts with a specific group according to the *homogeneous* line shape of that group in the atom's frame of reference and not the laboratory line shape that we know.

Thus the $(0, 0, q)$ mode interacts most strongly with that group of atoms that "think" that $\nu_{0,0,q}$ is the center frequency of a homogeneous line-shape function $g_h(\nu)$ for the gas. If the homogeneous line width $\Delta\nu_h$ is much less than the Doppler width, then this interaction

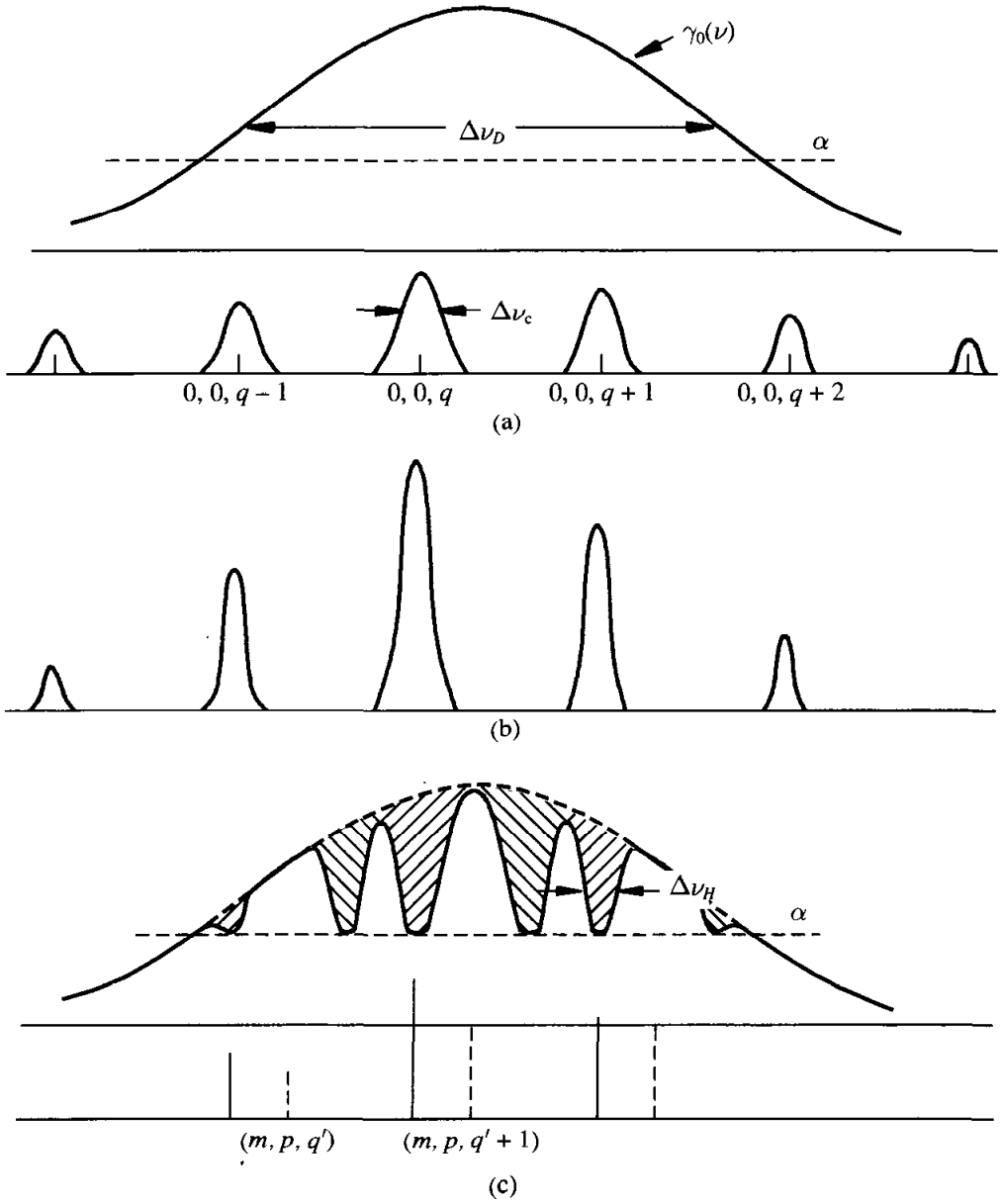


FIGURE 8.12. Evolution of oscillation in a Doppler-broadened transition.

occurs over a very limited portion of the laboratory line shape, leaving the remainder unaffected by the (0, 0, q) mode.

The same considerations apply to the stimulating waves at $\nu_{0,0,q-1}$ and $\nu_{0,0,q+1}$, which have positive values for gain minus loss. Each will burn a separate hole in the gain profile until the gain at each frequency saturates at the loss. As before, the cavity response narrows around the center resonance frequency until the spectral representation is nearly a delta function. In a loose sort of a way, the power in each mode is proportional to the area "burned" away in forming the hole.

For Doppler broadening, there is an additional complication that has most interesting consequences. Recall that the electromagnetic energy consists of fields running back and forth between the mirrors, obviously in opposite directions for a simple cavity. If ν_q is less than ν_0 (in the laboratory frame), the wave that travels in the positive z direction

interacts most strongly with atoms that have enough velocity in the opposite (or negative z) direction to make ν_q equal to ν_0 in their frame of reference. When the wave turns around and propagates in the negative z direction, it now interacts most strongly with atoms moving in the positive z direction. There are two distinct groups of atoms that are stimulated by one field, and consequently both contribute their energy to the same laboratory frequency. These two frequencies are images about the line center ν_0.

One of the most dramatic demonstrations of this effect occurs in a very short laser such that only one cavity mode has a gain-to-loss ratio greater than 1. This is shown in Fig. 8.13. By changing d slightly, we can tune the cavity mode across the line. This is usually done by mounting the mirror on a piezoelectric element and applying a sawtooth voltage to the electrodes. That voltage is shown on the lower trace of the oscilloscope picture. The upper trace is the power out of the laser. The letters at the bottom of the picture correspond to the cavity mode appearing at the corresponding figures above.

As expected, when the cavity mode is on wings (a and d) of the transition, the power is low; as it is tuned toward the line center—say b—the power increases as the area burned off of the gain curve increases. When, however, the hole at the oscillation frequency overlaps the image hole, the power decreases. It makes no difference to those atoms in the overlap region whether it gives up its energy in the form of a photon to the positive or the negative traveling wave. It only has *one* package of energy, $h\nu$, to give up and no more. Consequently the power drops, reaching a minimum at line center. This phenomenon is referred to as the "Lamb dip" after W. E. Lamb [8], who predicted its occurrence on theoretical grounds. It has the very practical application of enabling one to stabilize the oscillation frequency to the center of a transition with extreme precision.

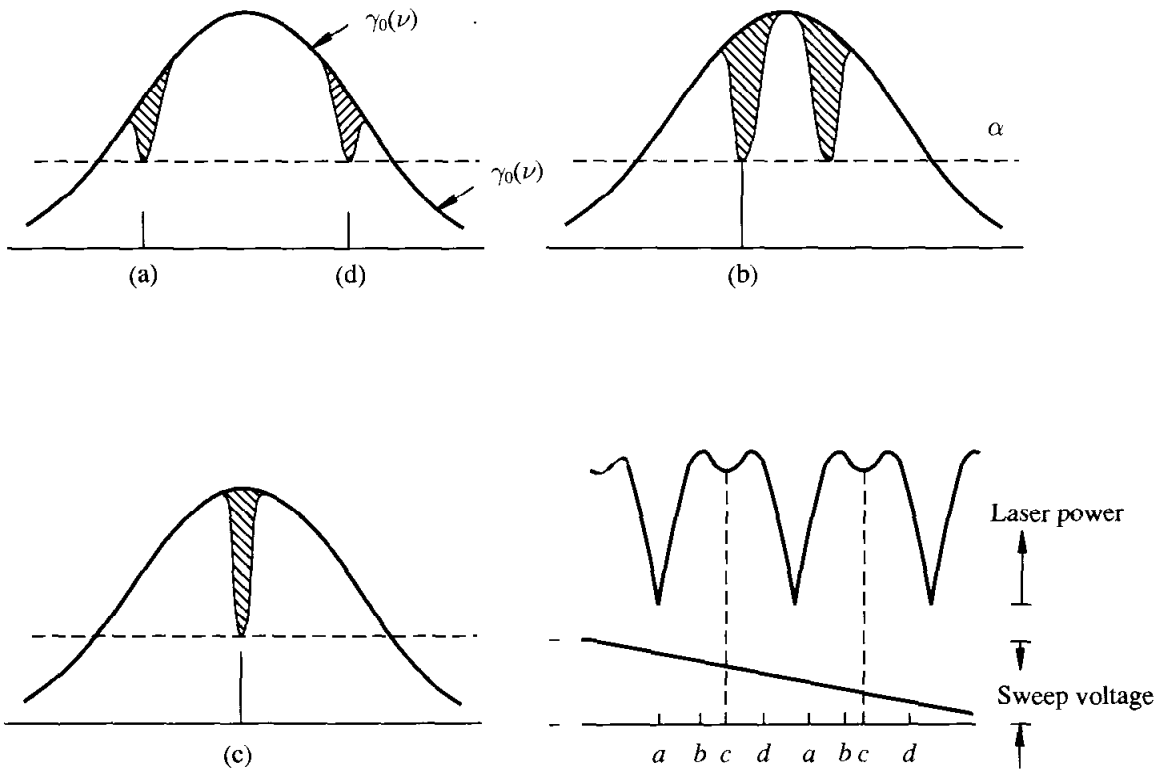


FIGURE 8.13. Lamb dip. (Courtesy of Spectra-Physics.)

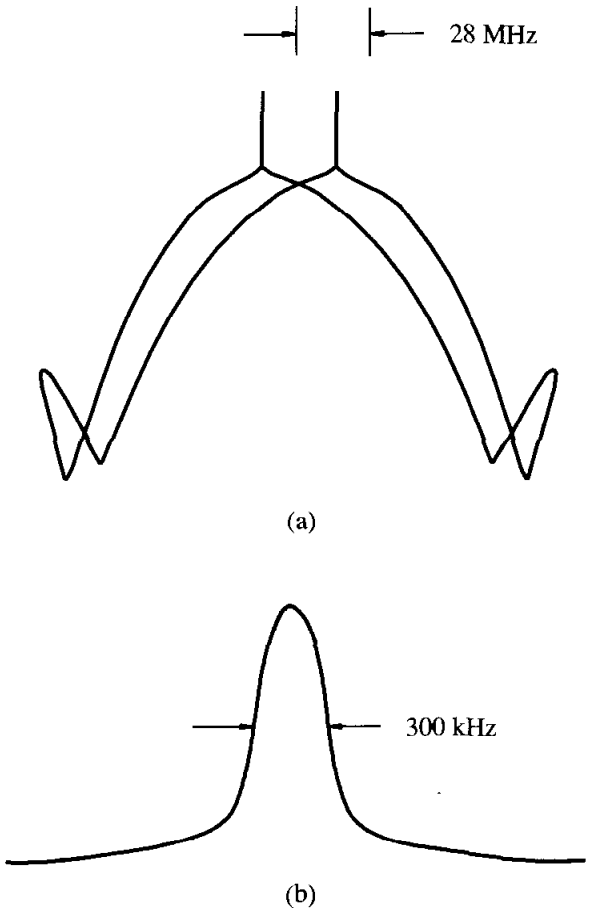


FIGURE 8.14. (a) Output power of a He:Ne laser at 3.39 μm in the near vicinity of the absorption of CH_4; (b) expanded frequency scale for the peak. (From R. L. Barger and R. L. Hall, *Phys. Rev. Lett.* 22, 4, 1969.) NOTE: Ref. 21 has identified the *frequency* of the peak to occur at $\nu = 88.3761816029$ THz ± 1.2 kHz.

Just as atoms have only one unit of energy, $h\nu$, which can be used by stimulated emission, they can also only absorb the same unit from either wave. This absorption can be saturated (i.e., reduced) just as "the amplification" in the discussion above. There is a fortuitous coincidence between the He:Ne transition at 3.39 μm and a vibrational-rotational transition in methane (CH_4). Thus, if a low-pressure methane cell is incorporated inside the laser cavity, there will be an inverted Lamb dip when the laser frequency coincides with the center frequency of the methane absorption: that is, the power reaches a maximum at the center. Because of the beautiful symmetry of the methane molecule, the saturated absorption is very narrow, on the order of 100 kHz or so. Such a scheme is easily reproduced, and a laser stabilized to this peak has become the international standard for length. An example of the output of a laser as it is tuned across the absorption line, taken from the work of R. L. Barger and J. L. Hall [17] of the National Bureau of Standards, is shown in Fig. 8.14.

8.5 MULTIMODE OSCILLATION

It is obvious from Fig. 8.12 that a laser can oscillate on more than one $TEM_{0,0,q}$ mode. A few numerical values are in order to illustrate this point. Consider the He:Ne transition at $\lambda_0 = 632.8$ nm, which is Doppler broadened with $\Delta\nu_D \sim 1.5$ GHz. A free-spectral range (the separation between $\nu_{0,0,q}$ and $\nu_{0,0,q-1}$ is $c/2d$, and, for a nominal mirror spacing

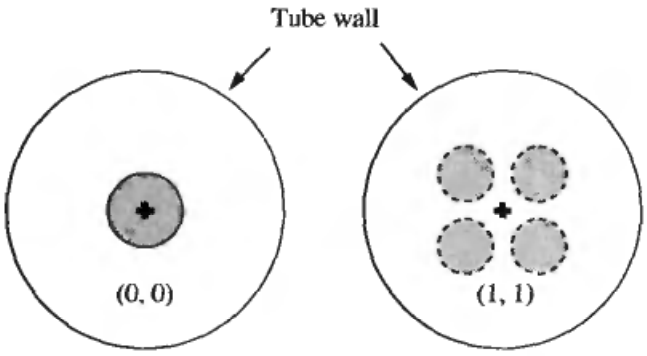


FIGURE 8.15. Dot patterns of (0, 0) and (1, 1) mode.

of $d = 1$ m, is 150 MHz. Thus as many as 10 different frequencies can be oscillating simultaneously on a single transition, assuming that $\gamma_0(\nu_0)/\alpha = 2$.

The situation becomes even more involved when the transverse modes are considered. Not only are the resonant frequencies of the $\text{TEM}_{m,p,q}$ modes different from those of the $(0, 0, q)$ mode—and thus still have net gain at their frequency—but the fields of these higher-order extend over a larger cross section and thus interact with different spatial groups of excited atoms.

This last point is illustrated in Fig. 8.15, where the "dot" patterns of the (0, 0) and the (1, 1) modes are shown. It is obvious that the (0, 0) mode interacts most strongly with the atoms near the axis of the tube, whereas the (1, 1) mode uses the atoms located at a distance away from the axis. Thus even a homogeneously broadened transition can support multimode oscillation by "spatial" hole burning.

One can suggest a situation whereby the field of one (m,p,q') mode competes with the field of another mode for the same group of atoms at the same location in the gain medium. Furthermore it is highly probable that the gain medium is *not* uniform in r (or ϕ). To handle this situation—even approximately—is not a trivial task and will be left to more advanced texts.

8.6 GAIN SATURATION IN DOPPLER-BROADENED TRANSITION: MATHEMATICAL TREATMENT

We have avoided any mathematics in the two preceding sections so as to gain a physical feeling for the saturation process. It is now time to put these feelings on a more quantitative basis.

The key points to recall are (1) the field interacts with the atoms according to the homogeneous line shape in the atom's rest frame and (2) the inhomogeneous line shape as observed in the laboratory is a weighted sum between the atom's homogeneous line shape and the probability that this group will occur.

Let $p(f)\,df$ be the fraction of the inversion density whose center frequency (in the laboratory frame) coincides with the frequency ν. This group of atoms interacts with the laser frequency at ν according to the homogeneous line-shape function and saturates according to the homogeneous law derived previously (8.3.12). Thus the saturated gain coefficient for

an inhomogeneously broadened transition is given by summing over all fractions.

$$\gamma(\nu, I_\nu) = A_{21} \frac{\lambda_0^2}{8\pi n^2} \left(N_2^0 - \frac{g_2}{g_1} N_1^0 \right) \int_0^{+\infty} \frac{p(f) g_h(\nu, f)\, df}{1 + (I_\nu / I_s)\overline{g}_h(\nu, f)} \tag{8.6.1}$$

where

$$\int_0^{+\infty} p(f)\, df = 1 \tag{8.6.2}$$

$$I_s = \frac{h\nu}{(\tau_1 + \tau_2 - \tau_1/\tau_2/\tau_{21})\sigma(\nu_0)} \tag{8.3.12}$$

$\overline{g}_h(\nu, f)$ = the Lorentzian normalized to 1 at $f = \nu$

or

$$\overline{g}_h(\nu, f) = \frac{\pi \Delta\nu_h}{2} g(\nu, f) = \frac{(\Delta\nu_h/2)^2}{(f - \nu)^2 + (\Delta\nu_h/2)^2} \tag{8.6.3}$$

N_2^0, N_1^0 = density of states 2,1 in the limit of *zero* amplitude of I_ν

To be specific, let us assume a thermal distribution of velocities that translates into a Doppler distribution of frequencies.

$$p(f) = \left(\frac{4 \ln 2}{\pi} \right)^{1/2} \frac{1}{\Delta\nu_D} \exp\left[-4(\ln 2) \left(\frac{f - \nu_0}{\Delta\nu_D} \right)^2 \right] \tag{7.6.18}$$

Anyone who attempts to substitute (8.6.3), (8.3.12c), and (7.6.18) into (8.6.1) is bound to be overwhelmed by the shear volume of symbols and letters. Let us at least keep $p(f)$ as an integral part of the expression and proceed carefully. If we examine (8.6.1), we see $g_h(\nu, f)$ and $\overline{g}_h(\nu, f)$ appearing in the numerator and denominator; hence, there will be some cancellation of common factors if numerator and denominator are multiplied by $[(\nu - f)^2 + (\Delta\nu/2)^2]$:

$$\gamma(\nu, I_\nu) = A_{21} \frac{\lambda_0^2}{8\pi n^2} \left(N_2^0 - \frac{g_2}{g_1} N_1^0 \right) \frac{\Delta\nu_h}{2\pi} \times \int_0^{\infty} \frac{p(f)\, df}{(f - \nu)^2 + (\Delta\nu_h/2)^2 (1 + I_\nu/I_s)} \tag{8.6.4}$$

Amazingly, the denominator has a Lorentzian frequency dependence with an intensity-dependent line width. Defining this last factor as a "hole width,"

$$\Delta\nu_H^2 = \Delta\nu_h^2 \left(1 + \frac{I_\nu}{I_s} \right) \tag{8.6.5}$$

(8.6.4) becomes much more palatable with the identification of a Lorentzian with an intensity-dependent width defined by (8.6.5). Rearranging (8.6.4) according to this logic leads us to

$$\gamma(\nu, I_\nu) = A_{21}\frac{\lambda_0^2}{8\pi n^2}\left(N_2^0 - \frac{g_2}{g_1}N_1^0\right)\frac{\Delta\nu_h}{\Delta\nu_H}\int_0^\infty p(f)\frac{\Delta\nu_H}{2\pi\left[(f-\nu)^2 + (\Delta\nu_H/2)^2\right]}\,df \tag{8.6.6}$$

We cannot go further without some sort of approximation procedure or a resort to a numerical evaluation of this integral. We take the former approach and make the same type of an approximation as was done in Sec. 7.6.2 [see 7.6.15 and 7.6.16]; namely, we approximate a Lorentzian in the integrand by a delta function:

$$\frac{\Delta\nu_H/2\pi}{(f-\nu)^2 + (\Delta\nu_H/2)^2} \to \delta(f-\nu) \tag{8.6.7}$$

Then the integration of (8.6.6) is trivial:

$$\gamma(\nu, I_\nu) = \left[A_{21}\frac{\lambda^2}{8\pi}\left(N_2^0 - \frac{g_2}{g_1}N_1^0\right)p(\nu)\right]\frac{\Delta\nu_h}{\Delta\nu_H} \tag{8.6.8}$$

The quantity in brackets is the small-signal gain coefficient, whereas the last term contains the effect of saturation:

$$\gamma(\nu, I_\nu) = \gamma_0(\nu)\frac{\Delta\nu_h}{\Delta\nu_H} \tag{8.6.9}$$

The ratio $\Delta\nu_h/\Delta\nu_H$ comes from (8.6.5) and can be expressed as

$$\frac{\Delta\nu_h}{\Delta\nu_H} = \frac{1}{(1 + I_\nu/I_s)^{1/2}} \tag{8.6.10}$$

Hence, the gain saturates according to

$$\gamma(\nu, I_\nu) = \frac{\gamma_0(\nu)}{(1 + I_\nu/I_s)^{1/2}} \tag{8.6.11}$$

Note that the saturation intensity as it appears in (8.6.11) is still given by $h\nu/\sigma(\nu_0)\tau_2$, where the line-width factor appearing in $\sigma(\nu_0)$ is $\Delta\nu_h$ (i.e., the homogeneous width) not $\Delta\nu_D$. (8.6.11) predicts that the gain of a Doppler-broadened transition saturates with a functional dependence that is slower than that of a homogeneous one. This is because the width of the hole burned in the inhomogeneous gain profile is changing simultaneously with the depth of the hole (or gain depression). In a homogeneous line, the width does not change and the profile of the line is maintained while the gain is depressed.

For the Doppler case, the saturation factor $(1 + I/I_s)^{-1/2}$ is frequency independent, where the normalized Lorentzian appears in (8.3.15). Whereas the gain coefficient in a homogeneously broadened transition is hard to saturate by stimulated emission in the wings (where $|\nu - \nu_0| > \Delta\nu/2$), the Doppler-broadened case can be saturated to the same degree with the one intensity over the band. Actually, however, the distinction is minor because seldom is an inhomogeneously broadened medium used as a power amplifier in the manner indicated in Fig. 8.7.

We should be very careful about blindly applying this saturation law for all values of the intensity. If the stimulating field is high enough, then the approximation that the intensity-dependent Lorentzian hole width is much narrower than that of the inhomogeneous line, $p(f)$, breaks down. In the intermediate case, the arithmetic complexity is overwhelming,

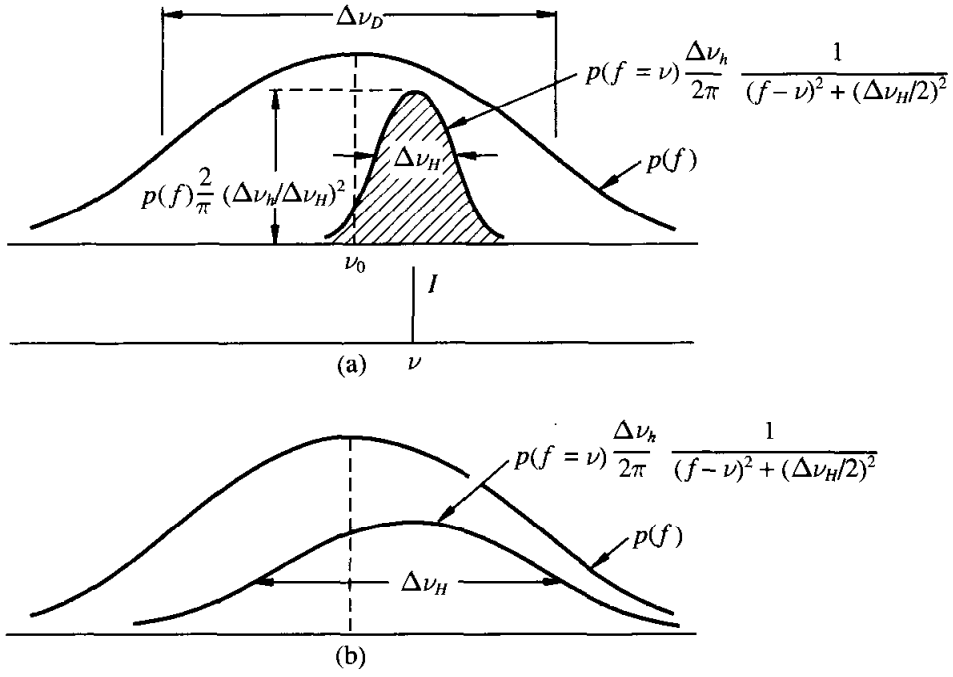


FIGURE 8.16. Saturation of an inhomogeneously broadened amplifier: (a) intermediate intensity; (b) high intensity.

so let us turn to a picture to see the physics of the situation. Fig. 8.16 shows the status.

If the hole width is much larger than the inhomogeneous width $\Delta\nu_D$ because of the intensity or because of the large value of homogeneous width $\Delta\nu_h$, then the Lorentzian can be pulled outside the integral in (8.6.6)

$$\gamma(\nu, I_\nu) = A_{21}\frac{\lambda_0^2}{8\pi n^2}\left(N_2^0 - \frac{g_2}{g_1}N_1^0\right)\frac{\Delta\nu_h}{2\pi\left[(\nu_0-\nu)^2+(\Delta\nu_H/2)^2\right]}\int_0^\infty p(f)\,df \tag{8.6.12}$$

where we have substituted the mean value of $f = \nu_0$ into the Lorentzian function.

Since the last integral is unity, we recover the prior formula for gain saturation in a homogeneous medium:

$$\gamma(\nu, I_\nu) = A_{21}\frac{\lambda_0^2}{8\pi n^2}\left(N_2^0 - \frac{g_2}{g_1}N_1^0\right)\frac{\Delta\nu_h/2\pi}{(\nu-\nu_0)^2+(\Delta\nu_h/2)^2(1+I_\nu/I_s)} \tag{8.6.13}$$

Rewriting makes the results more familiar:

$$\gamma(\nu, I_\nu) = \left[A_{21}\frac{\lambda_0^2}{8\pi n^2}\left(N_2^0 - \frac{g_2}{g_1}N_1^0\right)\frac{\Delta\nu_h/2\pi}{(\nu-\nu_0)^2+(\Delta\nu_h/2)^2}\right] \times \left[\frac{1}{1+\dfrac{I_\nu}{I_s}\dfrac{(\Delta\nu_h/2)^2}{(\nu-\nu_0)^2+(\Delta\nu_h/2)^2}}\right] \tag{8.6.14}$$

The first set of brackets is just $\gamma_0(\nu)$, whereas the second set can be rewritten using the definition of $\overline{g}_h(\nu)$:

$$\gamma(\nu, I_\nu) = \frac{\gamma_0(\nu)}{1 + (I_\nu/I_s)\overline{g}_h(\nu)} \qquad (8.3.15) \rightarrow (8.6.15)$$

It is important to realize that physics *demands* this limit in the extreme of very large intensities. As was argued previously (8.3.20) we can extract power from an optical system only at the effective rate that we can pump it. In other words, we can get out only as much as we put in.

8.7 AMPLIFIED SPONTANEOUS EMISSION (ASE)

Although the primary emphasis of this chapter (and most of the book) has been on laser oscillation, it should be pointed out that the energy represented by the population inversion $h\nu[N_2 - (g_2/g_1)N_1]$ can be extracted as broad-band radiation and can be quite intense. This can be a blessing or a curse. As we will see, this enables us to generate intense radiation without an optical cavity, but it also limits the amount of gain that we can design into an amplifier.

Consider the situation shown in Fig. 8.17, with an amplifier being pumped by an external source. Although no externally injected signal is shown there, the atoms in state 2 still radiate spontaneously into the frequency interval that matches the gain profile of the remaining part of the amplifier.

Let us focus on the radiation $I^+(\nu, z)$ traveling in the positive z direction and contained within a constant solid angle $d\Omega/4\pi$. The atoms in N_2 at $z = 0$ radiate part of their energy spontaneously into this solid angle; and this radiation, in turn, is amplified by the inversion between $z = 0$ and $z = l_g$. Thus spontaneous emission is continuously added to I^+ along z, and, simultaneously, stimulated emission amplifies the power from previous lengths. (Remember, this is incoherent radiation, and thus powers rather than fields must be added.)

$$\frac{d}{dz}\left[I^+(\nu, z)\, d\nu\right] = \gamma_0(\nu) I^+(\nu, z)\, d\nu + h\nu A_{21} N_2 g(\nu)\, d\nu \frac{d\Omega}{4\pi} \qquad (8.7.1)$$

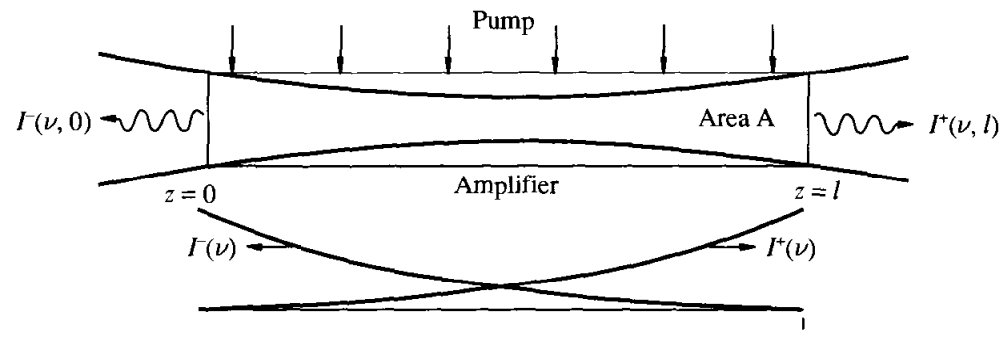


FIGURE 8.17. Optical amplifier generating broad-band incoherent radiation.

A solution to this linear first-order differential equation subject to the obvious boundary condition that $I^+(\nu, z = 0) = 0$ is readily obtained:

$$I^+(\nu, z = l_g) = \frac{h\nu A_{21} N_2 g(\nu)}{\gamma_0(\nu)} \left(e^{\gamma_0(\nu) l_g} - 1\right) \frac{d\Omega}{4\pi} \tag{8.7.2}$$

If we relate the small-signal gain coefficient $\gamma_0(\nu)$ to the population inversion and the A coefficient, we obtain a very important formula:

$$I^+(\nu, l_g) = \frac{8\pi n^2 h\nu^3}{c^2} \frac{N_2}{N_2 - (g_2/g_1)N_1} \left[G_0(\nu) - 1\right] \frac{d\Omega}{4\pi} \tag{8.7.3a}$$

where $G_0(\nu) = \exp[\gamma_0(\nu) l_g]$ is the small-signal gain of the amplifier. This formula applies equally well to an absorptive system with $N_2 < (g_2/g_1)N_1$ and $G_0 < 1$ (i.e., the cell is an attenuator).

Equation (8.7.3a) is very important, and it is worthwhile to digress for a moment to appreciate its significance.

• • •

Case A: An Optically "Thin" Amplifier or Attenuator. If $G_0(\nu)$ is very close to 1, the amplifier (or attenuator) is said to be optically thin, and thus $\gamma_0(\nu) l_g$ is small. Therefore the Taylor series expansion of $\exp(\gamma_0 l_g) - 1$ yields $\gamma_0(\nu) l_g$, and we obtain a most logical result:

$$I^+(\nu, l_g) = A_{21} h\nu N_2 l_g g(\nu) \frac{d\Omega}{4\pi} \qquad \text{(optically thin)} \tag{8.7.3b}$$

This states that the power from $N_2 l_g$ atoms radiating into $d\Omega/4\pi$ as $g(\nu)$ add their radiation, a result that would be guessed from the start. In other words, each element dz along z contributes an equal amount to the power.

Case B: A Thermal Population. If the atomic populations are such that $N_2 < (g_2/g_1)N_1$, the amplifier is an attenuator and $G_0 < 1$. Furthermore, if N_2/N_1 can be related to a "temperature" by a Boltzmann relation,

$$\frac{N_2}{N_1} = \frac{g_2}{g_1} \exp\left(-\frac{h\nu}{kT}\right)$$

then (8.7.3a) becomes

$$I^+(\nu, l) = \left[\frac{8\pi n^2 h\nu^3}{c^2} \frac{1}{\exp(h\nu/kT) - 1}\right] \frac{d\Omega}{4\pi} \left(1 - e^{-|\gamma_0(\nu)| l_g}\right) \tag{8.7.3c}$$

We should immediately recognize the quantity in the brackets as the Planck formula for blackbody radiation for $I(\nu) = (c/n_g)\rho(\nu)$ [see (7.2.10)]. The remaining factor, $1 - \exp[-\gamma_0(\nu) l_g|]$ is a measure of the blackbody power available to the outside world. Its maximum value is, of course, unity for an optically thick medium (i.e., $\exp[-|\gamma_0(\nu) l_g|] \ll 1$).

Thus we see that the quantity in parentheses in (8.7.3c) is the *emissivity* of the system for a normal population ratio. Since the factor $1 - \exp[-|\gamma_0(\nu) l|_g]$ is also the absorption

of a wave passing through the attenuator, we have thus derived Kirchoff's radiation law, which states that a body can emit only as much blackbody radiation at a frequency ν as it can absorb.

If the lower-state population increases with z, as it does, for instance, in a high-pressure sodium lamp (with z treated as the radius), the central part of the spontaneous emission is heavily attenuated, whereas the wings escape more or less unimpeded. Under such circumstances we can easily obtain a *self-reversed* line. (Observe a common sodium vapor street lamp with a small hand-held spectroscope. We see a broad orange spectrum extending on either side of 589 nm, but a "dark" band at this wavelength, which is the *center* of sodium emission. This is a common example of a self-reversed line.)

• • •

Let us now consider some of the consequences of (8.7.3a). First, we should remind ourselves that it was derived under the assumption that the populations N_2 and N_1 were *not* saturated by stimulated emission. Under such circumstances and high gain, the spectral width of the amplified spontaneous emission is narrower than that predicted by the line shape $g(\nu)$ as given, for instance, by (8.7.3b).

A few numbers should convince us of this fact. Suppose that $G_0(\nu_0)$ (i.e., the gain at line center) were 100. Then the factor $G_0(\nu_0) - 1 = 99$. At $\nu = \nu_0 \pm \Delta\nu/2$, $\gamma(\nu) = \gamma_0(\nu_0)/2$, and thus $G_0(\nu) = [G_0(\nu_0)]^{1/2} = 10$. Thus this factor is now equal to 9, a reduction by a factor of 11 in changing the frequency by a mere $\Delta\nu/2$. Thus spectral narrowing is to be expected from a system with high gain. A graphic display of this is shown in Fig. 8.18, where (8.7.3a) is plotted for various values of the gain coefficient (times length) with an assumed Lorentzian line shape for $g(\nu)$.

We cannot carry this analysis to the extreme of letting the intensity become arbitrarily large. If the line is inhomogeneously broadened, then too large an intensity at line center will burn a hole at line center but leave the wings unaffected. When this happens, the line width begins to expand back toward its optically thin value, and the arithmetic becomes most unbearable. Equation (8.7.1) used the small-signal gain coefficient, whereas now one would need to include saturation.

A much more serious problem arises in addition to our unhappiness with the mathematics. If the amplified spontaneous emission (ASE) saturates the population inversion, that inversion energy *cannot* be extracted by an externally injected coherent laser signal. Thus the curse of ASE is that it limits the maximum gain that can be built into an amplifier.

To quantify this limit, we generalize the rate equation in Sec. 8.3 to account for a spectral distribution of radiation $I(\nu)$ rather than I_ν in (8.3.3) through (8.3.9b). This amounts to replacing factors of the form $\sigma(\nu)I_\nu$ by $\int \sigma(\nu')I(\nu')\,d\nu'$ (the details are left for a problem). The result is that the gain saturates according to

$$\gamma(\nu) = \frac{\gamma_0(\nu)}{1 + \left[\tau_2 + \left(\frac{g_2}{g_1}\right)\tau_1\left(1 - \frac{\tau_2}{\tau_{21}}\right)\frac{\sigma(\nu_0)}{h\nu}\int \bar{g}(\nu')\left[I^+(\nu') + I^-(\nu')\right]d\nu'\right]} \tag{8.7.4}$$

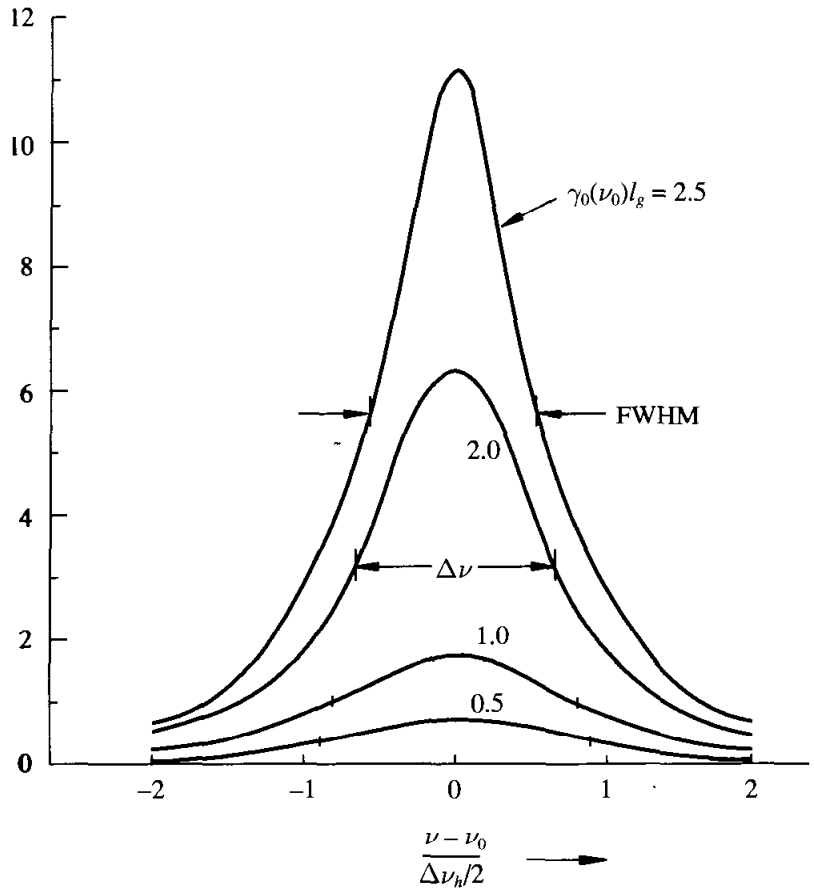


FIGURE 8.18. Spectral narrowing of spontaneous emission in an unsaturated amplifier.

Next we use the result of the other viewpoint expressed in Sec. 8.3; namely, that the saturation intensity can also be interpreted as that intensity that reduces the lifetime of state 2 by a factor of 2 (for a good laser with $\tau_1 \ll \tau_2$). For a spectral distribution of intensity, the lifetime of state 2 is reduced by a factor of 2 when

$$\frac{\sigma(\nu_0)}{h\nu} \int \overline{g}(\nu') \left[I^+(\nu', z) + I^-(\nu', z) \right] d\nu' = \frac{1}{\tau_2} \tag{8.7.5}$$

Saturation will occur first at $z = l_g$ or $z = 0$ of the amplifier of Fig. 8.17. If the condition expressed by (8.7.5) holds, further pumping of the amplifier is mostly futile; that is, stimulated emission by ASE extracts the energy from the inversion as fast as the pumping agent can create it.

Let us consider $z = l_g$ where $I^-(\nu, l_g) = 0$ and evaluate (8.7.5) with $I^+(\nu, l_g)$ given by (8.7.3).

$$\frac{\sigma(\nu_0)}{h\nu} \int \overline{g}(\nu') \frac{8\pi n^2 h\nu^3}{c^2} \frac{N_2}{N_2 - (g_2/g_1)N_1} \left[e^{\gamma_0(\nu_0) l_g \overline{g}(\nu')} - 1 \right] \frac{d\Omega}{4\pi} d\nu' = \frac{1}{\tau_2} \tag{8.7.6a}$$

All frequency variations expressed in (8.7.6a) are insignificant compared to the violent variation of $\overline{g}(\nu)$ appearing in the exponential and as a multiplier. Thus the center frequency, $\nu = \nu_0$, is dominant in the exponential where $\overline{g}(\nu_0) = 1$ and (8.7.6a) becomes

$$e^{\gamma_0(\nu_0) l_g} - 1 = \frac{\lambda_0^2}{8\pi n^2} \cdot \frac{1}{\sigma(\nu_0)} \left[\frac{N_2 - (g_2/g_1)N_1}{N_2} \right] \frac{1}{\tau_2} \cdot \left(\frac{d\Omega}{4\pi} \right)^{-1}$$

or

$$G_0(\nu_0) - 1 < \frac{\tau_{21}}{\tau_2}\left(\frac{d\Omega}{4\pi}\right)^{-1}\left(1 - \frac{g_2}{g_1}\frac{N_1}{N_2}\right) \tag{8.7.6b}$$

where τ_{21} is used instead of A_{21} in the expression for $\sigma(\nu_0)$. We can estimate $d\Omega/4\pi$ by the solid angle subtended by the aperture A on a sphere centered at $z = 0$, or $d\Omega/4\pi \sim A/4\pi l_g^2$. For a good laser medium, $N_1 \ll N_2$ and $\tau_2 \sim \tau_{21}$, and thus the maximum permissible gain is dictated by geometry:

$$G_0(\nu_0) - 1 = \frac{4\pi l_g^2}{A} \tag{8.7.7}$$

Although this derivation suffers from the approximations used, it gives a good physical picture of the limitations that spontaneous emission imposes on the amplifier design. But this limitation is on the small-signal gain of the system and does not limit the amount of energy that can be extracted. As shown in Sec. 8.3, we can extract as much power as the effective pumping rate can supply with the proviso that ASE does not use it first. Thus, for many high-energy applications such as laser fusion, one prefers a *low gain* but high energy storage capability in the amplifiers.

8.8 LASER OSCILLATION: A DIFFERENT VIEWPOINT

Let us close this chapter with a review of the basic concepts and simultaneously obtain a slightly different picture of laser oscillation. Incidentally, these considerations also apply to classical oscillators, such as microwave tubes and transistor circuits.

We start with a simple system, such as that shown in Fig. 8.1, with an active medium, characterized by a small-signal gain $G_0 = \exp[\gamma_0(\nu)l_g]$ and a suitable feedback system in the form of a cavity with mirror reflectivities R_1 and R_2. As discussed previously, the laser oscillation builds up from the spontaneous emission "noise" emitted from the upper state until the coherent photon flux saturates the gain. We can describe this buildup and the necessity of gain saturation with a few lines of transparent mathematics.

The fact that a population inversion exists automatically guarantees a noise source in the form of spontaneous emission into 4π steradians and into a whole band of frequencies in proportion to the line shape $g(\nu)\,d\nu$. What we need is a precise description of how these noise photons are allotted to a high-Q cavity mode which will eventually become the intense laser radiation. We obtain this prescription by the following line of reasoning.

Consider the high-Q $\text{TEM}_{0,0,q}$ modes of the laser cavity. These modes are separated by a frequency interval $\Delta\nu = c/2nd$, and the fields exist over an effective "area" that overlaps the active medium. (As we will see, it is not necessary to be precise about this area, but we must recognize that it exists.) Thus the number of active atoms in state 2 capable of being stimulated by these modes is $N_2 \times \text{area} \times \text{length} = N_2V$. Before stimulated emission becomes important, we must obtain the initial photons from the noise. We note that

1. The rate of generation of spontaneous photons by this group of atoms into all frequencies is $A_{21}N_2V$.

2. The fraction that appears in the frequency interval $c/2nd$, symmetrically placed around ν_q is $g(\nu)\Delta\nu$ (with $\Delta\nu = c/2nd$).
3. Only the $\text{TEM}_{0,0,q}$ mode has a high Q, but there are $(8\pi n^3\nu^2/c^3)V \cdot (\Delta\nu = c/2nd)$ electromagnetic modes in that volume and in that frequency interval.

Thus the rate of increase of photons in this one $\text{TEM}_{0,0,q}$ cavity mode caused by spontaneous emission is

$$\left.\frac{dN_p}{dt}\right|_{\text{spont.}} = (A_{21}N_2V)\left[g(\nu)\frac{c}{2nd}\right]\frac{1 \text{ mode}}{\text{no. modes} = (8\pi n^2\nu^2/c^3)(c/2nd)V} \tag{8.8.1}$$

where N_p is the number of photons in the cavity mode volume. Canceling common factors of V and $c/2nd$ and recognizing the remaining collection of factors as belonging to the stimulated emission cross section lead to a compact formula:

$$\left.\frac{dN_p}{dt}\right|_{\text{spont.}} = N_2 c\left[A_{21}\frac{\lambda^2}{8\pi}g(\nu)\right] = N_2 c\sigma_{\text{SE}} \tag{8.8.2}$$

Once the photons are in the cavity mode, they bounce back and forth between the two mirrors, being amplified by a factor G per pass, with some of them escaping by transmission through (or absorption by) the mirrors. If N_p starts at an arbitrary point inside the cavity, then $GR_1GR_2N_p$ returns after a round trip taking $2nd/c$ seconds. Thus the time rate of change of photon number caused by stimulated emission in the active cavity is given by

$$\left.\frac{dN_p}{dt}\right|_{\text{cavity with gain}} = \frac{G^2R_1R_2 - 1}{2nd/c}N_p \tag{8.8.3}$$

The addition of (8.8.2) and (8.8.3) yields this simple mathematical model:

$$\frac{dN_p}{dt} = \frac{G^2R_1R_2 - 1}{2nd/c}N_p + N_2 c\sigma_{\text{SE}} \tag{8.8.4}$$

If $G^2R_1R_2 - 1 < 0$, then we have a very uninteresting situation. The equilibrium value of N_p is enhanced by the active cavity; but even so, the output power is insignificant (see the problem at the end of this section). Now the last term in (8.8.4) is quite small, and a few numbers will convince us of that fact. Let $\sigma = 3 \times 10^{-12}$ cm^2 (a very large stimulated emission cross section), $c = 3 \times 10^{10}$ cm/sec (as usual), and $N_2 = 10^{12}$ cm^{-3} (typical of such a high-gain laser). Then the rate of spontaneous emission into this cavity mode sounds big, $N_2 c\sigma_{\text{SE}} = 9 \times 10^{10}$ sec^{-1}, but if each photon has 1 eV of energy, this represents a paltry 14.4 nanowatts.

But if $G^2 > 1/R_1R_2$, the first term on the right-hand side of (8.8.4) is a *positive* coefficient of N_p, and the photon number grows exponentially with time. As long as N_p stays small enough so that G can be considered a constant, the number *grows* exponentially with time.

$$N_p(t) \doteq N_p(0)\exp\left[+\left(\frac{G^2R_1R_2 - 1}{2nd/c}\right)t\right] \tag{8.8.5}$$

But this equation cannot apply for all time—*we* cannot avoid saturation! As long as $G^2R_1R_2 - 1 > 0$, the photon number increases with time and the only way to reverse that behavior is for the sign of the first term in (8.8.4) to change (i.e., revert back to a negative sign). Thus the gain *must* saturate slightly *below* the loss so that the two terms on the right-hand side of (8.8.4) are equal and then we have a steady-state laser.

For any computational purpose, we can safely set $G_{\text{sat.}}^2 = 1/R_1R_2$. For instance, suppose that the laser just described produces 10 mW of power, has a length of 50 cm, and has mirrors with reflectivities $R_1 = 1$ and $R_2 = 0.9$. From the specification of the output power, we can compute the power impinging on the mirrors, which, when divided by $h\nu$, yields the number of photons hitting M_1 or M_2 (per second). This number is also equal to the number of photons inside the cavity divided by the round trip transit time $2nd/c$ (since each of the photons hits each mirror in that time interval).

$$\frac{N_p}{2nd/c} = \frac{P_{\text{out.}}}{h\nu}\frac{1}{1 - R_1R_2} \tag{8.8.6}$$

Thus we can compute the saturated gain, $G_{\text{sat.}}$, by requiring $dN_p/dt = 0$ in (8.8.4). We must also recognize that the upper-state population is saturated by the laser oscillation and thus one must use the saturated value for N_2.

$$(1 - G_s^2R_1R_2)\frac{P_0}{h\nu}\frac{1}{1 - R_1R_2} = N_2^{(s)}c\sigma_{\text{SE}} \tag{8.7.7}$$

From the numbers chosen, we obtain

$$\begin{aligned} 1 - G_s^2R_1R_2 &= \frac{h\nu}{P_0}(1 - R_1R_2)N_2^{(s)}c\sigma_{\text{SE}} \\ &< 1.44 \times 10^{-7} \end{aligned} \tag{8.8.8}$$

Surely, then, setting the saturated gain equal to the loss is an excellent approximation!

Even though spontaneous emission is ignorable in the final state of laser oscillation, it is crucial to its initiation. As we shall see in Chapter 9, there are some laser media where the spontaneous emission is so low that it takes appreciable time for the laser pulse to develop. Indeed, the spontaneous emission is so low that the population inversion can buildup faster than the photon number. This will lead to a "gain-switched" pulse.

The fact that the gain does saturate at a value slightly less than the loss (for this or any oscillator) implies that the laser will have a finite, nonzero, spectral width caused by the "noise" contributed by the spontaneous emission. We can compute this width by recognizing that the radiation is distributed in frequency around ν_q according to the Fabry–Perot function for the *active* cavity. Modifying (6.3.1) to account for the saturated gain G_s leads to the following expression for the spectral distribution of power *inside* the laser cavity for $n = 1$.

$$P(\nu) = \frac{K}{\left[1 - G_s(R_1R_2)^{1/2}\right]^2 + 4G_s(R_1R_2)^{1/2}\sin^2\left[2\pi(\nu - \nu_q)\,d/c\right]} \tag{8.8.9}$$

We now use a whole sequence of tricks to derive various formulas for the spectral width. The first is a straightforward analysis to obtain the full width at half maximum of the saturated

cavity response function from (8.8.9).

$$\Delta\nu_{\text{osc.}} = \frac{1 - G_s(R_1R_2)^{1/2}}{\pi\left[G_s(R_1R_2)^{1/2}\right]^{1/2}} \frac{c}{2d} \tag{8.8.10a}$$

Now the mathematical manipulations start:

$$1 - G_s^2R_1R_2 = \left[1 - G_s(R_1R_2)^{1/2}\right]\left[1 + G_s(R_1R_2)^{1/2}\right] \simeq 2\left[1 - G_s(R_1R_2)^{1/2}\right]$$

or

$$1 - G_s(R_1R_2)^{1/2} = \tfrac{1}{2}(1 - G_s^2R_1R_2)$$

where the saturated gain is set equal to the loss except when the difference appears. Substituting (8.8.8) for $(1 - G_s^2R_1R_2)$ leads to another version of the oscillation bandwidth.

$$\Delta\nu_{\text{osc.}} = \frac{h\nu}{P_{\text{out.}}}\left[\frac{c}{4\pi d}(1 - R_1R_2)\right] N_2^{(s)}c\sigma_{\text{SE}} \tag{8.8.10b}$$

Multiply the quantity in the brackets by ν/ν, convert $c/\nu = \lambda$, and recognize the residue as ν/Q or the passive cavity width $\Delta\nu_{1/2}$—see (6.3.5).

$$\Delta\nu_{\text{osc.}} = \frac{h\nu}{P_{\text{out.}}}\Delta\nu_{1/2}N_2^{(s)}c\sigma_{\text{SE}} \tag{8.8.10c}$$

Now use the one final bit of trickery: we set the saturated gain *coefficient* $(N_2^{(s)} - g_2/g_1N_1^{(s)})\sigma$, equal to the loss, which is prorated over the length d:

$$N_2^{(s)}\sigma_{\text{SE}}\left(1 - \frac{g_2N_l^s}{g_1N_2^s}\right) = \frac{1}{2d}\ln\left(\frac{1}{R_1R_2}\right) = \frac{1}{2d}\ln\left[\frac{1}{1 - (1 - R_1R_2)}\right]$$

$$\simeq \frac{1 - R_1R_2}{2d}$$

Therefore

$$N_2^{(s)}\sigma_{\text{SE}} \simeq \frac{1 - R_1R_2}{2d}\left(1 - \frac{g_2}{g_1}\frac{N_1^{(s)}}{N_2^{(s)}}\right)^{-1}$$

Use this expression in (8.8.10c), insert factors of $2\pi/2\pi$ and ν/ν, and again identify some of the factors as $\nu/Q = \Delta\nu_{1/2}$.

$$\Delta\nu_{\text{osc.}} = 2\pi\frac{h\nu}{P_{\text{out.}}}\left(\Delta\nu_{1/2}\right)^2\left(1 - \frac{g_2}{g_1}\frac{N_1^{(s)}}{N_2^{(s)}}\right)^{-1} \tag{8.8.10d}$$

All forms of (8.8.10) are equivalent, with some being more convenient than others for computation purposes. Equation (8.8.10d) is most widely quoted and was originally derived by Schawlow and Townes [12] and also by Gordon [13].

Let us close by emphasizing that this oscillation bandwidth is extremely small. For the numbers used as an example for this section and assuming an ideal system with $N_1 =$

0 (no lower state), we have $\nu = 2.42 \times 10^{14}$ Hz, $\lambda = 1.24\ \mu$m, $Q = 5.05 \times 10^7$, $\Delta\nu_{1/2} = 4.77$ MHz (for the passive cavity), $P_0 = 10$ mW (an arbitrary, but typical power). Thus $\Delta\nu_{\text{osc.}} = 2.29 \times 10^{-3}$ Hz! Obviously, we have ample room to allow for nonideal situations ($N_1 \neq 0$), and the limit to oscillation bandwidths is incredibly small. In practice, the wandering of the oscillation frequency caused by very slight perturbations in the mirror separation completely overwhelms the foregoing limit. But in any case, a delta function for the spectral representation of the laser is an excellent approximation.

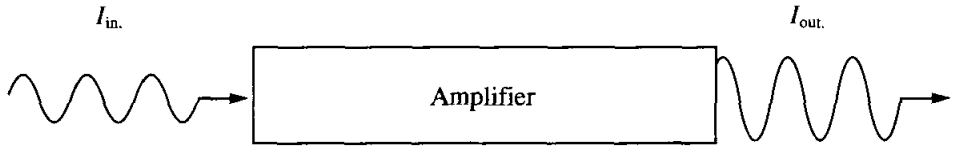
PROBLEMS

8.1. A homogeneously broadened laser transition at $\lambda = 10.6\ \mu$m (CO_2) has the following characteristics: $A_{21} = 0.34\ \text{sec}^{-1}$; $J_2 = 21$; $J_1 = 20$; $\Delta\nu_h = 1$ GHz.

(a) What is the stimulated emission cross section at line center?

(b) What must be the population inversion density $N_2 - (g_2/g_1)N_1$ to obtain a gain coefficient of 5%/cm. If the lifetime of the upper state is 10 μs and that of the lower state 0.1 μs, what is the saturation intensity?

8.2. An experiment involving a homogeneously broadened optical amplifier is depicted in the diagram below. For an input intensity of $1 W/\text{cm}^2$, the gain (output/input) is 10 dB. If the input intensity is doubled to $2 W/\text{cm}^2$, the gain is reduced to 9 dB.



(a) What is the small-signal gain (i.e., $I_{\text{in}} \to 0$) of this amplifier (in dB)?

(b) What is the saturation intensity?

(c) What is the maximum power (per unit area) that can be extracted from this amplifier (in limit of large input intensity)?

(d) What must be the input intensity to extract 50% of this maximum?

8.3. The purpose of this problem is to predict the saturated gain (or transmission) through an amplifier with a small-signal gain coefficient γ_0 that saturates according to the homogeneous law and a loss coefficient α that is not affected by the radiation. Hence the intensity changes with z according to

$$\frac{df}{dz} = \left(\frac{\gamma_0}{1+f} - \alpha\right) \cdot f \qquad 0 \leq z \leq l_g$$

where $f(z) = I(z)/I_s$ and l_g = length of the amplifier.

(a) Assume a small-signal gain $G_0 = \exp(\gamma_0 - \alpha)l_g = 4$ (i.e., 6 dB) and values of the ratio $\gamma_0/\alpha = 2, 5, 50$. Plot the saturated gain G_s (in dB) as a function of the input intensity (i.e., I_{in}/I_s). Use a semilog graph paper and plot G_s on the linear scale and I_{in}/I_s on the log scale covering the range from 10^{-2} to 10^3.

(b) If the input intensity is much larger than the saturation intensity, find an analytic expression for the output in terms of the input.

8.4. The model of Sec. 8.2 assumed an atomic system with equal degeneracies $g_1 = g_2$. Use the same logic path as used there to find an expression for the small-signal gain coefficient γ_0 and for the saturation intensity I_s for the case where $g_1 \neq g_2$.

ANSWER:

$$N_2 - (g_2/g_1)N_1 = \frac{R_2\tau_2\left[1 - (g_2/g_1)(\tau_1/\tau_{21})\right] - R_1\tau_1(g_2/g_1)}{1 + I_\nu/I_s}$$

$$I_s = \frac{h\nu}{\sigma\tau_2} \cdot \frac{1}{\left\{1 + (\tau_1/\tau_2)[1 - (g_2/g_1)(\tau_2/\tau_{21})]\right\}}$$

8.5. The ideas of gain saturation are equally applicable to absorption, and that is the purpose of this problem. Consider a single-frequency dye laser tuned to the center of the sodium D line at 5889.95 Å and irradiating a heated cell (630°K), 10 cm long, containing a mixture of sodium (Na) vapor at a density of 1.5×10^{15} cm^{-3} and helium (He) gas at a density of 6.53×10^{19} cm^{-3}. The self-broadening of this line, caused by collisions between sodium atoms, is 15 MHz for the conditions of this problem. The foreign gas broadening is due to collisions between sodium atoms and helium with a cross section estimated to be 10^{-14} cm^2. The following data for this transition are from NSRDS-NBS (vol. 11), U.S. Department of Commerce, National Bureau of Standards:

$N_1 \quad 3^2S_{1/2} \quad g_1 = 2 \quad E_1 = 0$

$N_2 \quad 3^2P_{3/2} \quad g_2 = 4 \quad E_2 = 16{,}978.07\ \text{cm}^{-1}; \quad A_{21} = 6.3 \times 10^7\ \text{sec}^{-1}$

(a) What are all of the pertinent line widths from the various causes ("natural," Doppler, self-broadening, or foreign gas [He] collisions)?

(b) If a "small-signal" laser is tuned to line center and propagates through the 10 cm length, what fraction emerges? Express the attenuation in dB.

(c) Let the input amplitude of the laser be a variable. Plot the transmission (in percent of the incident value) as a function of the input intensity normalized to a saturation value similar to that shown in Fig. 8.6. To find I_s, follow the procedure of Sec. 8.3, but remember that $N_1 + N_2 =$[Na] (i.e., the total number of sodium atoms is conserved).

8.6. **(a)** The response of an amplifier with a 6 dB small-signal gain to varying input intensities for the case of homogeneous broadening is covered in the text, with the results plotted as the solid curve in Fig. 8.8. Consider this amplifier, but assume that inhomogeneous broadening applies. Show that the equation relating the output to the input is given by

$$\ln\left[\frac{(1 + y_2)^{1/2} - 1(1 + y_1)^{1/2} + 1}{(1 + y_1)^{1/2} - 1(1 + y_2)^{1/2} + 1}\right] + 2\left((1 + y_2)^{1/2} - (1 + y_1)^{1/2}\right) = \gamma_0 l_g$$

where

$$y_2 = \frac{I_{\text{out.}}}{I_s} \qquad \text{and} \qquad y_1 = \frac{I_{\text{in.}}}{I_s}$$

This equation is plotted as the dashed curve in Fig. 8.8.

(b) Show that one recovers the small-signal amplification law for $(y_2, y_1) \ll 1$.

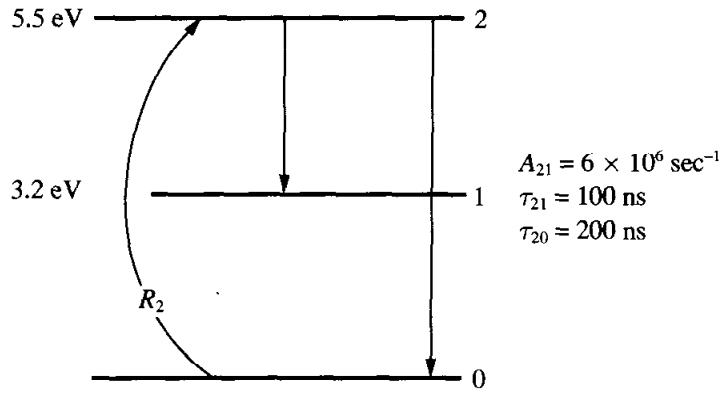
8.7. Consider the ideal laser medium shown below. The pump excites the atoms to state 2 at a rate R_2, which then decays to state 1 at a rate τ_{21}^{-1} and back to state 0 at a rate τ_{20}^{-1}. State 1 decays back to 0 so fast that the approximation $N_1 \approx 0$ is appropriate. The radiative rate for the $2 \rightarrow 1$ transition is 6×10^6 sec^{-1}, and its width is 10 GHz. (Assume a Lorentzian profile and steady state.)

(a) What is the stimulated emission cross section?

(b) What must be the pump rate R_2 in order to obtain a *small*-signal gain coefficient of 1%/cm?

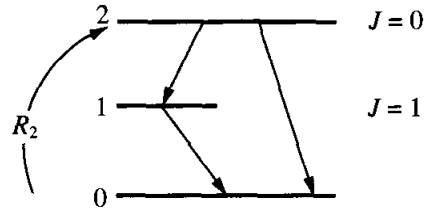
(c) What is the saturation intensity for the $2 \rightarrow 1$ transition?

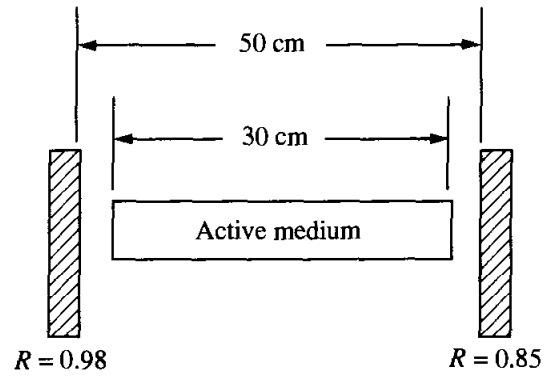
(d) How much power (in W/cm^3) is expended in creating the gain coefficient of (b)?



(e) Express the line width in Å units and cm^{-1} units.

8.8. In the laser system shown below, only state 2 is pumped directly from the ground (or 0) state with a pumping value of R_2 (cm^{-3}/sec). State 2 decays to 0 at a rate of 5×10^6 sec^{-1} and to 1 by spontaneous emission and quenching collisions at a rate of 10^7 sec^{-1}. The lifetime of state 1 is 50 ns. Assume homogeneous broadening; line





width $\Delta\nu = 2\,\text{cm}^{-1}$, $\lambda_0 = 4000$ Å; $A_{21} = 10^6\,\text{sec}^{-1}$; $M = 40 \times 1.67 \times 10^{-27}$ kg; $T = 500$ K; and that the medium fills the cavity in the manner shown below. Also assume steady state.

(a) What is the stimulated emission cross section at line center?

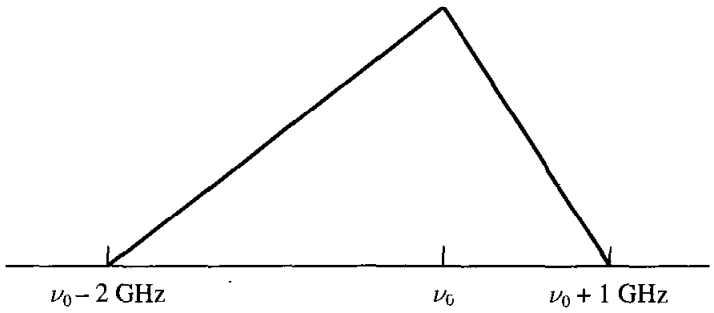
(b) What is the pump rate R_2 that brings the laser to threshold?

8.9. The following questions refer to an atomic system with $J_2 = 1$ and $J_1 = 2$.

(a) What is the ratio B_{12}/B_{21}?

(b) What is the formula for the small-signal gain coefficient for the $2 \rightarrow 1$ transition?

(c) If the line shape function could be approximated by the graph shown below, $A_{21} = 10^6$ sec, $\lambda = 6401$ Å, and $N_2 = N_1 = 10^{12}\,\text{cm}^{-3}$, what is the small-signal gain coefficient for the $2 \rightarrow 1$ transmission at $\nu = \nu_0$?



8.10. Consider a homogeneous broadened gain cell of Problem 8.2 being irradiated by an external laser of variable amplitude. The small-signal gain of the cell is 10 dB (at the line center), the full width at half maximum (FWHM) of the transition is 1 GHz, and the saturation intensity (at the line center) is 10 W/cm^2.

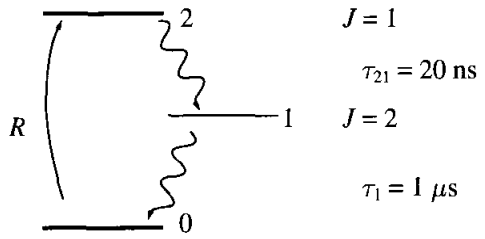
(a) Plot the gain (in dB) as a function of input intensity. (Assume that the frequency of the input coincides with the line center.)

(b) Repeat (a), but assume that the incoming frequency is detuned from the line center by 0.5 GHz. (Show as a dashed curve.)

(c) In the limit of large input intensities, the gain cell can, at most, add a photon flux to the signal. What is the maximum intensity that can be extracted from this gain cell?

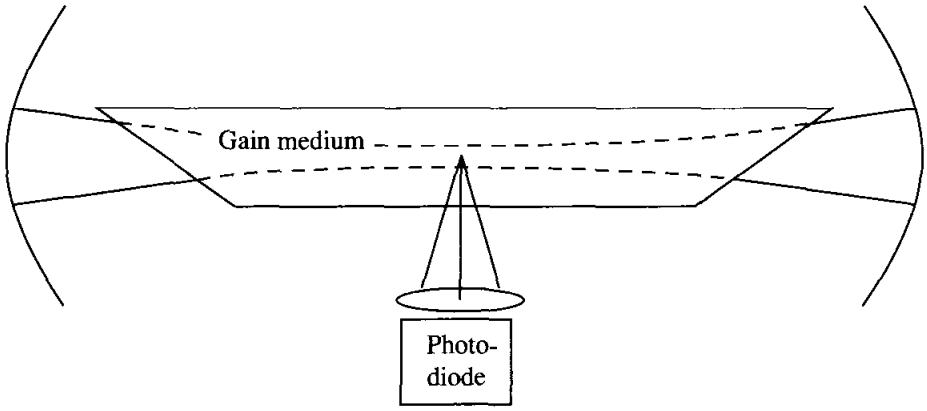
(d) What should be the input intensity to extract 95% of the maximum found in (c)?

8.11. Consider the following energy level diagram shown, which is representative of lasers such as the molecular nitrogen N_2.



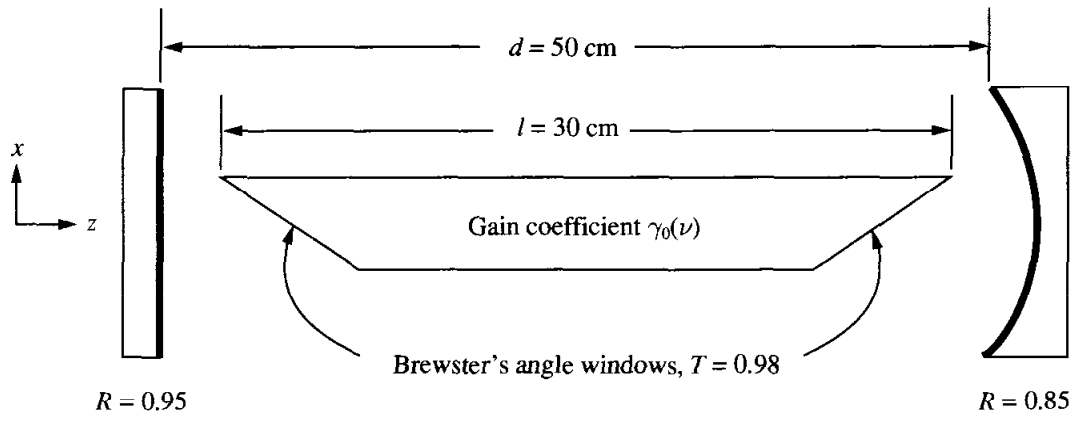
(a) If $R = 10^{20}$ cm^{-3}/sec, what are the equilibrium populations of states 2 and 1?
(b) Why is this system unsuitable for a CW laser?
(c) Suppose that the pump had the form of a step function of the amplitude given in (a). Sketch the time variation of the small-signal gain coefficient assuming all populations zero at $t = 0$ and a stimulated emission cross section of 10^{-12} cm^2.

8.12. The purpose of this problem is to point out a simple experimental method for estimating the saturation intensity, $I_{sat.}$, of a laser. You are given the experimental apparatus shown below, which is made up of a continuously pumped gain medium (small-signal gain coefficient γ_0), two nearly perfect reflecting mirrors, and a photodetector. The photodetector records the side fluorescence power emanating from a small volume of the gain medium. Assume that the laser transition is homogeneously broadened and that the lower laser level population is negligible compared to that in the upper state.



(a) If P_0 is the side fluorescence power (W/cm^{-3}) that is observed with one of the cavity mirrors blocked and P is measured when the laser is operating normally (i.e., mirror unblocked, everything else the same), then derive a simple expression that relates P/P_0 to the saturation intensity of the gain medium.
(b) If the side fluorescence is observed to be suppressed by 50% when the intercavity laser flux is 100 W/cm^2, what is $I_{sat.}$?

8.13. Consider the following laser cavity shown. The mirrors M_1, M_2 have a power reflectivity of 0.95 and 0.85, respectively, and the Brewster's angle windows transmit 98% of the power for the proper polarization (or minimum loss).



(a) If $\gamma_0 = 0$, then find the photon lifetime of the passive cavity. Assume a Brewster's angle polarization that yields a minimum loss.

(b) If $\gamma_0 = 4 \times 10^{-3}$ cm^{-1} (at line center), then will this system oscillate? Justify your answer.

(c) What is the orientation of the optical electric field for minimum loss in the cavity (i.e., does $\mathbf{E} = E_0\mathbf{a}_x$ or $E_0\mathbf{a}_y$ or $E_0\mathbf{a}_z$?)?

8.14. Suppose the distribution of center frequencies in an inhomogeneously broadened system is approximated by

$$p(f) = \frac{\Delta\nu_D}{2\pi\left[(f-\nu_0)^2 + (\Delta\nu_D/2)^2\right]}$$

where $\Delta\nu_D$ is the full width at half maximum of the "pure" inhomogeneously broadened line.

(a) Show that the saturated gain coefficient is given by

$$\gamma(\nu) = \frac{\gamma_0(\nu_0)}{\left(1+\dfrac{I_\nu}{I_s}\right)^{1/2}} \cdot \frac{1 + \left(\dfrac{\Delta\nu_h}{\Delta\nu_D}\right)\left(1+\dfrac{I_\nu}{I_s}\right)^{1/2}}{\left\{\left(\dfrac{\nu-\nu_0}{\Delta\nu_D/2}\right)^2 + \left[1+\left(\dfrac{\Delta\nu_h}{\Delta\nu_D}\right)\left(1+\dfrac{I_\nu}{I_s}\right)^{1/2}\right]^2\right\}}$$

In this expression, $\Delta\nu_h$ is the homogeneous line width, $\gamma_0(\nu_0)$ is the small-signal gain of the "pure" inhomogeneous transition evaluated at line center, I_s is the frequency-independent saturation intensity, and I_ν is the intensity of the wave stimulating the atoms.

(b) Show that one recovers the homogeneous saturation law if $\Delta\nu_h/\Delta\nu_D$ and/or $I_\nu/I_s \gg 1$.

(c) Identify the circumstances under which one recovers the "pure" inhomogeneous saturation behavior.

(d) Plot the saturated gain coefficient as a function of I/I_s for $\nu = \nu_0$ and $\Delta\nu_h/\Delta\nu_D = 1/2$. Show also the graph of the two limiting forms of the saturation: the extreme homogeneous limit and the inhomogeneous limit.

The problem involves considerable arithmetic so the following is given to alleviate some of the pain.

Define $f - \nu_0 = x$; $\delta = \nu_0 - \nu$; and therefore $(\nu - f)^2 = (x+\delta)^2$;

$$a = \Delta\nu_D/2; \qquad b = \Delta\nu_H/2 \text{ with } \Delta\nu_H = \Delta\nu_h\left(1+\frac{I_\nu}{I_s}\right)^{1/2}.$$

A partial fraction expansion will be needed:

$$\frac{1}{(x^2+a^2)\left[(x+\delta)^2+b^2\right]} = \frac{Ax+B}{x^2+b^2} + \frac{C(x+\delta)+D}{(x+\delta)^2+b^2}$$

where

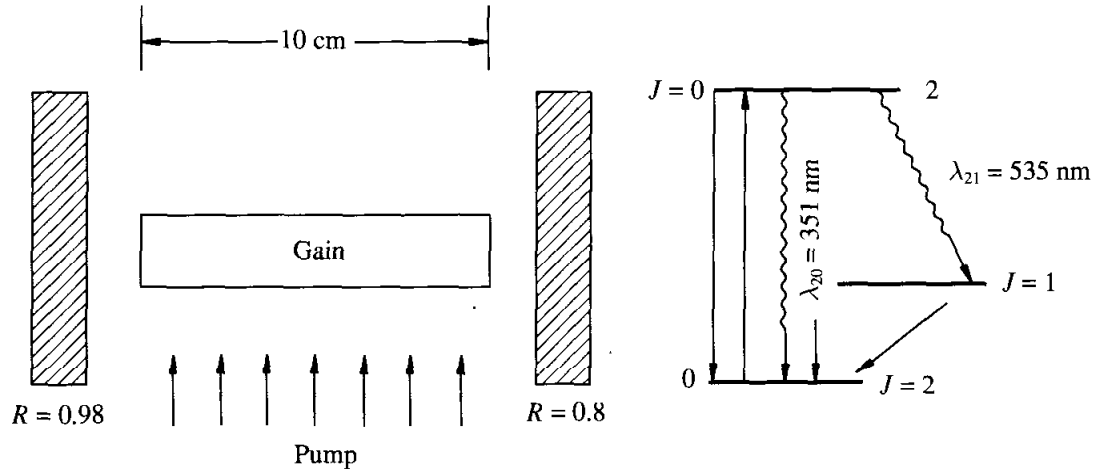
$$B = \frac{\delta^2 + b^2 - a^2}{\delta^4 + 2(b^2+a^2)\delta^2 + (b^2-a^2)^2}$$

$$D = \frac{\delta^2 - b^2 + a^2}{\delta^4 + 2(b^2 + a^2)\delta^2 + (b^2 - a^2)^2}$$

The terms involving A and C vanish after integration. Finally, note that

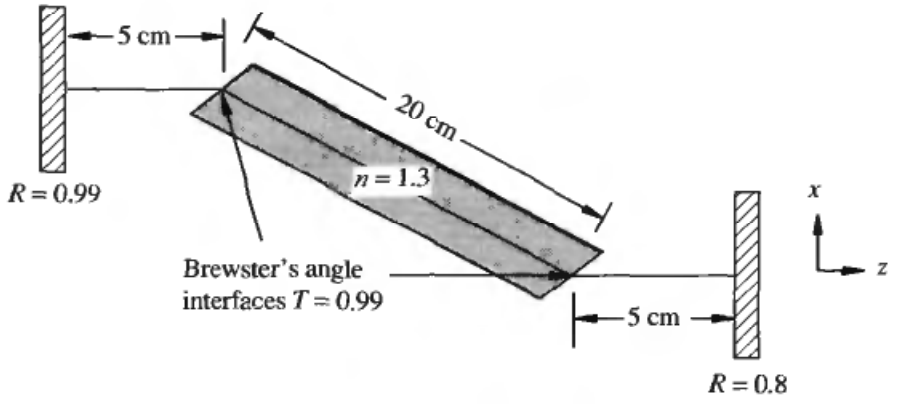
$$\delta^4 + 2(b^2 + a^2)\delta^2 + (b^2 - a^2)^2 = [\delta + (b - a)^2][\delta^2 + (b + a)^2]$$

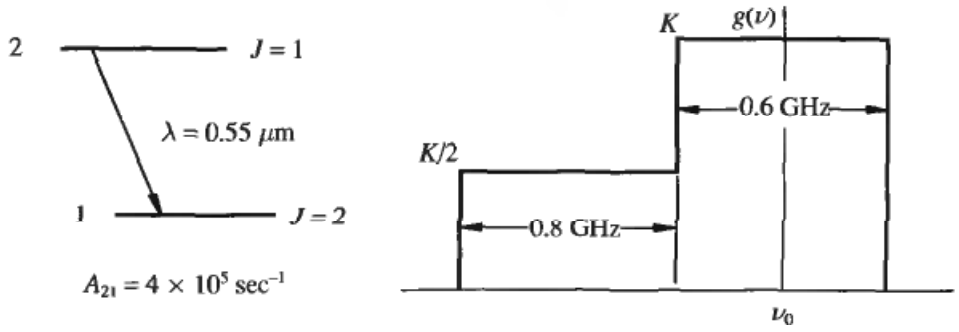
8.15. A model for one laser being optically pumped by another is shown in the diagram below. Compute the pump intensity (in W/cm^2) for the system to reach threshold at $\lambda_{21} = 535$ nm. The following information may be useful: $A_{20} = 5 \times 10^7$ sec^{-1}; $A_{21} = 1 \times 10^8$ sec^{-1}; $\tau_1 = 20$ ns; $M = 138$ amu; $M_p = 1.67252 \times 10^{-27}$ kg; $T = 300$ K; and $N_0 + N_1 + N_2 = 10^{14}$ cm^{-3}. The stimulated emission cross section at 351 nm is 10^{14} cm^2, and the $2 \rightarrow 1$ transition is homogeneously broadened with $\Delta\nu = 1$ GHz. Since this is a threshold calculation, we can neglect stimulated emission on the $2 \rightarrow 1$ line. You may neglect the attenuation of the pump as it propagates across the laser cavity. Assume steady state.



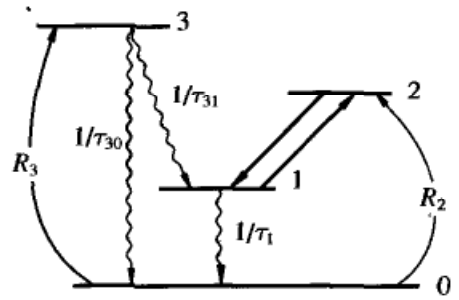
(a) What is the stimulated emission coefficient on the $2 \rightarrow 1$ transition?
(b) What gain coefficient on the $2 \rightarrow 1$ transition is required to reach threshold?
(c) What pump intensity (in W/cm^2) is required to reach threshold for oscillation at λ_{21}?

8.16. The laser shown below utilizes a transition that peaks at $\lambda = 0.55$ nm. The medium has an index of refraction of 1.3 with an inhomogeneous line shape that can be *approximated* by the graph shown.
(a) What is the value of the line shape function at $\nu = \nu_0$?
(b) What is the stimulated emission cross section at line center?
(c) What is the photon lifetime of the *passive* cavity?
(d) What is the orientation of the optical electric field?
(e) If the population of state 1 were 10^{12} cm^{-3}, what must be the density in state 2 to reach threshold?





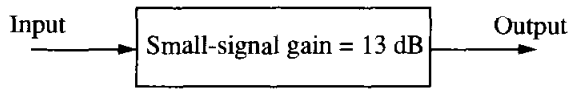
8.17. The following energy level diagram shown illustrates a situation that can occur in gas lasers. The upper state of the $3 \rightarrow 1$ transition is pumped at a rate R_3, and if the lifetime ratio τ_3/τ_1 is favorable, then we can obtain gain on the $3 \rightarrow 1$ transition. However, state 2 can also be excited (at a rate R_2, and if sufficient feedback is provided, oscillation can take place on the $2 \rightarrow 1$ transition, which raises the population of state 1 and lowers the gain at λ_{31}. Analyze this case by answering the following two questions. Assume steady state, assume all degeneracies equal to 1, and neglect spontaneous decay from $2 \rightarrow 1$. Define $\tau_3^{-1} = \tau_{30}^{-1} + \tau_{31}^{-1}$.



(a) Assume $R_2 = 0$ and thus no oscillation on the $2 \rightarrow 1$ transition. Find an expression of the small-signal gain coefficient on $3 \rightarrow 1$ in terms of the lifetimes, the pumping rates, and the stimulated emission cross section.

(b) Assume that $R_2 \neq 0$ and that there is a strong saturating signal at the $2 \rightarrow 1$ wavelength such that $N_1 = N_2$. (Neglect spontaneous decay from $2 \rightarrow 1$.) Find a new expression for the small-signal gain coefficient at λ_{31}.

8.18. A homogeneously broadened optical amplifier with a small-signal gain of 13 dB is irradiated with a wave with an intensity of $5\,\mathrm{W/cm^2}$. The output intensity is $30\,\mathrm{W/cm^2}$. What is the saturation intensity? If the saturation intensity were $20\,\mathrm{W/cm^2}$, what is the maximum power (per unit of area) extractable from this amplifier?



8.19. The saturated gain of a laser oscillator is approximated by

$$G_s^2 = \frac{1}{R_1 R_2}\left[1 - N_2^s c\sigma_{\mathrm{SE}} h\nu \frac{(1 - R_1 R_2)}{P_{\mathrm{out.}}}\right]$$

where N_2^s is the saturated value of the upper state density. This leads to the spectral distribution of power according to

$$S(\omega) = E \cdot E^* = \frac{E_0^2}{\left(1 - G_s\sqrt{R_1 R_2}\right)^2 + 4G_s\sqrt{R_1 R_2}\sin^2(\omega d/c)}$$

Evaluate the line width $\Delta\nu$ of the laser oscillating in the vicinity of $\omega d/c = q\pi$ for the following conditions: $\nu = 5 \times 10^{14}$ Hz, $R_1 R_2 = 0.9$; $d = 50$ cm; $\sigma = 10^{-12}\,\mathrm{cm^2}$, $N_2^s = 10^{12}\,\mathrm{cm^{-3}}$; and $P_0 = 1$ mW.

8.20. Consider an optical amplifier/attenuator that had a Lorentzian line shape with $\Delta\nu_h = 2$ GHz and a "gain" at line center that can be between -30 dB and $+30$ dB. Plot the FWHM of the spontaneous radiation emerging from the end of the cell [normalized to the line width of the transition (i.e., $\Delta\nu_{\mathrm{emiss}}/\Delta\nu_h$)].

8.21. The following questions refer to the laser system depicted in the following diagram on page 251.

(a) Find the photon lifetime of the passive cavity; that is, with $\gamma_0(\nu) = 0$.

(b) What is the passive cavity Q?

(c) What is the free spectral range of the cavity?

(d) What is the minimum gain coefficient $\gamma_0(\nu_0)$ necessary to sustain oscillation in this cavity?

(e) If the gain coefficient $\gamma_0(\nu_0)$ were $2 \times 10^{-2}\,\mathrm{cm^{-1}}$ (at line center) and the line shape of the transition was approximated by a Lorentzian with $\Delta\nu_h = 1.5$ GHz, then how many $\mathrm{TEM}_{0,0,q}$ modes are above threshold?

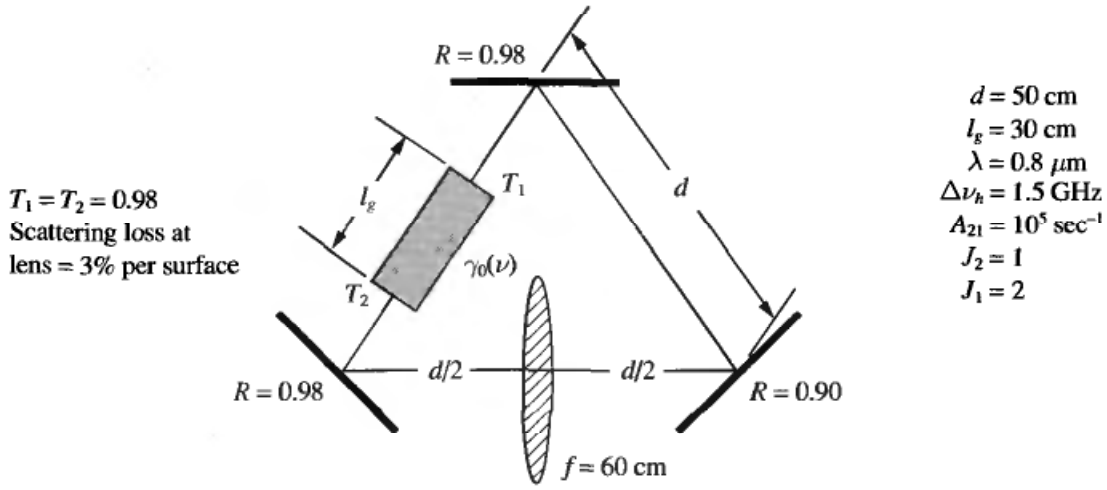
(f) What is the stimulated emission cross section (at line center)?

(g) What is the absorption cross section?

(h) The characteristic beam parameter z_0 is related to the dimensions of the cavity and the focal length of the lens by

$$z_0 = \frac{\pi w_0^2}{\lambda} = (3df)^{1/2}\left(1 - \frac{3d}{4f}\right)^{1/2}$$

where $d = 50$ cm and $f = 60$ cm.

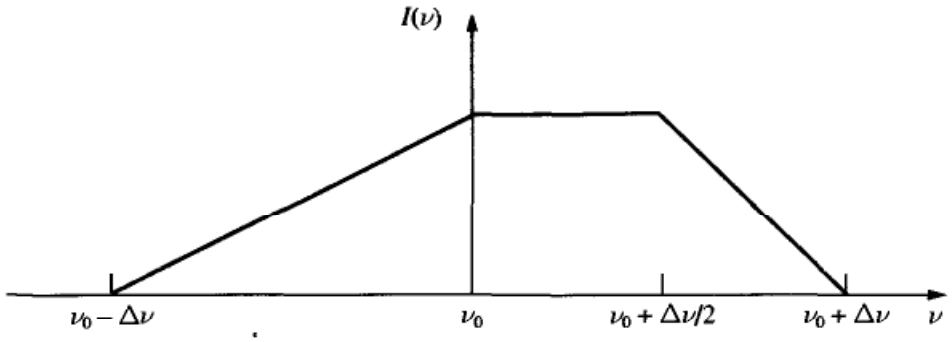


(1) Where is $z = 0$ in the cavity?

(2) Find a formula for the resonant frequency of the $\text{TEM}_{m,p,q}$ mode.

(3) What is the *difference* in resonant frequencies (in MHz) of the $\text{TEM}_{0,0,q}$ and $\text{TEM}_{1,0,q}$ modes?

8.22. The fluorescence from a certain system can be approximated by the sketch shown below. What is the value of $g(\nu_0)$?



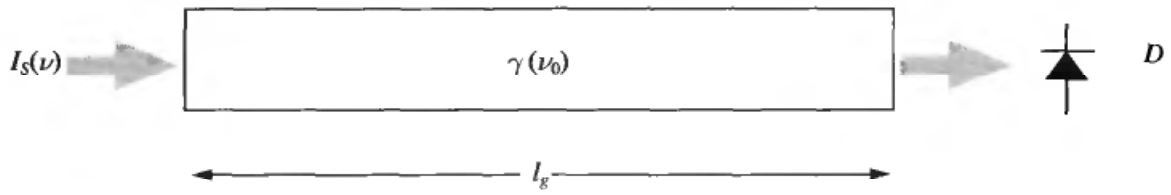
8.23. To analyze the laser system shown in Problem 8.8, assume the following: equal pump rates to states 2 and 1, zero lifetime for state 1, lifetime of state 2 is 100 ns, stimulated emission cross section is 1.3×10^{-17} cm^2 (at line center), no depletion of state 0, and CW operation.

(a) How much pump power in W/cm^3 is required to establish a gain coefficient of 0.05 cm^{-1}?

(b) What is the maximum power (per unit of volume) that can be extracted by stimulated emission?

8.24. The problem is to model the gain measurement indicated in the experiment sketched on the diagram below. The desired quantity is line-integrated gain at line center $\gamma(\nu_0)l_g$ in terms of a measured ratio between the detector (D) outputs with and without the gain present. If the source were a δ function in ν, the interpretation would be trivial, but no source meets that specification. Assume that the output of the detector

is proportional to the total number of photons in the transmitted signal integrated over the spectral bandpass of the source.



Construct a family of curves showing the "gain" = ratio D (with gain) to D (without) as a function of $\gamma(\nu_0)l_g$ for three values of $\alpha = \Delta\nu_s/\Delta\nu_g$, the ratio of the source to the gain linewidths. Plot this apparent gain for $0 < \gamma(\nu_0)l_g < 2$ for three cases:

(a) $\alpha = 0.1$ (the source is almost a delta function)
(b) $\alpha = 1.0$ (the source and gain widths are equal)
(c) $\alpha = 5.0$ (the source is almost a blackbody)

Assume that

$$I_s = I_0 \exp\left\{-4 \ln 2 \left[(\nu_0 - \nu)/\Delta\nu_s\right]^2\right\}$$

and

$$\gamma(\nu) = \gamma(\nu_0) \exp\left\{-4 \ln 2 \left[(\nu_0 - \nu)/\Delta\nu_g\right]^2\right\}.$$

(NOTE: This is a variation of the *line absorption technique* presented in [22]. The answer can be found on page 122 of the 1971 printing for an absorbing transition.)

This is an important technique for experimentalist. The line width of most sources—even a laser—is not zero and that fact makes the experiment somewhat difficult to interpret. For instance, a reasonably good dye laser would have a line width of 0.1cm^{-1} or 3 GHz, but that is larger than the line width of many gas transitions. (The neon transition at 6328 Å is only 1.5 GHz wide.) This problem is intended to teach you the importance, limitations, and pitfalls of a seemingly simple measurement.

8.25. Consider a laser system similar to that of Fig. 8.4 but only with pumping to the upper state 2 at a rate of R_2. If the lifetimes and branching ratio are proper, then this pumping creates a population inversion $(N_2^0 - N_1^0)$ where the superscript 0 indicates the value of the populations when there is no stimulating wave tuned to the $2 \rightarrow 1$ transition. When a stimulating wave is present, N_2 is less than N_2^0 and N_1 is greater than N_1^0, as is shown on the sketch below.

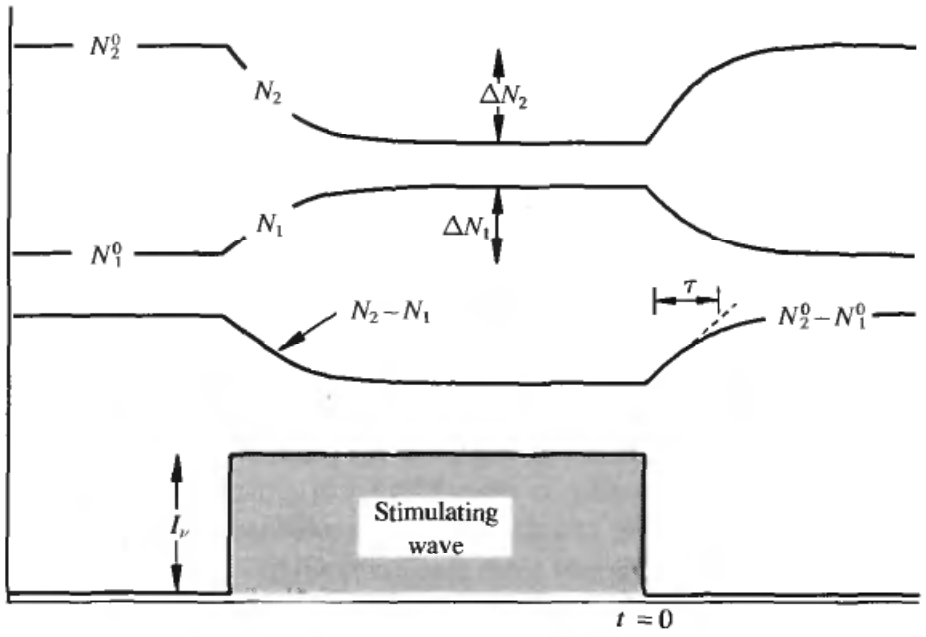
(a) Find an expression for the densities $N_{(1,2)}^0$ that involves R_2 and the lifetimes.
(b) Find an expression for $\Delta N_{(1,2)}$ that involves R_2 and the lifetimes. What is the ratio of $\Delta N_1/\Delta N_2$?

(c) Show that the rate of change of the inversion at $t = 0$ can be expressed as

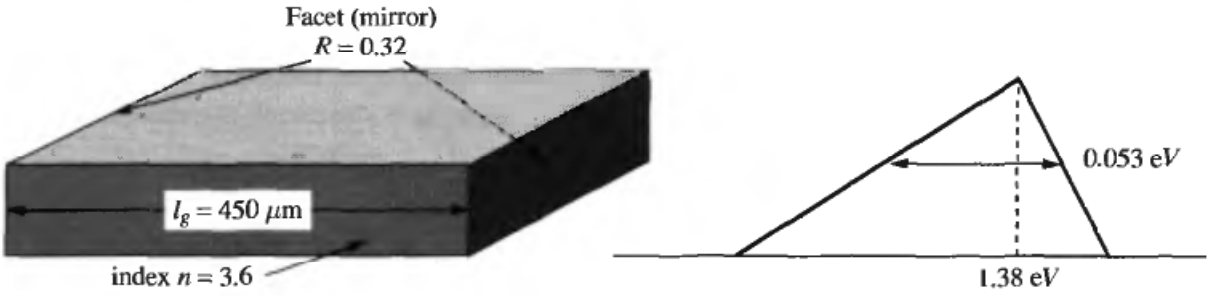
$$\left.\frac{d(N_2 - N_1)}{dt}\right|_{(t=0)} = \frac{\left\{[N_2^0 - N_1^0] - [N_2(0) - N_1(0)]\right\}}{\tau}$$

What is the value of τ in terms of the lifetimes and branching ratio.

(d) Find an expression for the time dependence of the populations in terms of the steady state densities $N^0_{(1,2)}$, the saturated changes $\Delta N_{1,2}$ and the lifetimes.



8.26. The spontaneous emission from a certain (semiconductor) laser can be approximated by the sketch shown below. The spontaneous lifetime is 5 ns.



(a) What is the stimulated emission cross section?

(b) If the facet (mirror) reflectivities (for power) were 0.32 and the length of the medium completely filling the cavity were 450 μm, then what must be the inversion density to obtain oscillation?

(c) What is the photon lifetime τ_p?

8.27. Consider the atomic system (sodium) shown below in which an external laser, tuned to the center of a transition terminating on the ground state, has a variable amplitude. If the input intensity were zero, *all* of the atoms would be in the ground state.

(a) If the input intensity were infinitely strong, then what fraction of the atoms would be in state 2?

(b) What should be the input intensity such that 25% of the atoms are in the upper state?

$E_2 = 16{,}978.07$ ——————— $g_2 = 4$ $\quad A_{21} = 6.3 \times 10^7 \text{ sec}^{-1}$

homogeneous broadened:

$\Delta\nu = 3.7$ GHz

$E_1 = 0$ ——————— $g_1 = 2$

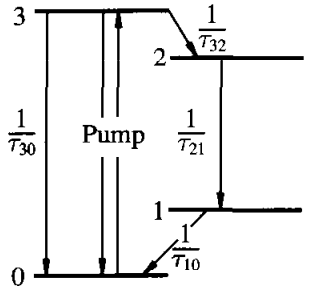
8.28. An external CW laser, tuned to the center of the $0 \rightarrow 2$ transition of Fig. 8.4, pumps the atomic system with the goal of creating a population inversion on the $2 \rightarrow 1$ wavelength.

(a) Formulate the rate equations neglecting stimulated emission on the $2 \rightarrow 1$ transition.

(b) Assume the conditions of (a), steady state, and solve for the maximum inversion possible (i.e., when the external laser is infinitely strong).

(c) Solve for the small signal inversion in terms of the pump intensity, the stimulated emission cross section for the pump, the indicated lifetimes, and the total number of atoms N. Do not forget to conserve atoms (i.e., $N_0 + N_1 + N_2 = N$) and thus only two simultaneous equations are necessary. (This makes the problem a bit different from that in the text.)

8.29. Assume steady state, equal degeneracies of all states, an *infinite pump* laser tuned to the $3 \rightarrow 0$ transition, and compute the *inversion* created on the $2 \rightarrow 1$ transition in terms of the stimulated emission cross section for the $3 \rightarrow 0$, the decay rates indicated, and the initial density of active atoms $N_0 + N_1 + N_2 + N_3 = N$.

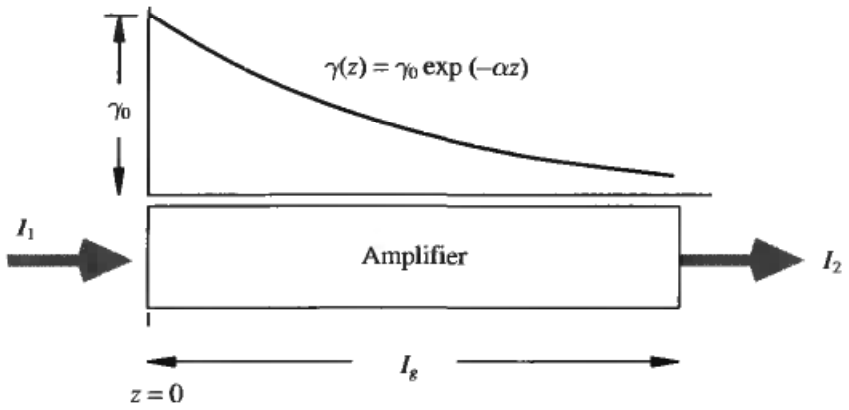


8.30. An optical amplifier is pumped by another laser, and as a consequence of the attenuation of the pump beam (α), the gain coefficient at the signal frequency is a function of z according to

$$\frac{1}{I_\nu}\frac{dI_\nu}{dz} = \frac{\gamma_0 e^{-\alpha z}}{1 + (I_\nu / I_s)}$$

(a) Derive an expression that *relates* the output I_2 to the input I_1 after the signal has propagated a length l_g.

(b) What is an expression for the maximum power (per unit of area) that can be extracted from this amplifier?



8.31. A fiber glass amplifier for 1.55 μm radiation has a small signal gain of 23 dB (i.e., 200X), a stimulated emission cross section of 10^{-19} cm^2, an upper state lifetime of 130 μs, and a very short lower state lifetime. Assume steady state, equal degeneracies, and that the lifetime of state 1 is very short.

(a) If the length of this fiber amplifier were 100 cm, what must be the inversion $N_2 - N_1$ to obtain this small signal gain?

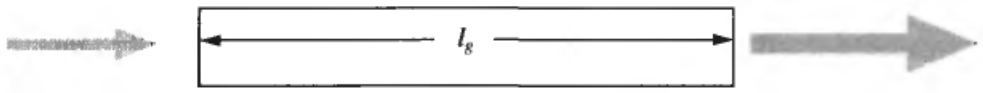
(b) How much reflection at the input and output of this amplifier can be tolerated before it breaks into oscillation?

(c) At what value of the input intensity will the gain of this amplifier be 20 dB (i.e., 100X)?

8.32. In some (if not most) optical amplifiers there are some residual losses (say, owing to scattering by imperfections), and thus the intensity changes with z according to

$$\frac{1}{I}\frac{dI}{dz} = \frac{\gamma_0}{1 + I/I_s} - \alpha$$

where $I_s = 16\text{watts/cm}^2$. Let the value of $\gamma_0 l_g = 2$ and the transmission through the amplifier in the absence of excitation (i.e., $\gamma_0 = 0$) be $0.85(l_g = 5\text{ m})$



(a) Compute the net small signal gain ($I_{\text{out.}}/I_{\text{input}}$) for this amplifier (in dB).

(b) What is the loss coefficient α (in cm^{-1})?

(c) If the input intensity were too large, then the amplifier becomes an attenuator. What maximum input intensity can be used and have unity gain?

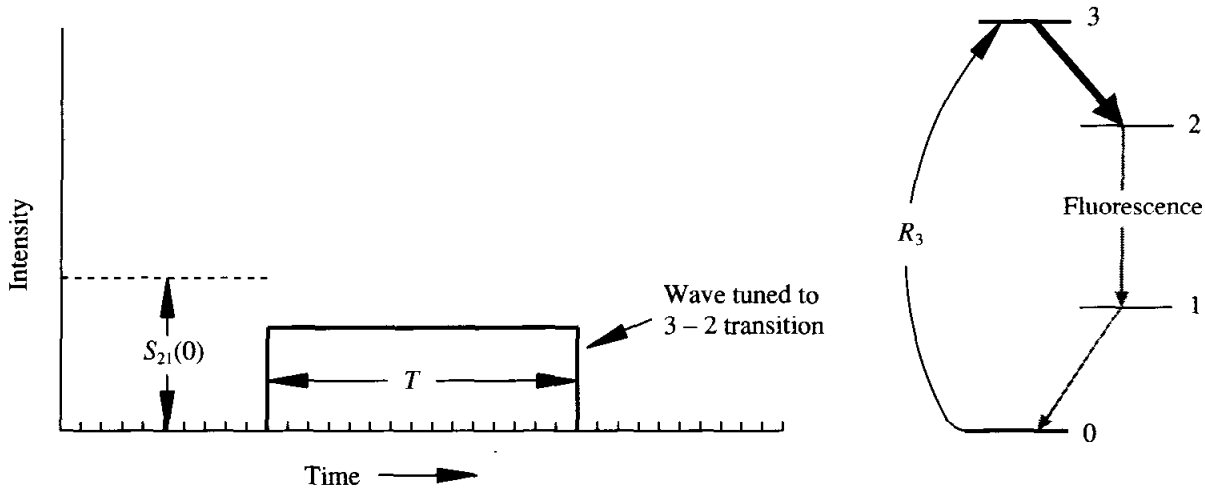
8.33. A square pulse of radiation tuned to the center of the $3 \rightarrow 2$ transition and equal to three times the saturation intensity illuminates a group of atoms whose energy level

diagram is shown below. Only state 3 is pumped by an external mechanism, but most of its spontaneous decay is to 1 and 0 with only 1% going to state 2. You may assume $\tau_2 \ll \tau_3 \ll T$, $N_2 \ll N_3$ and thus can neglect reabsorption from 2 to 3. An observer measures the time history of the side fluorescence $S_{21}(t)$ from state 2 with a photon detector and an oscilloscope.

(a) What are the rate equations for state 3 and 2?

(b) What is the ratio of the *steady state* fluorescence intensity on the $2 \rightarrow 1$ transition with and without the optical pulse tuned to $3 \rightarrow 2$ (i.e., $S_{21}(I \neq 0)/S_{21}(0)$).

(c) Sketch the time history of the fluorescence signal on the graph shown below indicating *relative* amplitudes and any characteristic time constants. (It is not necessary to solve the rate equation but merely interpret their form.)



8.34. A four-level system is optically pumped on the $0 \rightarrow 3$ transition with state 3 decaying to 2 at a rate of $(1\ \text{ns})^{-1}$. State 2 has a lifetime of 250 μs and decays to both 0 and 1 with a branching ratio of 0.4 and 0.6, respectively. State 1 decays back to 0 with a lifetime of 25 ns.

(a) Formulate the rate equations that describe the dynamics of the populations. Neglect stimulated emission on the $2 \rightarrow 1$ route.

(b) Compute the amount of absorbed power (per unit of volume) to maintain the steady state population in state 2 at $10^{19}\ \text{cm}^{-3}$.

8.35. Consider a simple laser system shown in Fig. 8.4 in which state 2 is selectively pumped from 0 at a rate $R_2(L^{-3}T^{-1})(R_1 = 0)$. The levels are close enough together such that there is thermal population in each of the states with Boltzmann factors being $N_2^e/N_0^e = 1/8$ and $N_1^e/N_0^e = 1/4$. Assume steady state and equal degeneracies for all levels and neglect stimulated emission on all transitions. The total density of the active atoms is $2.2 \times 10^{20}\ \text{cm}^{-3}$.

Please indicate the proper units along with the numerical answer for each of the following:

(a) What is an expression for the densities in each of the states?

(b) At thermal equilibrium, the $2 \rightarrow 1$ transition is absorptive. What is an expression for the absorption coefficient?

(c) Any *excess* populations in (2,1) created by the pump, above and beyond the thermal equilibrium values, decays to (1,0) with rates $(1/\tau_{2,1})$, respectively.

(1) Formulate the rate equations for the excess populations $(N_2 - N_2^e)$ and $N_1 - N_1^e)$. (The superscript e indicates thermal equilibrium values.)

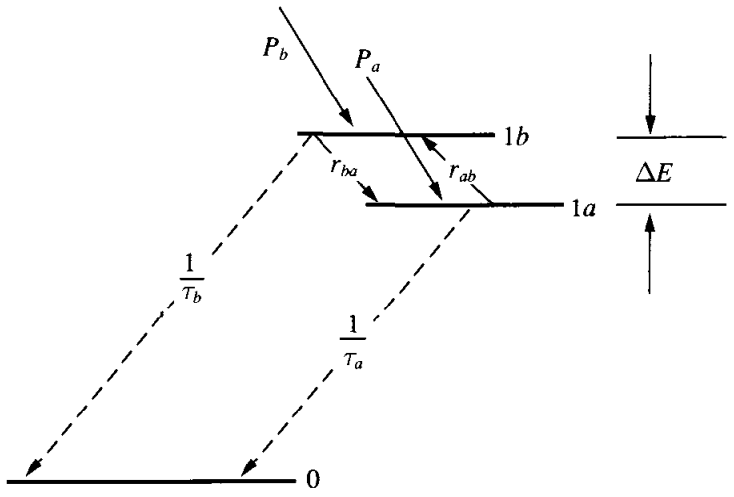
(2) Derive a formula for the pumping rate required to obtain optical transparency (i.e., $\alpha = 0$) on the $2 \rightarrow 1$ transition.

(3) Should the ratio, τ_2/τ_1, be greater or less than 1 in order to reach optical transparency? Sketch the functional dependence of the absorption coefficient on the pumping rate for the two cases where the lifetime ratio is favorable or unfavorable.

8.36. Consider the system shown below depicting two states ($1a$, $1b$) which are closely coupled by very fast thermal processes shown on the diagram and which are related by

$$r_{ab} = r_{ba} \exp[-\Delta E/kT]$$

where $r_{ba} = 10^{12}$ sec^{-1}. This coupling rate is much faster than the natural decay rate of each state: $\tau_a^{-1} = 10^4$ s^{-1} and $\tau_b^{-1} = 10^3$ s^{-1} and thus the sum of the populations decays with a time constant which is different from either τ_a or τ_b. Assume that τ_a, τ_b are independent of temperature.



(a) What is the decay rate of the manifold $1a + 1b$ if the Boltzmann factor $\exp[-\Delta E/kT] = 1/9$ at room temperature (300K)?

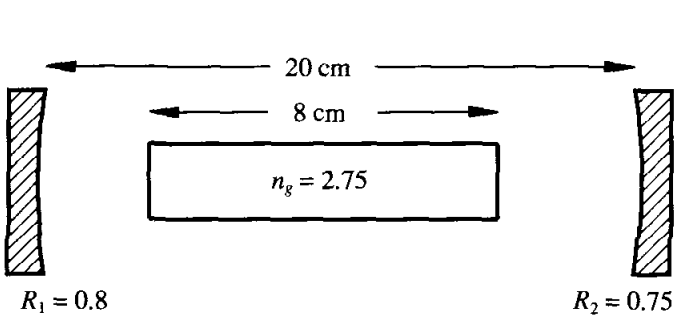
(b) If the temperature were increased to 600 K, then what is the decay rate of the populations in the manifold?

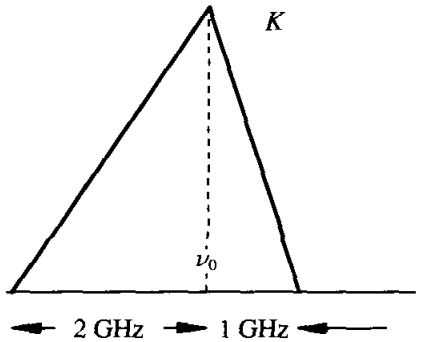
(c) Derive a general expression for the decay rate of the manifold for arbitrary temperatures and values for τ_a and τ_b.

8.37. The *inhomogeneous* lineshape function for the laser shown below can be approximated by the triangle shown below. Use the given information to compute

(a) Stimulated emission cross section.

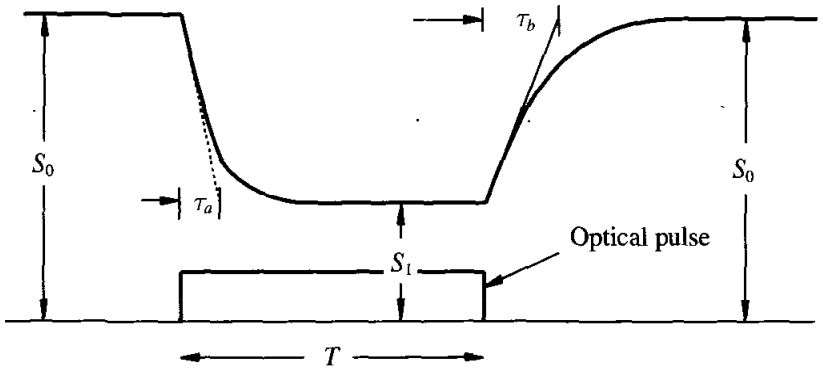
(b) The inversion to reach threshold for oscillation.
(c) The number of longitudinal modes capable of oscillation if the system were pumped to 1.25 times threshold.

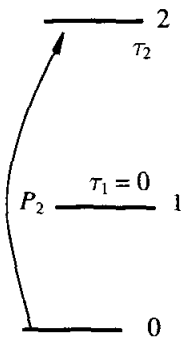




$A_{21} = 2.5 \times 10^{+6}\ s^{-1}$; $g_2 = 3$; $g_1 = 1$; $\bar{\nu}_0 = 12{,}500\ cm^{-1}$; $n_g = 2.75$

8.38. Consider a two-level system shown below where some sort of a pumping mechanism maintains a population in state 2. State 1 decays so fast that $N_1 \approx 0$ under any and all circumstances. An observer measures the *side* fluorescence with a detector and an oscilloscope and records the following wave shape when the cell is irradiated by an optical wave tuned to line center with a "square" envelope so that $I(t) = I_0[u(t) - u(t - T)]$.



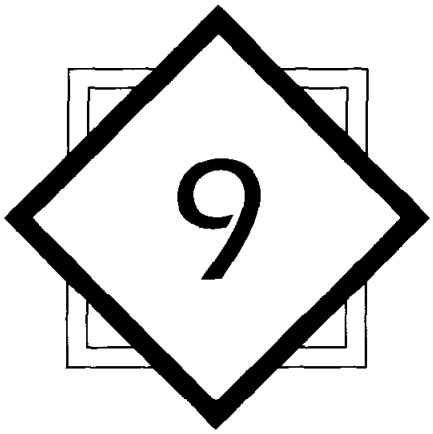


(a) What is an *expression* for the ratio of sidelight fluorescent amplitude S_0/S_1?
(b) The side fluorescence is suppressed during the optical pulse, with the two limits being connected by an exponential. What is an expression for the time constant τ_a?
(c) Once the pulse is removed for $t > T$, the side fluorescence recovers to S_0 exponentially. What is the time constant τ_b?

REFERENCES AND SUGGESTED READINGS

1. A. E. Siegman, *Introduction to Lasers and Masers* (New York: McGraw–Hill, 1971).
2. A. Yariv, *Introduction to Optical Electronics*, 2nd ed. (New York: Holt, Rinehart and Winston, 1971).

3. A. Maitland and M. H. Dunn, *Laser Physics* (Amsterdam: North–Holland, 1969).
4. A. Yariv, *Quantum Electronics*, 2nd ed. (New York: John Wiley & Sons, 1975).
5. W. V. Smith and P. P. Sorokin, *The Laser* (New York: McGraw–Hill, 1966).
6. W. R. Bennett, Jr., "Gaseous Optical Masers," *Appl. Opt., Suppl. Opt. Masers*, 24–61, 1962.
7. W. R. Bennett, Jr., "Inversion Mechanisms in Gas Lasers," *Appl. Opt., Suppl. 2 Chem. Lasers*, 3–33, 1965.
8. W. E. Lamb, Jr., "Theory of an Optical Maser," *Phys. Rev. 134A*, 1429, 1964.
9. W. R. Bennett, Jr., "Hole-Burning Effects in a He-Ne Optical Laser," *Phys. Rev. 126*, 580, 1962.
10. A. L. Bloom, "Gas Lasers," *Proc. IEEE 54*, 1262, 1966.
11. W. S. C. Chang, *Principles of Quantum Electronics* (Reading, Mass.: Addison–Wesley, 1969). This book has an extensive bibliography.
12. A. L. Schawlow and C. H. Townes, "Infrared and Optical Masers," *Phys. Rev. 112*, 1940, Dec. 1968.
13. E. I. Gordon, "Optical Maser Oscillators and Noise," *Bell Syst. Tech. J. 43*, 507–539, 1964.
14. A. Yariv, *Quantum Electronics* (New York: John Wiley & Sons, 1975).
15. K. A. Jones, *Introduction to Optical Electronics* (New York: Harper & Row, 1987).
16. W. Demtroder, *Laser Spectroscopy*, 2nd Printing (New York: Springer–Verlag, 1982).
17. R. L. Barger and R. L. Hall, "Pressure Shift and Broadening of Methane Line at 3.39 μm Studied by Laser Saturated Molecular Absorption," *Phys. Rev. Lett. 22*, 4, 1969.
18. C. J. Borde, J. L. Hall, C. V. Kunasz, and D. G. Hummer, "Saturated Absorption Line Shape: Calculation of the Transit-Time Broadening by a Perturbation Approach," *Phys. Rev. A 14*, 236–262, 1976.
19. J. B. Anderson, J. Maya, M. W. Grossman, R. Laqueskenko, and J. F. Waymouth, "Monte Carlo Treatment of Resonant Radiation Imprisonment in Fluorescent Lamps," *Phys. Rev. A, 31*, 2968, 1985.
20. T. Paoli, "Saturation Behavior of the Spontaneous Emission from Double-Heterostructure Junction Lasers Operating High Above Threshold," *IEEE J. Quant. Electron. QE-9*, 267, 1973.
21. V. P. Chebotayev, "Super High Resolution Spectroscopy," in *The Laser Handbook*, Vol. 5, Article 3, Eds. M. Bass and M. L. Stitch (New York: North Holland, 1985). NOTE: The frequency of helium-neon laser at 3.39 μm has been measured to be ν(3.39 μm line) = 88.3761816029 THz $\pm$ 1.2 kHz.
22. A. C. G. Mitchell and M. W. Zemansky, *Resonance Radiation and the Excited State* (New York: Cambridge University Press, 1934). Reprinted 1971.



General Characteristics of Lasers

9.0 INTRODUCTION

Chapter 8 addresses general issues of the approach to oscillation and the steady state saturation of the gain but did not discuss the actual amplitude of CW oscillation and did not address possible transient effects where the optical power may be a function of time. These are topics to be addressed here. We will also take a first cut at the pumping process so as to illustrate and calculate the rates $R_{1,2}$ of the previous chapter in terms of an easily identified process. This will enable us to make a general evaluation as to the ease and limiting efficiency of a laser.

9.1 LIMITING EFFICIENCY

9.1.1 Factors in the efficiency

To create a laser, a population inversion must first be established, and this requires power from the pumping source. An obvious conclusion is that the optical energy generated by stimulated emission cannot be more than that expended by the pump with the overall efficiency of the laser being the ratio of the output to the input. The overall (or wall-plug)

efficiency is a product of factors separated by the physics involved:

$$\eta = \eta_{qe} \cdot \eta_{cpl} \cdot \eta_{\text{pump}} \tag{9.1.1}$$

η_{qe}: : *Quantum efficiency.* This depends solely on the position of the energy levels of the laser and cannot be "designed." It is the upper limit for the performance of any laser.

η_{cpl}: : *Coupling efficiency.* This is an electromagnetic or cavity problem affected by the stimulated emission issue. Optical components are not perfect, and thus we must choose the arrangement of cavity components to maximize the stimulated emission and the output (in a desired direction) while minimizing the useless conversion of photons into heat.

η_{pump}: : *Pumping efficiency.* This is the fraction of the total pump power that is useful in creating the population inversion and thus contributes to the output. It is the most complicated of all topics but is the most essential part of any laser.

9.1.2 Two, 3, 4 . . . , *n* level lasers

Lasers are classified according to the minimum number of levels needed for an approximate analysis, but any practical system has an exceedingly large number. Even if that number is less than 4, there are different permutations and combinations such as that shown in Fig. 9.1, and all can be identified with well known lasers to be discussed later. In all cases, the route for creating the upper level of the population inversion on (2 → 1) is by pumping from the lowest (ground) state to the highest one shown. Any pumping event leading to 2 is good, anything leading to 1 is bad, and a loss of state 2 by anything other than stimulated emission is bad.

The detailed analysis of these cases is reserved for a problem, but some general observations can be made without any mathematics. For starters, we must realize that most of the atoms will reside in the lowest state in the absence of pumping and that the total number of atoms must be conserved. If a pumping event can create an upper state from a

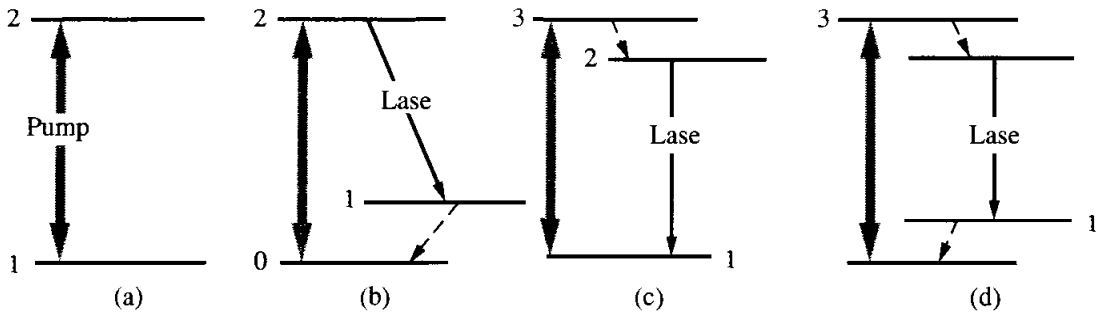


FIGURE 9.1. Possible arrangement of the energy levels of a laser: (a) represents a two level system, (b) and (c) are three level lasers, and (d) is a four-level laser. All double-headed arrows represent the pumping route, the dashed single-headed arrows represent relaxation by any cause, and the solid arrows between 2 and 1 represent stimulated emission by the laser radiation.

lower one, then that same process can relax the upper back to the lower, an obvious fact if optical pumping is considered but holds without exception for other processes also. Hence, Fig. 9.1(a) will not lase. At best we can force $N_2/N_1 = g_2/g_1$ even with an infinite optical pump. Figures 9.1(b) and 9.1(c) appear to be three-level lasers, but they have dramatically different performance characteristics. For Fig. 9.1(c), we must maintain roughly half of the population in state 2 to even think about obtaining a laser on $2 \rightarrow 1$. If the level $E_1 \gg kT$ and there is a fast relaxation of 1 to 0 in Fig. 9.1(b), then the required population in state 2 for optical gain is much less than in Fig. 9.1(c). If $E_1 \sim kT$, then we must pump much harder to obtain gain.

We can quickly compute the quantum efficiency for each case (in Fig. 9.1) by evaluating the photon energy emitted $(E_2 - E_1)$ and dividing it by the energy required to create 2.

$$\eta_a = 1 \qquad n_b = \frac{(E_2 - E_1)}{E_2} \qquad \eta_c = \frac{(E_2 - E_1)}{(E_3 - E_1)} \qquad \eta_d = \frac{(E_2 - E_1)}{(E_3 - E_0)} \tag{9.1.2}$$

The case shown in Fig. 9.1(d) can be used to illustrate the complexity of computing the pumping efficiency, with a similar analysis for the others saved for problems. This is called a "4-level" laser for obvious reasons. Let us presume an optical pump I_p tuned to the 0–3 transition with state 3 relaxing at a rate τ_3^{-1} (by whatever process) to 2, 1, and back to 0 with branching ratios ϕ_{32}, ϕ_{31}, ϕ_{30}, respectively. Assume state 2 decays to 1 at a rate τ_2^{-1} by radiation with a branching ratio of ϕ_{21}, with the rest going to 0, and that state 1 decays at a rate τ_1^{-1} back to zero. Such an "academic" model is very close to that of a semiconductor laser pumping a Nd:YAG crystal, which will be analyzed in detail in chapter 10. The goal here is to formulate the problem (and also to practice using the rate equations).

$$\frac{dN_3}{dt} = \frac{\sigma_{30} I_p}{h\nu_{30}} \left(\frac{g_3}{g_0} N_0 - N_3 \right) - \frac{N_3}{\tau_3} \tag{9.1.3}$$

$$\frac{dN_2}{dt} = \phi_{32} \frac{N_3}{\tau_3} - \frac{N_2}{\tau_2} - \frac{\sigma_{21} I_1}{h\nu_{21}} \left(N_2 - \frac{g_2}{g_1} N_1 \right) \tag{9.1.4}$$

$$\frac{dN_1}{dt} = \phi_{31} \frac{N_3}{\tau_3} + \phi_{21} \frac{N_2}{\tau_1} - \frac{N_2}{\tau_1} + \frac{\sigma_{21} I_1}{h\nu_{21}} \left(N_2 - \frac{g_2}{g_1} N_1 \right) \tag{9.1.5}$$

$$[N] = N_0 + N_1 + N_2 + N_3 \qquad \text{(conservation of atoms)} \tag{9.1.6}$$

where σ_{30} is the stimulated emission cross section at the pump wavelength, σ_{21} is that for the laser at ν_{21}, and the g's are the degeneracies of the respective states. There are various pitfalls that inflict the uninitiated and are noted in the following.

1. Some do not include the stimulated emission rate (by the pump) returning atoms from 3 back to 0. If $N_3 \ll (g_3/g_0)N_0$, then this is legitimate but the results are then limited to weak pumping. If $I_p \gg h\nu_{30}/(\sigma_{30}\tau_3)$, then $N_3/N_0 \rightarrow g_3/g_0$, a result to be proved in a problem. Thus even in the limit of an infinite pump intensity, we do not put all the atoms in state 3.

2. If we are only interested in a threshold calculation (and that is the *first* quantity to be computed for any laser), then we can neglect stimulated emission by the laser at ν_{21}, the last terms in (9.1.4) and (9.1.5). As has been stated previously, stimulated emission is only important above threshold.
3. The positive terms in (9.1.3) and (9.1.4) correspond to the pumping rates R_2 and R_1 of Chapter 8.
4. Thus the analysis of Sec. 8.3 is applicable (for a steady state situation). For transient cases, we must solve (9.1.3) to (9.1.5) directly.
5. It is not necessary to have a separate rate equation for N_0 provided we conserve atoms as is expressed by (9.1.6).

Notice that the optical pump contributes atoms to state 2 (which is good) and indirectly to state 1 (which is bad). If we use a nonselective pump, such as a flash lamp with a distribution of wavelengths or a discharge with electron of many energies, then state 1 can be pumped directly. The fraction of the input power used to excite 2 constitutes the pumping efficiency.

If we assume steady state and ignore stimulated emission on the $2-1$ route (threshold calculation) then the populations are

$$N_3 = \frac{g_3}{g_0} \cdot \frac{I_p/I_{sp}}{1 + I_p/I_{sp}} \cdot N_0 \tag{9.1.7}$$

$$N_2 = \phi_{32}\frac{\tau_2}{\tau_3} N_3 = \phi_{32}\frac{\tau_2}{\tau_3} \cdot \left[\frac{g_3}{g_0}\frac{I_p/I_{sp}}{1 + I_p/I_{sp}}\right] \cdot N_0 \tag{9.1.8}$$

$$N_1 = \phi_{31} \cdot \frac{\tau_1}{\tau_3} \cdot N_3 + \phi_{21}\frac{\tau_1}{\tau_2} \cdot N_2 = (\phi_{31} + \phi_{32}\phi_{21}) \cdot \frac{\tau_1}{\tau_3} \cdot N_3 \tag{9.1.9}$$

where $I_{sp} = h\nu_p/\sigma_{30}\tau_3$, a saturation intensity for the pump. The small-signal gain is given by

$$\frac{\gamma_0}{\sigma_{21}} = \left(N_2 - \frac{g_2}{g_1}N_1\right)_0 = \left[\phi_{32}\frac{\tau_2}{\tau_3} - \frac{g_2}{g_1}\left(\phi_{31} + \phi_{32}\phi_{21}\right)\frac{\tau_1}{\tau_3}\right] \cdot N_3 \tag{9.1.10}$$

Notice that N_2 can be made much larger than even N_0 if $\phi_{32} \sim 1$ (a favorable branching ratio) and $\tau_2/\tau_3 \gg 1$ (a favorable lifetime ratio). This is a distinct advantage of the system shown in Fig. 9.1(d) over that in Fig. 9.1(b).

9.2 CW LASER

Our next task is to relate the output power (or energy in the case of pulsed excitation) to the small-signal gain coefficient and the saturation intensity (or energy) of the system. In all of the following parts of this section we assume the model of Sec. 8.3 for the atomic physics part of the problem and consider γ_0 and I_s as being known and thus the pumping efficiency question is bypassed. We also assume homogeneous broadening and oscillation close to line center so that $\overline{g}(\nu) = 1$. (The case of inhomogeneous broadening and $\nu \neq \nu_0$

is reserved for more advanced textbooks.) Our goal is to compute the output intensity for various types of laser configurations.

Before we jump into too much mathematics, it is appropriate to talk our way through the "starting" of a laser so as to appreciate the physics involved. (The details will be addressed in Sec. 9.3).

Lasers operate by using stimulated emission that adds photons at the same frequency, phase, polarization, and direction of those doing the stimulation. Obviously, we need an initial "seed" of photons, which is usually provided by spontaneous emission from the upper state, although there are cases where it is supplied by another laser.

However, the fraction of the spontaneous emission appearing in the frequency bandwidth of the cavity mode with the proper polarization is an exceedingly small fraction of the total spontaneous power (which goes into 4π). For the system to be considered a laser, this seed has to be amplified by making multiple passes through the gain medium so that the stimulated emission rate becomes comparable to or larger than the other processes. Traditional wisdom (roughly the intuition of the early workers) indicates that the seed must be amplified by a factor of exp[20 to 30] $\sim 10^8$ to 10^{13}, and this takes time.

If the net round trip amplification is expressed as $\exp[+g_{\text{net}}]$, then the number of round trips required is

$$n \sim \frac{20 \text{ to } 30}{g_{\text{net}}} \tag{9.2.1}$$

and thus it takes $n \cdot \tau_{RT}$ to establish a laser. For instance, consider a low gain laser with a net amplification of 1.1 per round trip (and thus $g_{\text{net}} \sim 0.1$) and thus it would take 200 to 300 round trips to establish a laser. If the round trip distance were 24 inches, then $\tau_{RT} \sim 2$ and the time scale is 0.4 to 0.6 μs.

This is a fundamental problem for all pulsed excited lasers: The product of the gain and its duration must be large enough to allow the lasing to develop. If the seed is provided by another laser and is much larger than that provided by spontaneous emission, the time scale can be shortened accordingly.

9.2.1 Traveling Wave Ring Laser

The simplest case that yields an exact analytic solution is that of a unidirectional traveling wave ring laser shown in Fig. 9.2. It is presumed that there is some device incorporated inside the cavity that has an anisotropic attenuation coefficient, and as a result oscillation is constrained to be in *one* direction, such as the counterclockwise one shown. As will become apparent, the one direction is a key simplification used to minimize the mathematical complexity. It is also assumed that there is no intrinsic loss inside the gain medium, but there can still be internal losses inside the cavity because of imperfect reflectivities of the turning mirrors, losses in the "diode" that favors the counterclockwise oscillation, and other parasitic losses. All of these losses are assumed to be independent of intensity.

Fig. 9.2(b) is intended to be a guide for the mathematics and sketches the intensity of the wave as it propagates around the loop. We assume an intensity I_1 at $z = 0$, just inside the gain cell. It obviously gets bigger going the distance l_g to the output of the gain cell

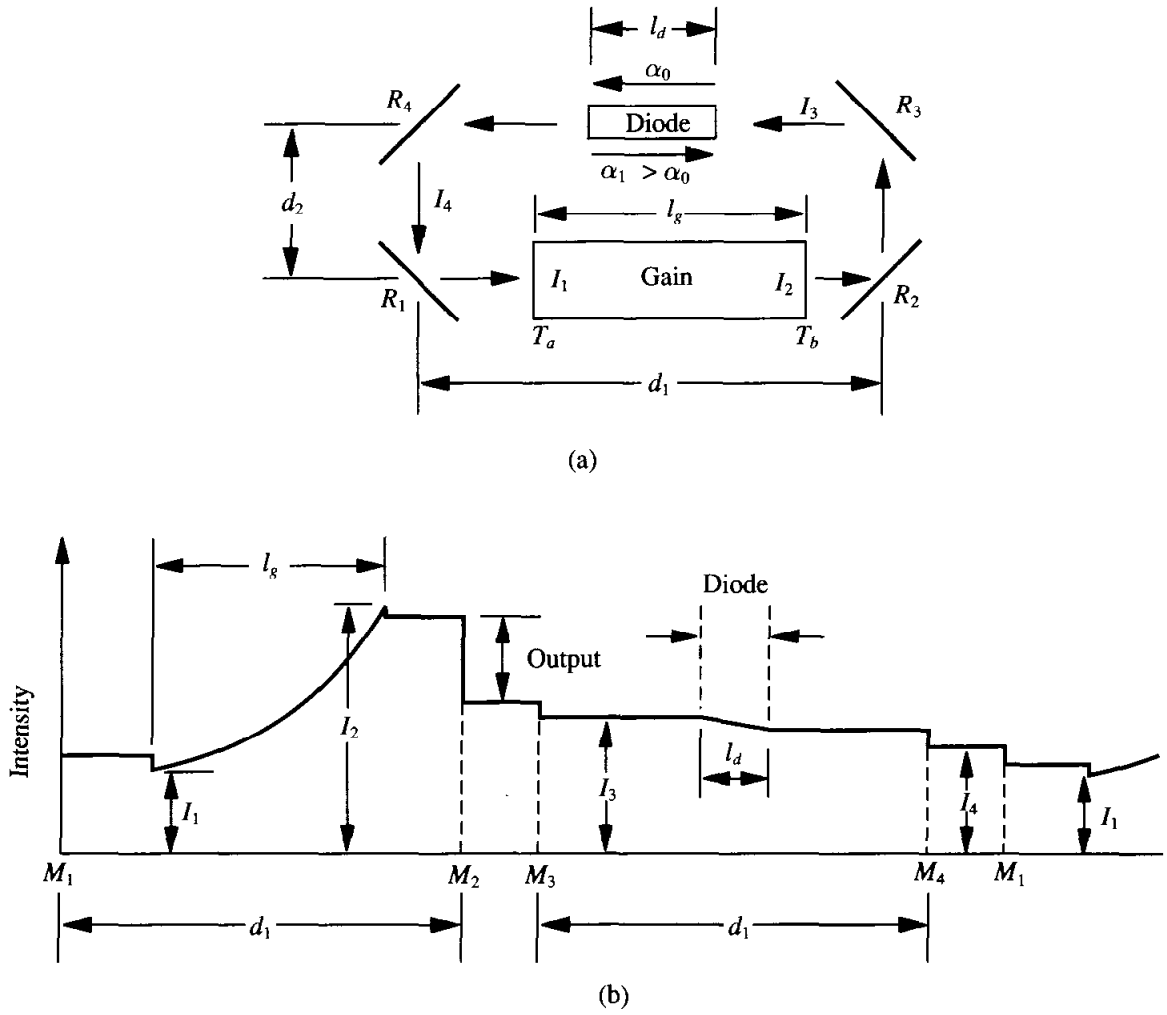


FIGURE 9.2. Unidirectional traveling wave laser. (a) The geometry. (b) A self-consistent variation of the intensity inside the laser cavity.

where it is I_2. Then it gets progressively smaller because of the mirror reflectivities and the attenuation of the "diode." In order to have a steady state or CW laser, the product of all the survival factors for the photons in the remainder of the cavity times I_2 must reproduce the assumed initial intensity I_1. In other words, the solution must be self-consistent.

Now let us assume that all transients have died out and demand self-consistency. In steady state, intensity in the gain cell is amplified according to (8.3.15) [with $\overline{g}(\nu) = 1$].

$$\frac{dI_{ccw}}{dz} = \frac{\gamma_0 I_{ccw}}{1 + I_{ccw}/I_s} \tag{9.2.2}$$

and thus I_2 is related to I_1 by the following [see the mathematics used in going from (8.3.15) to (8.3.17):

$$\ln\left(\frac{I_2}{I_1}\right) + \frac{1}{I_1}(I_2 - I_1) = \gamma_0 l_g = \ln\left(\frac{I_2}{I_1}\right) + \frac{I_2}{I_s}\left(1 - \frac{I_1}{I_2}\right) \tag{9.2.3}$$

Now we follow the intensity I_2 around the loop and back to the entrance of the gain medium to find:

$$I_1 = \left[\prod_i R_i \prod_j T_j \exp[-\alpha_d l_d]\right] I_2 \triangleq S I_2 \tag{9.2.4}$$

where S is the fraction of photons surviving a round trip in the *passive cavity*. Substitute (9.2.4) into (9.2.3)

$$\ln\left(\frac{1}{S}\right) + \frac{I_2}{I_s}(1 - S) = \gamma_0 l_g$$

or

$$\frac{I_2}{I_s} = \frac{\gamma_0 l_g - \ln(1/S)}{(1 - S)} \tag{9.2.5a}$$

The logarithm in the numerator can be expressed in a more meaningful form by recognizing that the threshold is defined by R.T.G. $= 1$ or $\{S \cdot \exp[\gamma_{th} l_g]\} = 1$ and hence $\ln(1/S) = \gamma_{th} l_g$.

$$\therefore \frac{I_2}{I_s} = \frac{[\gamma_0 l_g - \gamma_{th} l_g]}{(1 - S)} = \frac{[\gamma_0 l_g - \gamma_{th} l_g]}{1 - \exp[-\gamma_{th} l_g]} \tag{9.2.5b}$$

The output intensity (through M_2) is found by propagating the intensity I_2 through the window b and through M_2 with transmission T_2

$$I_2 = T_b T_2 I_s \frac{[\gamma_0 l_g - \gamma_{th} l_g]}{(1 - S)} = I_{\text{out}} \tag{9.2.6}$$

This equation exhibits a very important and characteristic variation with external excitation. Let us list some of the more obvious features.

1. The equation makes absolutely no sense if the small signal gain coefficient is less than the threshold value (one cannot have a negative intensity!). Thus the computation of the threshold gain coefficient, γ_{th}, is always the first step in any analysis.
2. The threshold value γ_{th} and the photon survival factor S are functions of the passive cavity.
3. There is always some "knob" that increases the power input and small signal gain coefficient γ_0 when it is twisted clockwise. However, external power must be expended to bring the system to the threshold and to the point of being interesting. The efficiency is essentially zero until threshold is crossed.
4. The power output of any laser will exhibit dramatic increase in intensity (say, by factors of 10^8) for a small fractional increase in pumping above the threshold value. The sharp break is just a manifestation of the use of any excess population by stimulated emission into one mode as opposed to allowing the spontaneous to go into 4π directions.

9.2.2 Optimum Coupling

A problem of interest is to determine the value of the transmission coefficient of M_2 that yields the maximum output intensity. If T_2 is too large, then the losses exceed the gain and no lasing is possible. If T_2 is too small, then the laser may be oscillating "brightly" inside the cavity but little is coupled out. Clearly there is an optimum coupling for maximum output. For those of masochistic tendencies, the exact answer requires differentiating (9.2.6) with respect to $T_2 = 1 - R_2$, setting the expression equal to zero, and solving for T_2, a most unpleasant task to do exactly. However, with some judicious approximations, this task can be simplified considerably, yielding reasonably accurate answers:

1. Low loss or high Q approximation. If the optical cavity has a high Q, the survival factor S is close to 1, and thus $(1 - S)$ = loss per round trip is a small fraction. Hence, the logarithm in (9.2.5a) can be rewritten and expanded in a Taylor series.

$$\ln \frac{1}{S} = \ln \frac{1}{1 - (1 - S)} \approx \ln[1 + (1 - S)] \approx 1 - S$$

Hence (9.2.6) becomes

$$I_{\text{out}} = (T_b I_s) \left\{ T_2 \left[\frac{\gamma_0 l_g - (1 - S)}{(1 - S)} \right] \right\} = (T_b I_s) \left\{ T_2 \left[\frac{g_0}{L + T_2} - 1 \right] \right\} \tag{9.2.7}$$

where $g_0 = \gamma_0 l_g$; the small signal line integrated gain L is the fraction wasted internally per round trip caused by imperfect alignment, bad optical components, dirt, dust, residual absorption; and T_2 is the fraction coupled through M_2 and it is assumed that S is sufficiently close to 1 that the following approximation is valid.

$$(1 - S) = 1 - \prod_i R_i \prod_j T_j \exp[-\alpha_d l_d] = 1 - R_1(1 - T_2)R_3R_4T_aT_b \exp[-\alpha_d l_d]$$

$$\approx 1 - R_1R_3R_4T_aT_b \exp[-\alpha_d l_d] + T_2 = L + T_2$$

Then the output is given by

$$I_{\text{out}} = (T_b I_s) \left\{ T_2 \left[\frac{g_0}{(L + T_2) - 1} \right] \right\} \tag{9.2.8}$$

Now one can differentiate to find a maximum without a lot of fuss.

$$\frac{dI_{\text{out}}}{dT^2} = T_b I_s \left[\frac{g_0 L}{(L + T_2)^2} - 1 \right] = 0 \qquad \text{for max}$$

or

$$T_2(\text{opt}) = -L + (g_0 L)^{1/2} \tag{9.2.9}$$

It is instructive to reinsert (9.2.9) into (9.2.8) to find an expression for the output intensity of the laser with optimum coupling.

$$I_{\text{out}}(max) = T_b I_s \left\{ g_0^{1/2} - L^{1/2} \right\}^2 \tag{9.2.10a}$$

Again we note that this only makes sense provided the integrated gain g_0 is greater than the losses L. If we go one step further in our approximations and assume $g_0^{1/2} \gg L^{1/2}$ (i.e., high gain compared to internal losses) then the output becomes

$$I_{\text{out}}(max) \cong T_b I_s g_0 = T_b[\gamma_0 l_g I_s] \tag{9.2.10b}$$

We should recognize the term in the brackets as the limiting additive intensity one can extract from an amplifier by stimulated emission (compare with (8.3.19)). Although the simplified analysis suffers from the approximations, it is reasonably accurate and is usually used for the first guess, even for a standing wave laser covered in Sec. 9.2.3.

2. High gain and high loss. Some lasers have very high gains and thus permit a significant output coupling. A semiconductor laser is a prime example. Although such lasers are seldom made in the ring configuration, it is appropriate to consider that limit because the conclusions are valid for the simple two mirror lasers.

For a heavily coupled laser or one with high residual internal losses, the fraction of the photon surviving a round trip is small, and hence the denominator of (9.2.5a) can be replaced by 1. If we define an internal loss coefficient $\alpha_{\text{int}} l_g$ to account for the waste of photons, then S can be written as

$$\begin{aligned} S &= \{R_1 R_3 R_4 T_a T_b \exp[-\alpha_d l_d]\}\,(1 - T_2) \\ &\triangleq \{\exp[-\alpha_{\text{int}} l_g]\}\,(1 - T_2) \\ \therefore \ln \frac{1}{S} &= \alpha_{\text{int}} l_g + \ln \frac{1}{1 - T_2} \end{aligned}$$

Hence, (9.2.6) becomes

$$I_{\text{out}} \approx (T_b I_s)\left\{T_2\left[(\gamma_0 l_g - \alpha_{\text{int}} l_g) - \ln\left(\frac{1}{1 - T_2}\right)\right]\right\} \tag{9.2.11}$$

After differentiating to find an optimum, we obtain a transcendental equation for T_2:

$$\frac{T_2}{1 - T_2} + \ln \frac{1}{1 - T_2} = (\gamma_0 l_g - \alpha_{\text{int}} l_g) = \text{net integrated gain} \tag{9.2.12}$$

A solution for this coupling is shown in Fig. 9.3. Note that the optimum transmission is quite high for a system with high integrated gain.

A semiconductor laser has such high integrated gain for injected currents above threshold. Consequently, many are made with an antireflection coating on the output facet. Of course, we could not use zero reflection for then there would be no feedback. (More to the point is the fact that it is extremely hard to obtain perfect antireflection coatings for the modes generated by a semiconductor laser.)

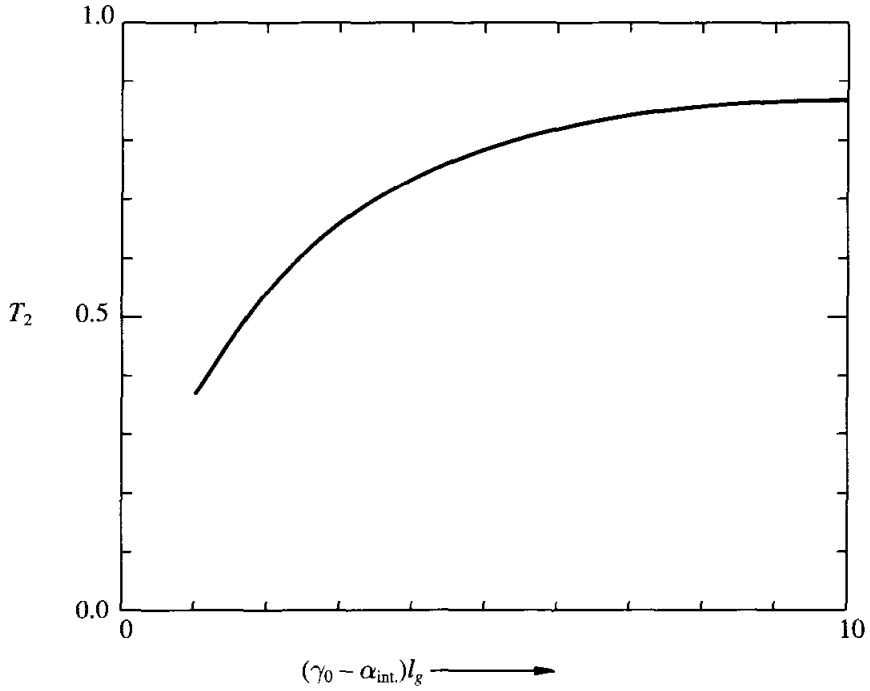


FIGURE 9.3. Optimum coupling for a high gain–high loss laser.

9.2.3 Standing Wave Lasers

Most lasers are of the simple variety with just two mirrors for feedback in the manner shown on page 270 in Fig. 9.4(a), with Fig. 9.4(b) showing the forward (I^+) and reverse (I^-) waves inside the cavity, and Fig. 9.4(c) illustrating a zeroth-order guess as to the spatial variation of those intensities.

We start a wave $I^+(z = M_1)$, just to the right of M_1 that propagates to the entrance of the gain cell where its amplitude drops because of the imperfect transmission of the gap (windows, scattering, or absorption in that space). Once the forward-going wave is inside the gain cell, it is amplified because of the population inversion until that wave exits at the other end with a corresponding decrease at that window. The wave propagates with a transmission coefficient T_b to the mirror M_2 where only a fraction $R_2 I^+(z = M_2)$ starts the return path and is now identified as I^-. The wave propagating in the negative direction undergoes a similar set of attenuations and amplifications as did I^+ until it is reflected by M_1 and that amplitude must *reproduce* the assumed initial value of I^+. The object of what follows is to translate this narrative into a simplified mathematical model.

Within the gain medium, the rate of change of the forward and reverse intensities with z is given by

$$\frac{1}{I^+}\frac{dI^+}{dz} = \frac{\gamma_0}{1 + (I^+ + I^-)/I_s} = -\frac{1}{I^-}\frac{dI^-}{dz} \tag{9.2.13}$$

where the minus sign on the right-hand side of (9.2.13) reflects the fact that I^- increases as z decreases. Much to our dismay, we see that the amplification of the forward wave I^+ depends on both I^+ and I^-. Now each atom has *one* package of energy to give; it can give it to the forward wave or the reverse but not to both. This unpleasant mathematical complication has to be faced and makes the exact analysis of this simple laser much more

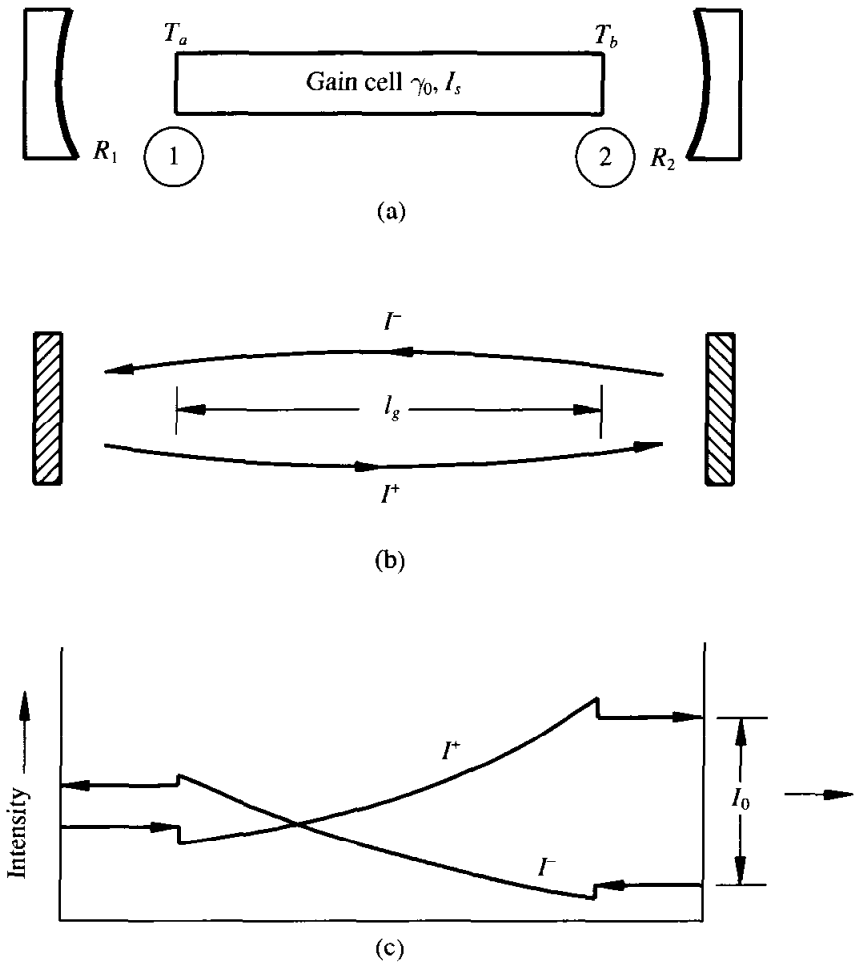


FIGURE 9.4. A simple standing wave laser. (a) The cavity showing the mirror reflectivities and the transmission coefficients between the mirrors and the gain cell. (b) The counterpropagating intensities. (c) The spatial variation of the waves inside the cavity.

complicated than that of the traveling wave ring laser of the previous section. First it is appropriate to search for an approximate but revealing approach.

9.2.3(a) High *Q* approximate solution. A key simplification results from assuming a high Q passive cavity, because then the losses are small, and as a result the variation of the intensities within the cavity in order to have a self-consistent solution is also small. Thus it is reasonable to neglect the variation of the intensities with z and set the saturated gain coefficient equal to the losses, which are to be prorated on a per unit of length of the active medium (even though most of those losses appear elsewhere inside the cavity).

$$\frac{dI^+}{dz} = \frac{\gamma_0 I^+}{1 + (I^+ + I^-)/I_s} - (\alpha_{\text{int}} + \alpha_{\text{ext}})I^+ \simeq 0 \tag{9.2.14}$$

There are several points to be noted about this equation. First, as was noted previously, we have assumed oscillation close to line center so that the frequency dependence of the saturation and the gain coefficient can be ignored [$\overline{g}(\nu) = 1$]. Second, both the forward and reverse waves saturate the gain of the forward one, and similar considerations apply

to the variation of the reverse wave.* Third, the total loss coefficient is composed of two parts: (1) that which is internal to the laser medium itself or due to internal parts of the laser cavity such as imperfect window transmission, scattering from misaligned or rough surfaces, or anything that represents a useless conversion of photons to heat and (2) the external losses that represent the coupling to the outside world by the intentional choice of a lower reflectivity for mirror M_2 [see (9.2.7)].

In the spirit of this high Q approximation, we can assume that the forward and reverse waves are equal in amplitude, and thus the saturating term in the denominator of (9.2.14) is just twice one of them. Solving that equation for the mean circulating intensity yields

$$I = \frac{I_s}{2}\left[\frac{\gamma_0 \cdot 2l_g}{\alpha_T \cdot 2l_g} - 1\right] \tag{9.2.15a}$$

and the output through M_2 is

$$I_{\text{out.}} = \frac{T_b I_s}{2}\left[T_2\left(\frac{\gamma_0 \cdot 2l_g}{\alpha_T \cdot 2l_g} - 1\right)\right] = \frac{T_b I_s}{2}\left[T_2\left(\frac{g_0}{L+T} - 1\right)\right] \tag{9.2.15b}$$

Equation (9.2.15b) is identical to (9.2.8) for the ring laser, aside from the extra factor of 2. Even that difference can be explained quite easily:

> The internal loss terms express the integrated gain or loss over a round trip $2l_g$. The external factor of 2 appears because stimulated emission goes into two directions for the simple lasers but only into one for the ring laser. This points out one of the advantages of the unidirectional geometry.

All of the conclusions from Sec. 9.2.2 regarding optimum coupling apply here and thus will not be repeated.

9.2.3(b) "Exact" analysis (after Rigrod). If the output mirror has a relatively high transmission coefficient, the approximation of setting the derivative $\partial/\partial z = 0$ is not valid, and we must face up to the fact that the gain coefficient is saturated by both waves and suffer through the mathematical complications [8]. We return to (9.2.13) and normalize the intensities to the saturation value and use a set of mnemonic symbols: $\mathbf{f}(\text{forward}) = I^+/I_s$ and $\mathbf{r}$ (reverse) $= I^-/I_s$. Thus (9.2.13) becomes

$$\frac{1}{\mathbf{f}}\frac{d\mathbf{f}}{dz} = \frac{\gamma_0}{1+\mathbf{f}+\mathbf{r}} = -\frac{1}{\mathbf{r}}\frac{d\mathbf{r}}{dz} \tag{9.2.16}$$

where the minus sign acknowledges the fact that the reverse wave is amplified as it propagates in the negative z direction. Combining the outside equality yields a very important relationship between $\mathbf{f}$ and $\mathbf{r}$.

$$\frac{1}{\mathbf{f}}\frac{d\mathbf{f}}{dz} = -\frac{1}{\mathbf{r}}\frac{d\mathbf{r}}{dz} \qquad \frac{d}{dz}[\ln(\mathbf{f}\cdot\mathbf{r})] = 0$$

*Actually there is interference between the fields associated with I^+ and I^-, which leads to a spatial variation of the saturation. However, this is a complication best left for more advanced treatments.

or

$$\mathbf{f} \cdot \mathbf{r} = k^2 \qquad \text{(a constant independent of } z\text{)} \tag{9.2.17}$$

Thus even though both waves are functions of z, their product is an (unknown) constant to be evaluated from the boundary conditions at the mirrors. From Fig. 9.4(c), we note that

$$\mathbf{f}(1) = R_1'\mathbf{r}(1) \tag{9.2.18a}$$

$$\mathbf{r}(2) = R_2'\mathbf{f}(2) \tag{9.2.18b}$$

where the number in parentheses refers to the planes of Fig. 9.4 for which the boundary conditions apply, and the primes on the reflectivities are used to indicate the *fraction* of the intensity returning from each mirror. For instance, if the one-way transmission coefficient between M_1 and the gain cell were T_a then the quantity R_1' would be $T_a^2 R_1$. These boundary conditions can be combined to evaluate the constant k^2.

$$k^2 = \mathbf{f}(1)\mathbf{r}(1) = \mathbf{r}(2)\mathbf{f}(2) \tag{9.2.19a}$$

$$= \frac{\mathbf{f}^2(1)}{R_1'} = R_2'\mathbf{f}^2(2) \tag{9.2.19b}$$

$$= R_1'\mathbf{r}^2(1) = \frac{\mathbf{r}^2(2)}{R_2'} \tag{9.2.19c}$$

The combination of (9.2.19c) and (9.2.19b) yields the noteworthy fact that

$$\frac{\mathbf{f}(2)}{\mathbf{f}(1)} = \frac{1}{\sqrt{R_1'R_2'}} = \frac{\mathbf{r}(1)}{\mathbf{r}(2)} \tag{9.2.20}$$

Equation (9.2.20) merely states that the net one-way amplification factor of each of the waves, $\mathbf{f}(2)/\mathbf{f}(1)$ or $\mathbf{r}(1)/\mathbf{r}(2)$, must be equal to the mean fraction of the photons returning from the mirrors.

To integrate (9.2.16) in terms of simple and uncomplicated functions, the loss coefficient internal to the laser medium α_0 is set equal to zero. Substitute $\mathbf{r} = k^2/\mathbf{f}$ for the left-hand side of (9.2.16) (or $\mathbf{f} = k^2/\mathbf{r}$ on the right) and integrate between the planes (1) and (2) to obtain

$$\ln\left[\frac{\mathbf{f}(2)}{\mathbf{f}(1)}\right] + [\mathbf{f}(2) - \mathbf{f}(1)] + k^2\left[\frac{1}{\mathbf{f}(1)} - \frac{1}{\mathbf{f}(2)}\right] = \gamma_0 l_g \tag{9.2.21a}$$

$$\ln\left[\frac{\mathbf{r}(2)}{\mathbf{r}(1)}\right] + [\mathbf{r}(2) - \mathbf{r}(1)] + k^2\left[\frac{1}{\mathbf{r}(1)} - \frac{1}{\mathbf{r}(2)}\right] = -\gamma_0 l_g \tag{9.2.21b}$$

Now we are primarily interested in $\mathbf{f}(2)$, the forward wave I^+/I_s, so (9.2.19b) is used to eliminate k^2 and (9.2.20) is used to express $\mathbf{f}(1)$ in terms of $\mathbf{f}(2)$. The solution for $\mathbf{f}(2)$ is

$$\frac{I^+}{I_s} = \mathbf{f}(2) = \frac{\gamma_0 l_g - \frac{1}{2}\ln[1/R_1'R_2']}{\left(1 - \sqrt{R_1'R_2'}\right)\left(1 + \sqrt{R_2'/R_1'}\right)} \tag{9.2.22}$$

Now it is just a matter of propagating this wave through the gap between the active medium and M_2 (call it a transmission coefficient T_b) where only a fraction $1 - R_2$ is available for the output. Eliminating the primes on the effective reflectivities yields an expression for the output intensity in terms of the gain coefficient, saturation intensity, and cavity parameters.

$$\frac{I_{\text{out.}}}{I_s} = \frac{T_b T_2 \left[\gamma_0 l_g - \frac{1}{2} \ln\left(1/T_a^2 T_b^2 R_1 R_2\right)\right]}{\left(1 - \sqrt{T_a^2 T_b^2 R_1 R_2}\right)\left(1 + \sqrt{T_b^2 R_2 / T_a^2 R_1}\right)} \tag{9.2.23}$$

It is not a pleasant task to differentiate this equation to find the coupling that yields maximum output power; rather, the result is shown graphically in Fig. 9.5 and compared to the high Q approximate analysis given earlier in this section. It should be clear from this figure that the approximate analysis is fairly accurate overall but especially good at predicting the optimum coupling.

Example

Choose a simple laser with

$$T_a = 0.995; \quad R_1 = 0.99$$

$$T_b = 0.995; \quad R_2 \text{ variable from 1.0 to 0.5}; \quad T_2 = 1 - R_2$$

$$\gamma_0 l_g = 0.30 \quad \text{(i.e.,the integrated gain is equal to 30 \% per pass)}$$

The total losses are

$$\alpha_T 2l_g = \ln\left[\frac{1}{T_a^2 T_b^2 R_1}\right] + \ln\frac{1}{R_2}$$

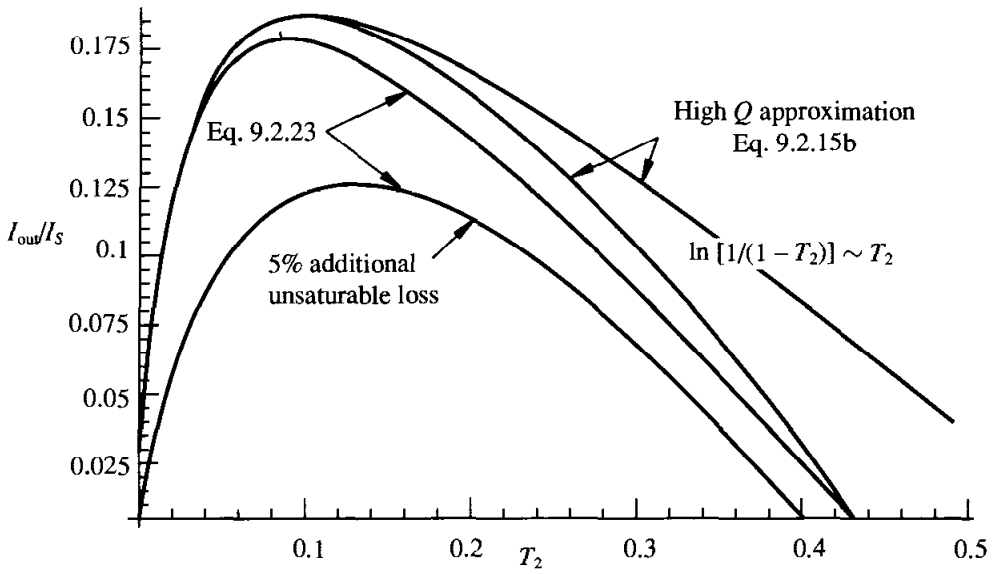


FIGURE 9.5. Output power as a function of coupling.

Thus the parameters for use in (9.2.15b) are

$$L = \ln \frac{1}{T_a^2 T_b^2 R_1} \qquad T = \ln \frac{1}{R_2} \sim T_2$$

Also demonstrated in Fig. 9.5 is the sensitivity of the power to the internal nonsaturable losses inside the laser cavity. Even though most of the total loss per round trip is due to the coupling through M_2 (i.e., T_2), the change of 5% in additional internal losses makes an appreciable change in the maximum. Although the example considers the effect of window losses or mirror imperfections, it should be clear that any nonsaturable loss placed anywhere would have a similar detrimental effect.

9.3 LASER DYNAMICS

9.3.1 Introduction and model

The prior sections presumed a steady state (CW) situation where there is a balance between the pumping mechanism and the use of the inversion by stimulated emission. There are some insights to be gained by examining dynamic phenomena in lasers to include the buildup of the coherent radiation caused by pulsed excitation or a "small" change in excitation, either as a step function or as a sinusoidal modulation. A unidirectional ring laser is chosen for the analysis so as to avoid the coupling problem found in the last section and side pumping is assumed so that the small signal gain coefficient is independent of z. The model is shown in Fig. 9.6 where many of the characteristics will be chosen to be close to that of the popular optically pumped YAG laser.

Let us assume that the upper state of the laser is pumped optically directly from the ground state by a diode laser with an intensity I_p rather than to state 3 (> 2) followed by relaxation back to 2 as it occurs for the practical laser. This scenario simplifies the problem and avoids the dynamics of the relaxation from a state above 2 (and also saves something for a problem). The atoms in state 2 can be utilized by stimulated emission or relax by other causes to 1 or 0. To make the analysis as simple and transparent as possible, assume $\tau_1 \ll \tau_2$

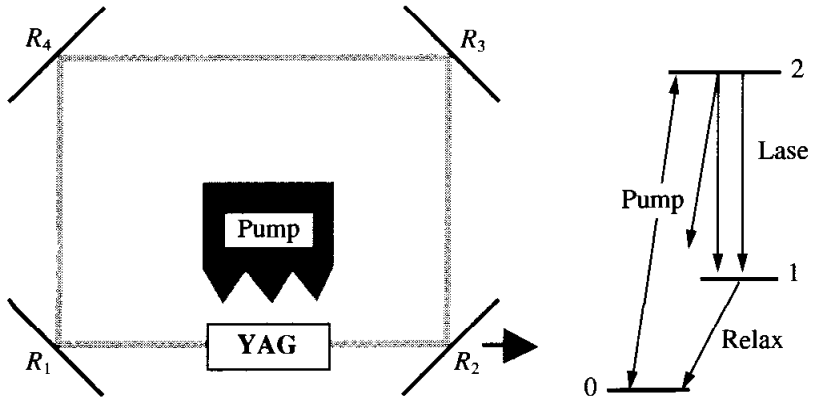


FIGURE 9.6. The laser system to be analyzed.

and thus $N_1 \ll N_2$ and hence the line-integrated gain is related to N_2 by $g(t) = N_2(t)\sigma_L l_g$*.

$$\frac{dN_2}{dt'} = \frac{\sigma_p I_p}{h\nu_p} N_0 - \frac{N_2}{\tau_2} - \frac{\sigma_L I_L}{h\nu_L} N_2 = \frac{\sigma_p I_p N_0}{h\nu_p} - \frac{N_2}{\tau_2}\left(1 + \frac{I_L}{I_s}\right) \tag{9.3.1}$$

where σ_p, λ_p, and I_p are the absorption cross section, the pump wavelength ($0 \rightarrow 2$), and the pump intensity, respectively. While the Nd:YAG system is emphasized here, the model applies equally well for many types of lasers. For instance, (9.3.1) applies to a semiconductor laser provided we assign the proper interpretation to the parameters. Our goal is to follow the inversion as well as the photons with time.

The dynamics of the photons are described by a formulation similar to that used in Sec. 6.4 for the definition of the photon lifetime. The cavity is treated as a circuit element by averaging changes occurring on the time scale of a round trip transit time. Thus, the energy inside the cavity w (per unit of area) is increased by the net stimulated emission by the energy present in the cavity and also by that added by spontaneous emission.† To avoid the crush of symbols, denote $S = 1 - L$ as the survival factor of the photons inside the cavity representing residual absorption, coupling, and imperfect reflectivities.

$$\frac{dw}{dt'} = \frac{(S\exp[+N_2\sigma_L l_g] - 1)}{\tau_{RT}} w + \beta h\nu \frac{N_2 l_g}{\tau_2} \tag{9.3.2}$$

where β is the fraction of the decay of state 2 (by all causes) that yields a photon of energy, $h\nu$, into the cavity mode and w (joules/cm^2) $= I_L \cdot \tau_{RT} = h\nu N_p$.

It is important to agree with the functional form of (9.3.2) before proceeding. In one round trip, the initial photons inside the cavity are amplified by $\exp(+N_2\sigma_L l_g)$ but only a fraction S of those survive a round trip. The net change added is the difference between the start and finish of a round trip taking τ_{RT} seconds. During this time, the upper state ($N_2 \cdot l_g$) is radiating (into 4π steradians) and a (very) small fraction is captured by the cavity mode. Typically, β ranges from 10^{-2} (for a microcavity) to 10^{-8} (for a macrocavity), and thus the last term can safely be neglected under some easily identified circumstances. The spontaneous emission "seed" is important—however small—for the laser will not start without it.

Some obvious definitions and normalizations (using mnemonic symbols to help us remember their meaning) are useful to avoid endless repetitions of constants.

1. $t = t'/\tau_{RT}$: a normalized time.
2. $g = N_2\sigma_L l_g$: the line integrated gain.
3. $P = w/w_s = I_L/I_s$: the relative number of photons inside the cavity and $w_s = I_s \cdot \tau_{RT}$ is a convenient energy for normalization.
4. $R = (N_0\sigma_p l_g) \cdot (\lambda_p/\lambda_L) \cdot (I_p/I_s)$ is a relative pump rate normalized to the saturation intensity of the laser transition. Note the appearance of the line integrated absorption of the pump, $N_0\sigma_p l_g$, and the quantum efficiency λ_p/λ_L.

*(The time is written as t' so that we use a normalized variable t, which will be t'/τ_{RT}. These assumptions are well obeyed for the Nd:YAG system as will be shown in Chapter 10.)

†The energy per unit of area is equal to $h\nu$ times the number of photons/cm^2 or the running wave intensity times the round trip time

5. $a = \tau_{RT}/\tau_2$: a ratio of characteristic times. Usually $a \ll 1$.
6. S = survival factor $= 1 - L$, where L is the fractional *loss* per round trip.

Thus, a multiplication of (9.3.2) by τ_{RT} and a division by $I_s \cdot \tau_{RT}$ yields

$$\boxed{\frac{dP}{dt} = (S \cdot e^g - 1)P + \beta g} \tag{9.3.3}$$

Multiply (9.3.1) by $\sigma_L l_g$ and τ_{RT}; use the definition of $I_s = h\nu/\sigma\tau_2$ from Sec. 8.3 and identify $P = I_L/I_s$, the relative number of photons; set $g = N_2\sigma_L l_g$; and extract the ratio τ_{RT}/τ_2 from the right-hand side to obtain

$$\boxed{\frac{dg}{dt} = a \cdot [R - g \cdot (1 + P)]} \tag{9.3.4}$$

Equations (9.3.3) and (9.3.4) contain a lot of information so it is appropriate to build a base of knowledge as the complexity is increased. (The subscript on the result will denote the case.)

9.3.2 Case a: A sub-threshold system

Suppose the pumping was of the form of a step function $R = R_a u(t)$ but the resulting peak gain is insufficient to bring the laser to threshold. Thus, stimulated emission can be ignored in (9.3.4) ($P \ll 1$), and the time history of the gain is easily found.

$$\frac{dg}{dt} + ag = aR_a u(t) \tag{9.3.5}$$

or

$$g(t) = R_a(1 - \exp[-at]) \tag{9.3.6}$$

and as $t \to \infty$, $g_a = R_a$. Since this case is really not very interesting, we will not beat it to death but merely find the steady state solution for the photons P from (9.3.3).

$$P_a = \frac{\beta R_a}{1 - S\,e^{+g_a}} \xrightarrow[g_a \ll 1 \text{ and} < L]{} \frac{\beta R_a}{1 - S} \tag{9.3.7}$$

Thus, if the pump rate is too small, then the relative number of photons is enhanced by the cavity, but P_a is insignificant compared to 1 because of the small value of β.

9.3.3 Case b: A CW laser: threshold conditions

The pumping rate has to be bigger than a threshold value so as to make $g \geq g_{th}$

$$S\exp[g_{th}] \geq 1 \qquad \text{or} \qquad g_{th} \geq \ln\frac{1}{S} = \ln\frac{1}{1 - L} \approx \ln(1 + L) \approx L \tag{9.3.8}$$

which just expresses the obvious fact that the line integrated gain per round trip must exceed the losses. (This is modified, ever so slightly, by what follows.)

The threshold pumping rate is found from combining (9.3.6) and (9.3.8), and hence the pumping rate must be larger than this minimum value.

$$R_b > g_{th} \tag{9.3.9}$$

For a CW laser, the relative number of photons is significant compared to 1 and is very large compared to that contributed by spontaneous emission for g_s, the steady state saturated gain, whatever it might be.

$$\frac{dP}{dt} = 0 = (S\,e^{+g_s} - 1)P + \beta g_s$$

or

$$(1 - S\,e^{+g_s}) = \beta \frac{g_s}{P} \tag{9.3.10}$$

The right-hand side of (9.3.10) is surely positive and thus so also is the left, which implies that the factor $S \exp[g_s] < 1$, or that the steady state saturated gain g_s must be *slightly less* than the losses. However, the words, *slightly less*, can be quantified a bit further; that is, g_s cannot be much different from $g_{th} = L$, $\beta \approx 10^{-8}$, or so, and for $P \sim 1$:

$$\therefore \frac{\beta g_s}{P} \approx \frac{\beta L}{P} \approx 10^{-8}$$

Thus, with very little approximation, we can set the right-hand side of (9.3.10) equal to zero and thus the *steady-state saturated gain* is equal to the threshold value $g_s = g_{th}$, irrespective of the pumping rate as long as it is bigger than the threshold value. Since the saturated gain is clamped at threshold and is equal to g_{th}, this also implies that

> the inversion density is always clamped at the threshold value irrespective of the pumping rate (for a CW laser).

(This is a very important point that many students resist. You should be comfortable with this statement.) Now we return to (9.3.4) to find an expression for P_b

$$P_b = \frac{R_b}{g_{th}} - 1 \tag{9.3.11}$$

which again expresses the idea that the laser intensity varies linearly with the pump above threshold.

9.3.4 Case c: A sinusoidal modulated pump

Assume adequate pumping, $R_c > R_{th}$, a sufficient time to establish a steady state for both the inversion and the relative photon number, and then allow a time dependent change in the pump specified by

$$R(t) = R_c + \Delta r(t) \tag{9.3.12a}$$

The photons will also change to

$$P(t) = P_c + \Delta p(t) \tag{9.3.12b}$$

where
$$P_c = \frac{R_c}{g_{th}} - 1 \tag{9.3.12c}$$

by (9.3.11) and

$$g(t) = g_{th} + \Delta g(t) \tag{9.3.12d}$$

where it is assumed that the pump was turned on in the distant past so as to obtain a CW laser as in case b.

Equations (9.3.3) and (9.3.4) are coupled and nonlinear, and hence a numerical solution is required for a time-dependent solution with arbitrary conditions but, even so, the CW or DC terms reach an equality as in case (b). If Δr were small, then a linearization technique whereby *products* of small quantities, $\Delta g \cdot \Delta p$, are neglected is useful. The photon equation (9.3.3) becomes

$$\frac{d\Delta p}{dt} = \left[\left(S\, e^{g_{th}} \cdot e^{\Delta g}\right) - 1\right] \cdot (P_c + \Delta p) + \beta(g_{th} + \Delta g)$$

Now $S \exp(g_{th}) \equiv 1$ and $\Delta g \ll 1$ then $e^{\Delta g} \sim 1 + \Delta g$ and the time-dependent part becomes (after neglecting the product $\Delta g \cdot \Delta p$):

$$\frac{d\Delta p}{dt} = \Delta g \cdot (P_c + \beta) \tag{9.3.13}$$

The population inversion equation (or gain) becomes

$$\begin{aligned}\frac{d\Delta g}{dt} &= a\left[R_c + \Delta r - (g_{th} + \Delta g)(1 + P_c + \Delta p)\right] \\ &= a\left\{\left[R_c - g_{th}(1 + P_c)\right] + \left[\Delta r - \Delta g(1 + P_c) - g_{th}\Delta p\right]\right\}\end{aligned}$$

or

$$\frac{d\Delta g}{dt} = a\left[\Delta r - \Delta g(1 + P_c) - g_{th}\Delta p\right] \tag{9.3.14}$$

where the fact that the CW terms vanish has been used.

Let us eliminate $\Delta g(t)$ by differentiating (9.3.13) with respect to time and substituting (9.3.14) to obtain a single equation for Δp.

$$\frac{d^2\Delta p}{dt^2} = (P_c + \beta)a\left[\Delta r - \frac{(1 + P_c)}{(\beta + P_c)} \cdot \frac{d\Delta p}{dt} - g_{th}\Delta p\right]$$

and collect terms in order of descending derivatives to obtain

$$\frac{d^2\Delta p}{dt^2} + a(1 + P_c)\frac{d\Delta p}{dt} + g_{th}a(P_c + \beta)\Delta p = a(P_c + \beta)\Delta r \tag{9.3.15}$$

A useful check is to consider the CW limit where all time derivatives vanish and the answer from case b is applicable

$$P = P_c + \Delta p = \frac{R_c + \Delta r}{g_{th}} - 1$$

and the above yields

$$\lim_{\Delta t \to \infty} \Delta p = \frac{\Delta r}{g_{th}}$$

Now suppose $\Delta r(t)$ had a sinusoidal variation. Since (9.3.15) is linear, we can use the phasor representation for $\Delta r(t)$ with a reminder to take the real part of the result:

$$\Delta r(t) = \{r_m \exp[j\omega_m t]\} \tag{9.3.16a}$$

where ω_m is normalized modulation rate.

NOTE: Since we are dealing with normalized quantities—even time—we are thus faced with normalized or dimensionless modulation frequencies. To go back to real dimensional quantities, we recall that $t = t'/\tau_{RT}$ and that $\omega_m' t'$ must be dimensionless. Hence, the laboratory modulation frequency $\omega' = \omega_m/\tau_{RT}$.

The response must have the same functional form:

$$\Delta p(t) = \{p_m \exp[j\omega_m t]\} \tag{9.3.16b}$$

and thus

$$\frac{g_{th} p_m}{r_m} = \frac{g_{th} a(\beta + P_c)}{\left\{\left[g_{th} a(\beta + P_c) - \omega_m^2\right] + j\omega_m \left[a(1 + P_c)\right]\right\}} \tag{9.3.17}$$

The optical transfer function (i.e., p_m/r_m) has the characteristic behavior (with the constant scale factor g_{th} ignored) shown in Fig. 9.7:

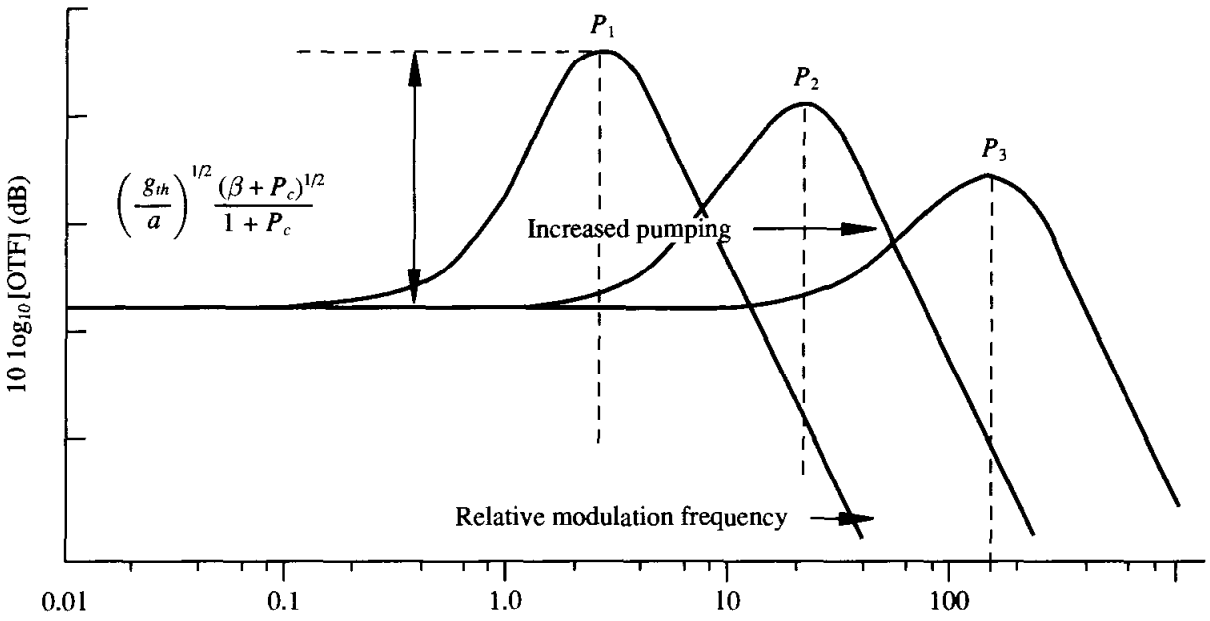


FIGURE 9.7. The relative response of a laser to a sinusoidal modulation of the pump. As the pumping is increased, the power, the resonant frequency, and the resonant width all increase. The "standard" method of plotting the modulation "gain" (in dB relative to that at a low frequency) as a function of the modulation frequency on a logarithmic frequency scale was used.

Equation 9.3.17 and Fig. 9.7 indicate a "resonance" at a normalized modulation frequency given by

$$\omega_r^2 = g_{th}a(\beta + P_c) \approx g_{th}aP_c \tag{9.3.18a}$$

At that resonant frequency, the optical transfer function (OTF) has a value of

$$\begin{aligned}\frac{p_m}{r_m} &= \text{OTF}(\omega_r) = -j\frac{a(P_c+\beta)}{[g_{th}a(P_c+\beta)]^{1/2}[a(1+P_c)]} \\ &= -j\frac{1}{g_{th}}\cdot\left(\frac{g_{th}}{a}\right)^{1/2}\frac{(\beta+P_c)^{1/2}}{1+P_c}\end{aligned} \tag{9.3.18b}$$

The first factor, $-j$, indicates that the photons "lag" the modulation by 90°; the second term $1/g_{th}$, is the transfer function at a very slow modulation rate [(cf. (9.3.11)], and, finally, the product of the last two factors is the resonant (peak or gain) in the transfer characteristic and is given by

$$\text{resonant gain} = \left(\frac{g_{th}}{a}\right)^{1/2}\frac{(\beta+P_c)^{1/2}}{1+P_c} \tag{9.3.18c}$$

For $\omega_m \gg \omega_r$, the response drops off as ω^{-2} and approaches a "slope" of −20 dB/decade as shown in Fig. 9.7.

9.3.5 Case d. A sudden "step" change in excitation rate

Let $R = R_d + \Delta r \cdot u(t)$, with $u(t) =$ the step function, and $R_d > R_{th}$. Hence, $g(t) = g_{th} + \Delta g(t)$ and $P(t) = P_d + \Delta p(t)$, where $P_d = (R_d/g_{th}) - 1$ and the prior analysis, case b, has been used to establish the initial conditions on the photons [cf. (9.3.11)] and, again, the saturated gain is set equal to the threshold value. The Laplace transform ($\mathcal{L}$) technique can be used to solve (9.3.13) and (9.3.14).

$$\mathcal{L}\left\{\frac{d\Delta p}{dt} = \Delta g(P_d+\beta)\right\} \Rightarrow s\Delta p(s) = \Delta g(s)(P_d+\beta) \tag{9.3.19a}$$

$$\begin{aligned}&\mathcal{L}\left\{\frac{d\Delta g}{dt} = a\left[\Delta r(t) - \Delta g(1+P_d) - g_{th}\Delta p\right]\right\} \\ &\Rightarrow s\Delta g(s) = a\cdot\left[\frac{\Delta r}{s} - \Delta g(s)(1+P_d) - g_{th}\Delta p(s)\right]\end{aligned} \tag{9.3.19b}$$

where the initial values of Δg and Δp are both zero.

A rearrangement into matrix form yields

$$\begin{bmatrix} ag_{th} & s + a(1+P_d) \\ s & -(\beta+P_d)\end{bmatrix}\begin{bmatrix}\Delta p(s) \\ \Delta g(s)\end{bmatrix} = \begin{bmatrix}\dfrac{a\Delta r}{s} \\ 0\end{bmatrix} \tag{9.3.20}$$

Thus,

$$\Delta p(s) = \frac{a(\beta + P_d)\Delta r}{s\left[s^2 + a(1 + P_d)s + ag_{th}(\beta + P_d)\right]} \tag{9.3.20}$$

The poles of the response are located at $s = 0$ and

$$s = \frac{-a(1 + P_d)}{2} \pm j\left\{ag_{th}(\beta + P_d) - \left[\frac{a(1 + P_d)}{2}\right]^2\right\}^{1/2} \tag{9.3.21a}$$

$$= -\frac{1}{\tau_d} \pm j(\omega_0 \approx \omega_r) \tag{9.3.21b}$$

Note also that the imaginary part of (9.3.21) is more or less the resonant frequency of case c, [(c.f. (9.3.18a)] if we assumed that the factor $ag_{th}(\beta + P_d) \gg [a(1 + P_d)/2]^2$.

As a check, consider a semiconductor laser of length $d = 400\ \mu$m, $n = 3.6$, $\tau_2 \approx 1$ ns, $\beta \approx 10^{-3}$, $g_{th} \approx L = 0.64$, and a normalized photon number of 2, all typical values (although stretching the limits of applicability of this analysis somewhat).

$\therefore \tau_{RT} = 2nd/c = 9.6$ ps; $a = \tau_{RT}/\tau_2 = 9.6 \times 10^{-3}$

$ag_{th}(\beta + P_d) = 1.23 \times 10^{-2}$; $\left[a(1 + P_d)/2\right]^2 = 2.07 \times 10^{-4} \ll 1.23 \times 10^{-2}$

Hence, the assumption is reasonable.

A partial fraction expansion of (9.3.20) yields

$$\Delta p(s) = \frac{\Delta r}{g_{th}}\left[\frac{1}{s} - \frac{s + 1/\tau_d}{(s + 1/\tau_d)^2 + \omega_0^2} - \frac{1/\tau_d}{(s + 1/\tau_d)^2 + \omega_0^2}\right]$$

where the normalized damping rate $(1/\tau_d)$ and ω_0^2 from (9.3.21b) have been used. The inverse Laplace transform yields:

$$\Delta p(t) = \frac{\Delta r}{g_{th}}\left\{u(t) - e^{-t/\tau_d}\left[\cos\omega_0 t + \frac{1}{\omega_0\tau_d}\sin\omega_0 t\right]\right\} \tag{9.3.22}$$

The first factor in (9.3.22) should be expected because this is the CW solution, and the last two express the time-dependent interchange of population and photons. Let us evaluate the term $1/\omega_0\tau_d$ by using the semiconductor example discussed earlier.

$$\omega_0^2 \approx ag_{th}(\beta + P_d) = 1.23 \times 10^{-2} \quad \therefore \omega_0 \approx 0.111 \qquad \text{(dimensionless)}$$

$$\tau_d = \frac{2}{a(1 + P_d)} = 69.4 \quad \therefore \omega_0\tau_d = 7.7$$

and hence, the sine term in (9.3.22) is much less than the cosine.

Thus, the approximate time history of the photons is given by

$$\Delta p(t) = \frac{\Delta r}{g_{th}}\left[1 - e^{-t/\tau_d}\cos\omega_0 t\right] \tag{9.3.23}$$

which is graphed in Fig. 9.8 using the numerics given previously.

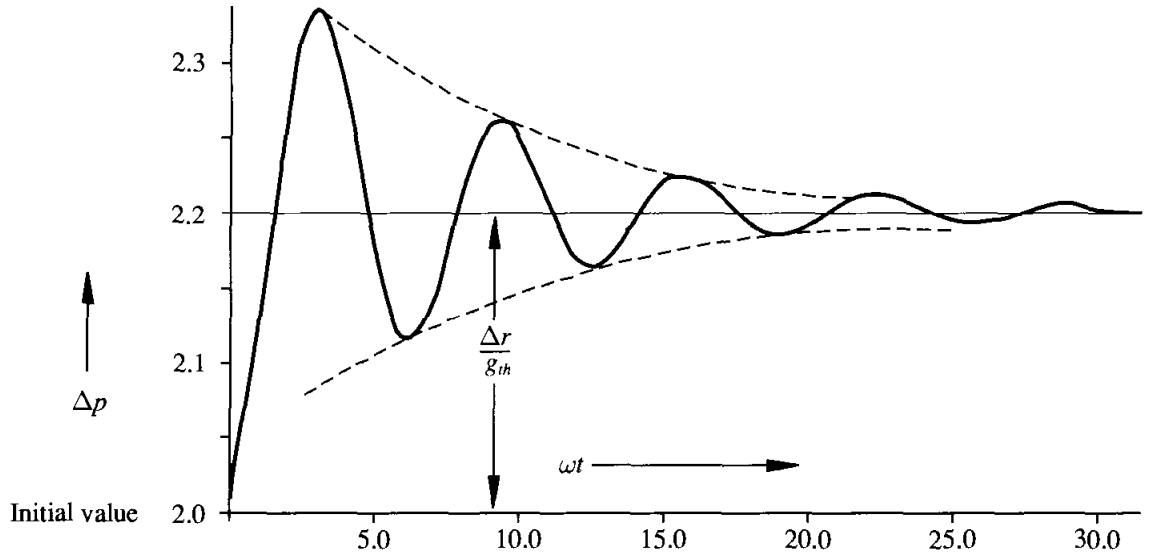


FIGURE 9.8. Damped oscillation while a laser approaches a new steady state. ($\Delta r = 0.2 \times 0.64 = 0.128$)

9.3.6 Case e: Pulsed excitation → gain switching

Assume a laser system, starting from zero excitation and with zero photons, excited by a pump whose amplitude is well above that required to exceed threshold on a steady state basis. The goal is to follow the development of the laser amplitude with time.

One of the first points to be made is that the product of the pump rate and its duration, if it is in the form of a pulse, must be large enough to make the relative photon number grow from its rather insignificant value found in case a to something appreciable and interesting, say $P \sim 0.1$. In this time interval, (9.3.3) and (9.3.4) can be solved by analytic techniques

$$\frac{dg}{dt} = a(R_e - g)$$

$$\frac{dP}{dt} = (S \exp[g] - 1) + \beta g$$

Thus,

$$g = R_e(1 - \exp[-at]) \approx R_e$$

for $t \gg a^{-1}$ (or t' (in sec) $\gg \tau_2$), which indicates that the gain can exceed threshold before the photons have time to build up from spontaneous emission and use the excess inversion. For the photons, assume that g has reached a value in excess of g_{th} but can be treated as if it were a constant.

$$P(t) \approx \delta P_0 \exp\left[(S \exp[g] - 1)t\right]$$

If δP_0 is very small as indicated by case a, then the exponential factor must be quite large to make P comparable to 1. Traditional wisdom (intuition of the early workers) suggested that this factor in the exponent should be $\sim$ 15 to 20 before we can consider the system to be a laser.

If we choose $\delta P_0 = 10^{-7}$, then $\exp[(S\exp[g]-1)t] = 10^6$ or $[(S\exp[g]-1)t] = \ln(10^6) = 13.8$ for P to reach the "interesting" level of 0.1 where the equations become nonlinear.

$$(S\,e^g - 1)\Delta t \approx 13.8$$

If the excitation is merely just above threshold—say 10% above threshold $g = 1.1 \cdot g_{th}$—and g_{th} is small, say ~ 0.1, then an estimation for time to reach oscillation is

$$\Delta t \sim \frac{\ln(10^6)}{\exp\left[(g - g_{th})\right] - 1} \rightarrow \frac{13.8}{(g - g_{th})}$$

where S was expressed as $\exp[-g_{th}]$. For this example, it will take a normalized time of 1380 (the number of round trips) before the laser reaches an amplitude that is beginning to be interesting. If the perimeter of the cavity were 12 inches and $n = 1$, then $\tau_{RT} = 1$ ns and the buildup time is 1.38 μs. Obviously, the pumping must be maintained for a long enough time to permit this evolution. If we "seed" the cavity with a reasonable value of photons—say from another laser—we can speed up this process and shrink the factor of 1380 down to a smaller value by starting with a larger value of δP_0.

The nonlinear and coupled differential equations must now be faced. Case a indicated that nothing interesting happens until the gain exceeds threshold and even then the photons take a considerable time to build up from spontaneous emission to a significant value. Hence the analytic integration approach can be used for the initial part.

$$\frac{dP(t)}{dt} = \left\{S\exp[g(t)] - 1\right\} P(t) = \left\{\exp[(g(t) - g_{th})] - 1\right\} P(t) \tag{9.3.3}$$

$$\frac{dg(t)}{dt} = a\left\{R_e - g(t)[1 + P(t)]\right\} \tag{9.3.4}$$

with $P(0) = 10^{-8}$ or so. It is worthwhile to minimize the number of "free" parameters for the model and also to choose a more convenient time normalization. Let $T = t'/\tau_p$, the time normalized to the photon lifetime of the passive cavity; $\tau_p = (\tau_{RT}/1 - S)$; $S = \exp[-g_{th}]$; and $b = \tau_p/\tau_2$. Our equations then become

$$\frac{dP}{dT} = \frac{\left\{\exp[g(T) - g_{th}] - 1\right\} P(T)}{1 - \exp[-g_{th}]} \tag{9.3.24}$$

$$\frac{dg}{dT} = b[R_e - g(1 + P)] \tag{9.3.25}$$

A numerical integration of the latter two equations is shown in Fig. 9.9 for a value of the pumping parameter $m = R_e/R_{th} = 4.0$ (and thus the CW photon number will be 3), $b = 0.05$ (which indicates that τ_2 is 20 times the photon lifetime), and a line integrated threshold gain of $0.1 = g_{th}$, which specifies the losses in the cavity or alternatively the fraction of the photons surviving a round trip in the passive cavity ($\therefore S = 0.9048$). The numerical integration started when the gain was twice the threshold value with the relative photon number equal to 10^{-6}, an arbitrary but typical choice.

The interesting part of this figure appears for $T > 8$ where the gain is above threshold by about a factor of ~ 2.5 (on its way toward the asymptotic value of 4) and the relative photon number still insignificant compared to 1. At $T > 9$, the photons become larger than

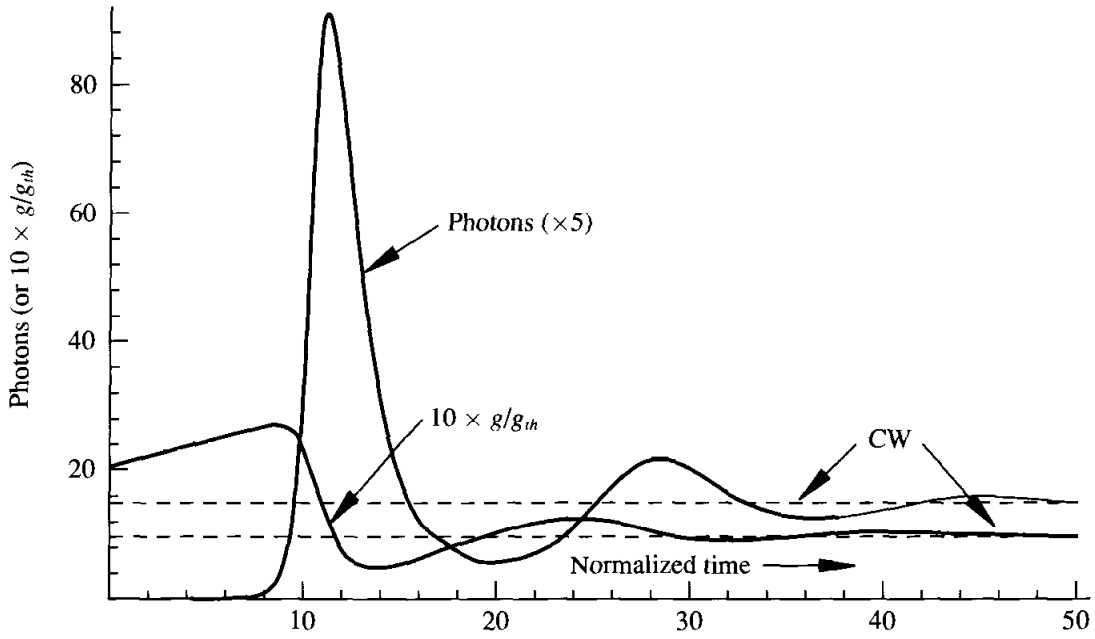


FIGURE 9.9. The time evolution of a "gain switched" pulse. The gain and photon number have been multiplied by 10 and 5, respectively, so as to separate the curves.

1 and begin to depress the population inversion (or gain) and even becomes larger than the CW value by a factor of ~ 6 at $T \sim 12$. (The CW value is shown as the dashed line at 19.3.) With the number of photons in the cavity large, stimulated emission reduces the gain and the photon number reaches a maximum when the gain crosses threshold, shown as the dashed line at 10. There are still photons in the cavity—a large number—and they continue to stimulate the atoms and thus reduce the gain to *below* the threshold and as a consequence the photon number decreases to below the CW value. This allows the gain to re-establish itself and leads to a secondary peak in the photons. Eventually, both the photons and the gain settle down to their CW values. The oscillatory behavior as both settle down is similar to that found in case e, although the numbers chosen for this example are different.

This large initial pulse is sometimes referred to as a "gain switched pulse" and occurs because the gain can build up faster than the photons. Depending upon the choice of the parameters, we can have one or more significant "pulses" with the initial one being rather intense. If we could build up the inversion to a larger value than was found here—say, by making the initial seed of the photons even smaller—we would expect a larger initial pulse. Techniques for accomplishing this and modifications to our model are covered in the next section on "Q switching."

9.4 *Q* SWITCHING, *Q* SPOILING, OR GIANT PULSE LASERS

The last section has indicated a rather complex but interesting interplay between the atomic populations and the photons inside the cavity with one case leading to the generation of short intense pulses of optical energy. There are many uses of such a laser as a few moments of arithmetic will illustrate.

Suppose we had a laser rated at 1 Watt of average power carried in a TEM_{00} beam with $w_0 = 0.5$ cm. The intensity is $\langle I \rangle = 1$ Watt $\div [\pi(0.25)\ \text{cm}^2/2] = 10.18$ Watts/cm^2 and the peak electric field is $E = (2\eta_0 \langle I \rangle)^{1/2} = 87.6$ V/cm, somewhat commonplace if not boring numbers. If, however, this same laser we mode locked (Sec. 9.5) with the output consisting of pulses of duration of 0.1 ps (10^{-13} sec) repeating every 10 ns (10^{-8} sec), then the peak intensity of each pulse is $I_p = \langle I \rangle \cdot$ (time between pulses is 10^{-8} sec) $\div$ (duration of pulse $= 10^{-13}$ sec) $= 1.018$ MW/cm^2 and the peak electric field is 27.7 kV/cm. If this laser were *Q* switched and produced this same average power in the form of a 1 joule, 10 ns pulse every 1 sec, then the peak power is 10^8 watts, the peak intensity is 1.018 GW/cm^2, and the peak electric field is 0.876 MV/cm.

The numerical values for the intensity and electric field become extreme and would become more so if the beam is focused. Techniques for the generations of such pulses are discussed here.

The idea of *Q* switching is a takeoff of some well-known techniques used in low-frequency electronics for the generation of and switching of high power. For instance, consider a radar set operating at 10 to 50 MW of peak radiated power. It is a nontrivial job of modulator design to devise a switch to handle this power level, and it is mind boggling to consider a "portable" power supply capable of delivering this amount of continuous power.

However, the energy contained in a typical radar pulse—say of duration 0.1 μs—is quite ordinary: $W = (10 \times 50 \times 10^6)W \times 0.1 \times 10^{-6}$ sec $= 1$ to 5 joules. The fact that the energy is so minuscule provides a clue to techniques used for its generation. Fig. 9.10 illustrates a simple technique for generating high peak pulsed power.

1. We extract energy from the primary source, the power supply, at a reasonable rate, limited by the charging resistor to be P_{max} (power supply) $= V^2/R$. Call this the *pumping cycle*.
2. The energy is stored by the capacitor, which is prevented from discharging by the open circuit of the switch.
3. Once the switch is closed, the only thing limiting the capacitor discharge current is the impedance of the load. Call this the power cycle.

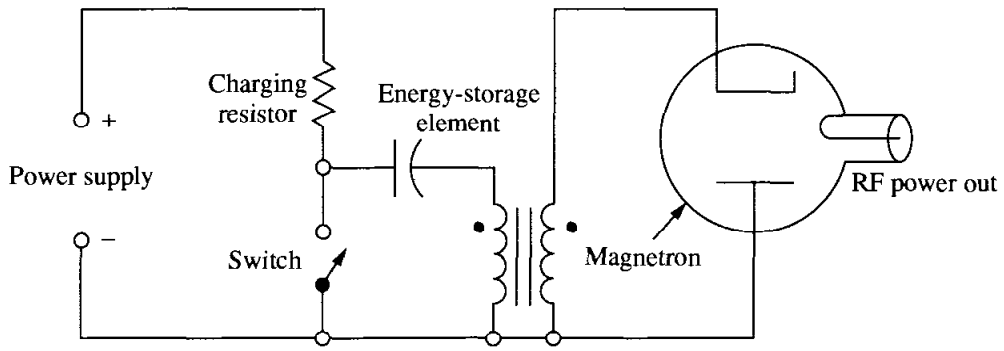


FIGURE 9.10. Elementary radar transmitter. (Replace the magnetron by a spark plug, and it is part of an electronic ignition system.)

It should be clear that the peak power delivered to the load can be many times the peak instantaneous power extracted from the source.

These same ideas, utilized in the "giant pulse" operation of a laser, use the fact that the energy can be stored for future use by creating a population inversion. Obviously, spontaneous emission out of the upper state represents a drain on the stored energy, much as would a leakage path on the capacitor.

However, spontaneous emission causes another difficulty; that is, it is amplified by the population inversion, and if the round trip gain exceeds 1, will build up to a steady state value whose intensity is limited by the rate at which energy can be pumped into the system [see (8.3.21)]. To avoid this drain on the population inversion, the laser is *prevented* from oscillating by making the loss per pass very high while pumping the system. If amplified spontaneous emission can be prevented from saturating the active medium with a single-pass gain length, then considerable energy can be stored in the population difference, $N_2 - N_1$.* This stored energy can be extracted by suddenly *lowering* the loss. Under these circumstances, the gain greatly exceeds the loss and the intensity rapidly builds up from spontaneous noise, reaching a level where further growth is impossible (i.e., when the gain per pass equals the loss per pass).

It is important to remember that *transient* phenomena are being discussed here. This intensity at which further growth is impossible represents the energy stored in the initial population inversion. It is *not* the intensity given by the CW oscillation condition. As we will see, the peak intensity can be many times the CW level.

Let us follow this sequence through with an educated guess as to the behavior before getting buried in the mathematics. As shown in Fig. 9.11, we assume a simple laser geometry with some sort of a closed shutter to spoil the cavity Q and thus prevent oscillation from taking place and using the population inversion. Under such circumstances, we can continually pump energy into the population inversion until some sort of equilibrium is reached between the pump and the spontaneous decay processes of the system. This initial population inversion may be many times that required for CW oscillation in the absence of the shutter.

Let us identify $t = 0$ as that point when the shutter is opened. At that instant we have a system that is far above threshold, and thus the spontaneous emission along the axis of the cavity is greatly amplified and, owing to the feedback provided by the cavity, builds up from its low value to one sufficiently strong to start depleting the population inversion. A few numbers will convince us that this occurs on a very short time scale. For instance, assume that the net single-pass gain is 5 [i.e., $(R_1 R_2)^{1/2} \exp(\gamma_0 l) = 5$]. Then in five round trips (10 single passes), the photon flux would be amplified by $5^{10} \sim 10^7$. As a consequence of this rapid time scale, we neglect any pumping that might occur after $t = 0$.

In view of this large increase in photon flux, we realize that the population inversion will become depleted as the photon number increases. Consequently, we must keep track of the number of excited states as well as the number of photons. (Such a dual bookkeeping situation can lead to a painful headache without careful attention to detail.)

*We tend to discount this stored energy until numbers are quoted. For instance, we can store $\sim$ 50 joules per liter-atmosphere in the population difference of the CO_2 laser transition at 10.6 μm, a figure comparable to the best capacitor available.

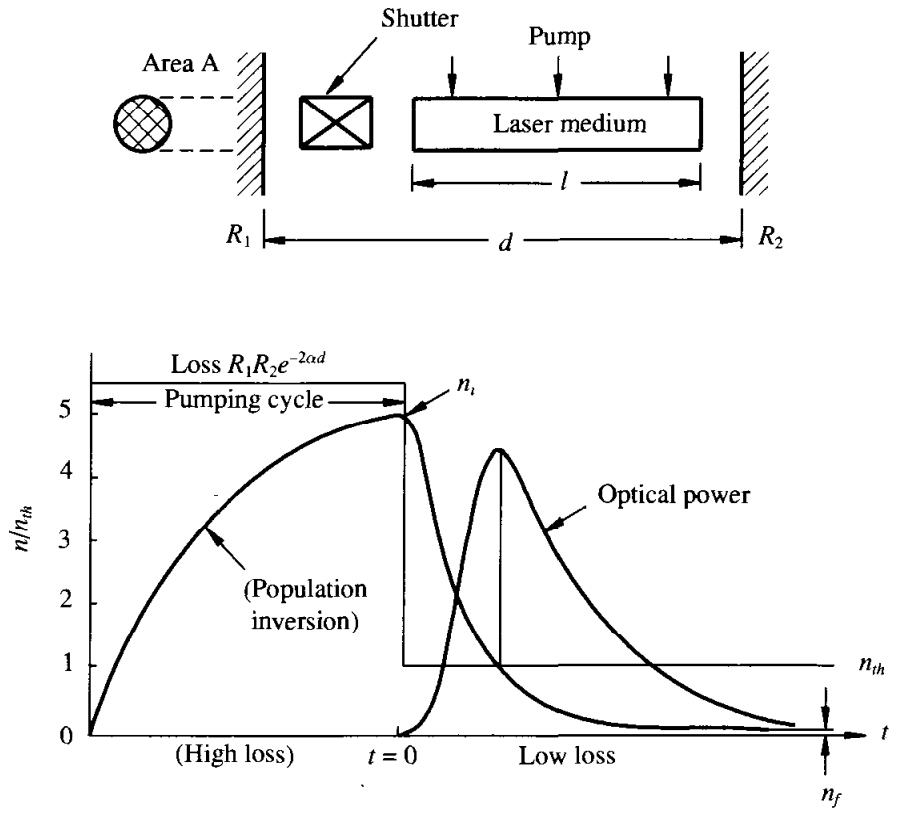


FIGURE 9.11. Guess at the sequence of events during a Q switch.

Let us now consider the task of formulating those ideas into a mathematical format so as to make numerical predictions about the amplitude of the intensity produced by this Q switching operation. The geometry for our analysis is shown in Fig. 9.11 where some of the details indicate that many of the lasers that are candidates for Q switching are solid-state ones and have an index of refraction that is significantly different from 1, and that the electro-optic shutter has still another index. The problem is exceedingly difficult to handle exactly because three spatial dimensions as well as time are involved. We can get rid of at least two of the transverse spatial ones by assigning an area A to them, but we are still left with two dimensions, z and t, and thus we make our first approximation: We treat the cavity as a circuit element and ignore any nonuniformities in the population densities or the photon density with distance inside the simple cavity shown in Fig. 9.12.

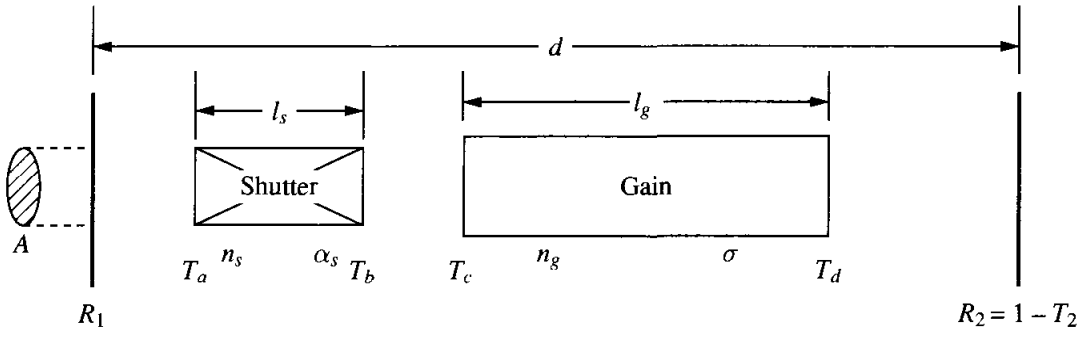


FIGURE 9.12. The geometry of a Q switched laser.

Let us follow the time evolution of the number of photons inside the cavity presuming that there are a few around to initiate the lasing process. In one round trip this number will increase by a factor $\exp[2(N_2 - N_1)\sigma l_g]$ and decrease owing to imperfect window transmission ($\Pi_j T_j$), finite reflectivities ($\Pi_i R_i$) and residual absorption (usually in the shutter) as $\exp -[2\alpha_s l_s]$. The net *change* in N_p during a round trip, the amplified result minus the starting value, divided by τ_{RT} is an excellent approximation to dN_p/dt.

$$\frac{dN_p}{dt} = \left\{ \frac{[(\prod_i R_i)(\prod_j T_j)\, e^{-2\alpha_s l_s}]\, e^{+2(N_2-N_1)\sigma l_g} - 1}{\tau_{RT}} \right\} N_p \tag{9.4.1}$$

There is an obvious need for abbreviations to eliminate the clutter of symbols. We should realize that the factor $[N_2(t) - N_1(t)]\sigma l_g = g(t)$ is the single pass line-integrated gain and must be bigger than a threshold value g_{th} in order for the photons to increase with time. Thus, define

$$[N_2(t) - N_1(t)]\sigma l_g \triangleq g(t) \qquad \text{(the line integrated gain)} \tag{9.4.2a}$$

$$e^{-2g_{th}} \triangleq (\prod_i R_j)(\prod_j T_j)\, e^{-2\alpha_s l_s} \qquad \text{(the threshold value of } g\text{)} \tag{9.4.2b}$$

and recall that the photon lifetime of the *passive cavity* is

$$\tau_p = \frac{\tau_{RT}}{1 - (\Pi_j R_j)(\Pi_j T_j)\, e^{-2\alpha_s l_s}} = \frac{\tau_{RT}}{1 - e^{-2g_{th}}} = \frac{\tau_{RT}}{1 - S} \tag{9.4.2c}$$

where $S = (\Pi_j R_j)(\Pi_j T_j)\, e^{-2\alpha_s l_s}$ (the fraction of photons surviving a round trip).

With these abbreviations, (9.4.1) becomes

$$\frac{dN_p}{dt} = \frac{\left\{e^{2[g(t)-g_{th}]} - 1\right\}}{[1 - e^{-2g_{th}}]} \frac{N_p}{\tau_p} \tag{9.4.3a}$$

The assumption of treating the active cavity as a circuit element forces us to restrict the analysis to cases where the change in N_p per round trip is small and thus the exponents are also small and a Taylor series expansion for $e^x = 1 + x$ is in order.

$$\frac{dN_p}{dt} = \left\{ \frac{1 + 2[g(t) - g_{th}] - 1}{1 - (1 - 2g_{th})} \right\} \frac{N_p}{\tau_p} = \frac{N_p}{\tau_p} \left\{ \frac{g(t)}{g_{th}} - 1 \right\} \tag{9.4.3b}$$

$$\boxed{\frac{dN_p}{dt} = \frac{N_p}{\tau_p} \left\{ \frac{n}{n_{th}} - 1 \right\}} \tag{9.4.3c}$$

where $\tau_p \approx \tau_{RT}/[2g_{th}]$ from (9.4.2c) has been used; $n = (N_2 - N_1)Al_g$ is the total number of inverted atoms in the cavity interacting with an optical mode of cross sectional area A; and n_{th} is total number at threshold.

All forms of (9.4.3) should appeal to our intuition: If the gain or inversion is larger than the threshold value, then the photons increase with time—otherwise, they decrease.

Stimulated emission increases the photon number but also changes the inversion. For every photon added, there is one atom changing its label from 2 to 1, which reduces the inversion by 2 (for equal degeneracies) and thus reduces the gain. Thus we must now shift our attention to the dynamics of the population inversion.

Usually the time scale for the build up and decay of the photons is on the order of a few photon lifetimes, which, in turn, is usually much shorter than the lifetime of state 2 and/or the characteristic time scale for pumping. This disparity in time scales permits us to concentrate only on change in populations caused by stimulated emission and ignore any other cause.

$$\frac{dN_2}{dt} = -\frac{\sigma(I^+ + I^-)}{h\nu}(N_2 - N_1) \tag{9.4.5}$$

$$\frac{dN_1}{dt} = +\frac{\sigma(I^+ + I^-)}{h\nu}(N_2 - N_1) \tag{9.4.6}$$

Subtract (9.4.6) from (9.4.5); multiply both sides by A, the area of the optical mode and l_g, the length of the gain medium; and identify $(N_2 - N_1)\sigma l_g = g(t)$

$$\frac{dn}{dt} = -2\left[(N_2 - N_1)\sigma l_g\right] \cdot \frac{(I^+ + I^-) \cdot A}{h\nu} \cdot \frac{\tau_{RT}}{2} \cdot \frac{2}{\tau_{RT}} \tag{9.4.7a}$$

$$\frac{dn}{dt} = -4g(t) \cdot \frac{N_p}{\tau_{RT}} \tag{9.4.7b}$$

where the number of photons inside the cavity N_p equals [the intensity of both waves $(I^+ + I^-)$] $\times$ [the mirror-to-mirror transit time] $\times$ [effective area of the mode] $\div$ $[h\nu]$.

$$N_p = \left\{ \frac{(I^+ + I^-) \cdot A}{h\nu} \cdot \frac{\tau_{RT}}{2} \right\} \tag{9.4.8}$$

Now use $\tau_{RT} \approx \tau_p(2g_{th})$ from (9.4.2c) and the fact that $g(t)/g_{th} = n/n_{th}$ to obtain

$$\boxed{\frac{dn}{dt} = -2\frac{n}{n_{th}} \cdot \frac{N_p}{\tau_p}} \tag{9.4.9}$$

Equation (9.4.9) could have been stated as a direct consequence of (9.4.3) since it merely states that if the photons increase by 1, then the inversion must decrease by 2 (for equal degeneracies).

> The serious student will consider the case of unequal degeneracies, a situation which modifies the above formulation somewhat. It is an excellent practice in the manipulation of rate equations, but the modification is not a serious change in what follows.

One last bit of manipulation is in order: to define a time scale normalized to the photon lifetime of the passive cavity by $T = t/\tau_p$. Thus our basic equations become

$$\frac{dN_p}{dT} = \left(\frac{n}{n_{th}} - 1\right) N_p \tag{9.4.10a}$$

$$\frac{dn}{dT} = -2\frac{n}{n_{th}} N_p \tag{9.4.10b}$$

Here we run into an obstacle to our progress with the mathematics. These equations are nonlinear and coupled and cannot be solved for N_p and n in terms of elementary functions of time. A computer is needed for this so the issue will be postponed. However, we can find an elementary solution for the photon number N_p in terms of the inversion n. Indeed, if we pay close attention to the physical picture presented in Fig. 9.11 and the physical "walk through" the Q switch pulse of the prior pages, we can obtain a very good answer for the peak of the power pulse, its energy, and a reasonable estimation for its FWHM without resorting to anything more complicated than a calculator. We start by dividing (9.4.10a) by (9.4.10b), which eliminates time from the equation.

$$\frac{dN_p}{dn} = \frac{1}{2}\left(\frac{n_{th}}{n} - 1\right) \tag{9.4.11}$$

Now we multiply both sides by dn and integrate the left-hand side from the initial value of the photon number $N_p(i)$ (which is negligible compared to what it will be) to the photon number $N_p(max)$ at the peak of the power pulse, which is the first physical parameter to be determined. But how are we to know when we have reached the peak of the pulse? This is where the walk-through associated with Fig. 9.11 is so important: Note that photon number reaches a peak when the inversion crosses the threshold value. Of course, the differential equation (9.4.10) states the same thing: $dN_p/dt = 0$ when $n = n_{th}$. Thus the limits of integration of the right-hand side are from the initial value of the inversion, n_i, to the threshold value n_{th}.

$$\int_{N_p(i)-0}^{N_p(\max)} dN_p = \frac{1}{2}\int_{n_i}^{n_{th}} \left(\frac{n_{th}}{n} - 1\right) dn$$

$$N_p\ (\text{max.}) = \frac{n_i - n_{th}}{2} - \frac{n_{th}}{2} \ln\left(\frac{n_i}{n_{th}}\right) \tag{9.4.12}$$

Now $h\nu \cdot N_p$ represents the total optical energy stored inside the cavity, but that energy is being lost in the Q switching element ($\exp[-2\alpha_s l_s]$), imperfect transmission coefficients at the air-element interfaces ($\prod_j T_j < 1$), and imperfect mirror reflectivities ($\prod_j R_j < 1$). (The quantities in the parentheses represent the fraction lost per round-trip for each process.) Only that loss representing the coupling by (or transmission through) one of the mirrors, $1 - R_2$, represents the useful loss to the outside world. The output power is equal to the stored energy multiplied by the fraction lost by coupling per round trip divided by the time for that trip.

$$P(t) = \frac{h\nu N_p(t) \cdot [\text{fraction lost through coupling (per round trip)}]}{[\text{time for a round trip}]}$$

$$= h\nu N_p(t) \cdot \frac{[\text{coupling loss per round trip}]}{[\text{total loss per round trip}]} \cdot \frac{[\text{total loss per round trip}]}{[\text{time for a round trip}]}$$

The first ratio is the coupling efficiency η_{cpl} and is the optical electromagnetic problem mentioned in Sec. 9.1.1. We design for, but never quite achieve, an absence of useless

conversion of photons into heat, and thus $\eta_{\text{cpl}} < 1$. The second ratio is the inverse of the photon lifetime for the low loss cavity being addressed here. Hence the peak power can be found from this definition and from (9.4.12).

$$P(\max) = \eta_{\text{cpl}} \frac{h\nu N_p(\max)}{\tau_p} \tag{9.4.13a}$$

where the coupling efficiency can be expressed as the ratio of the output mirror transmission coefficient to the total loss in one round trip

$$\eta_{\text{cpl}} = \frac{[T_2 = 1 - R_2]}{[1 - S]} = \frac{\{1 - \exp[-2 < \alpha_{\text{ext.}} > d]\}}{\{1 - \exp[-2\langle\alpha_T\rangle d]\}} \approx \frac{\langle\alpha_{\text{ext.}}\rangle}{\langle\alpha_T\rangle}$$

where the "spatially averaged" attenuation coefficients are *defined* by

$$\exp[-2\langle\alpha_{\text{ext.}}\rangle d] \triangleq R_2$$

and

$$\exp[-2\langle\alpha_T\rangle d] = (\prod_j T_j)(\prod_j R_i) \exp[-2\alpha_s l_s] \triangleq S$$

Thus, the maximum output power can be expressed in a compact fashion by

$$P(\max) = \eta_{\text{cpl}} \frac{h\nu N_p(\max)}{\tau_p} \tag{9.4.13b}$$

Let us return to (9.4.11) and integrate over the complete time interval of the pulse so as to estimate the fraction of the inversion that is converted to photons. The photon number N_p starts at a very low value, which we again approximate as zero, reaches a maximum, and returns to this low value at the final stage. Thus the limits of integration on N_p are from zero to zero. However, the limits on n corresponding to those "different" values of zero are n_i and n_f where the latter one is the final inversion number.

$$\int_0^0 dN_p = \frac{1}{2} \int_{n_i}^{n_f} \left(\frac{n_{th}}{n} - 1\right) dn$$

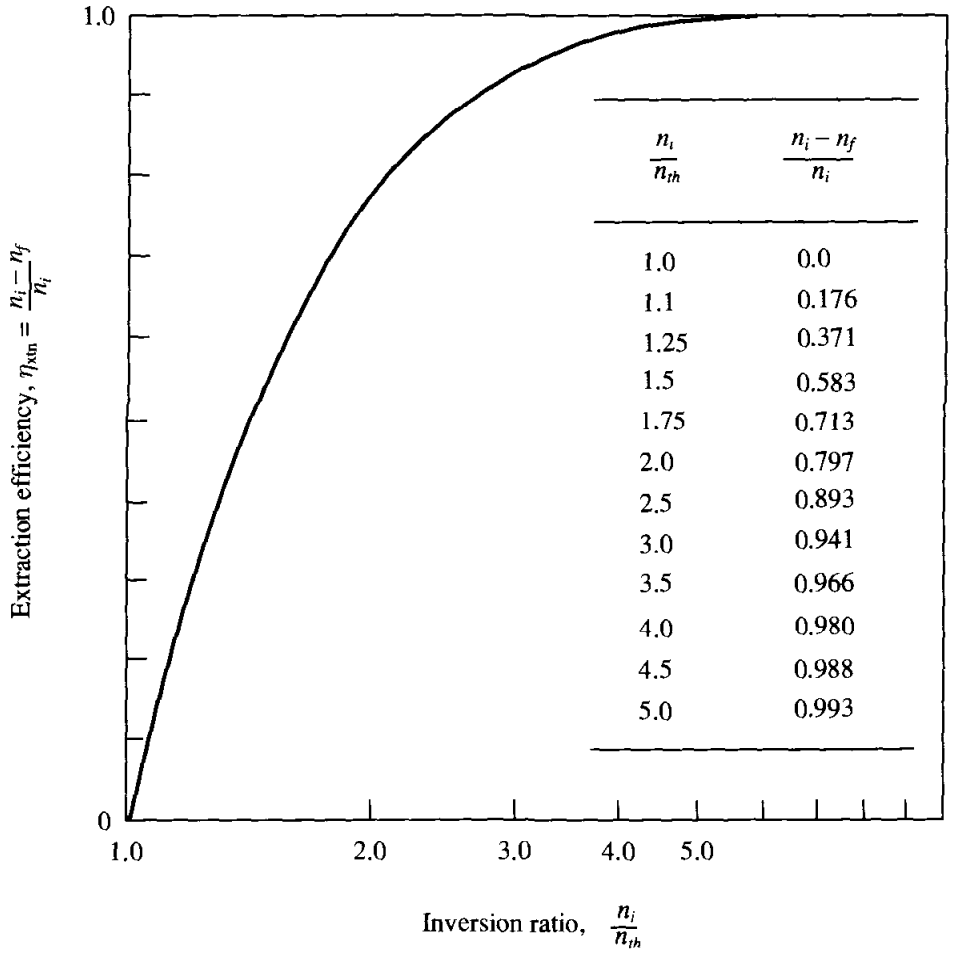
or

$$\frac{n_i - n_f}{2} - \frac{n_{th}}{2} \ln\left(\frac{n_i}{n_f}\right) = 0$$

Rearranging we get

$$\frac{n_f}{n_t} = \exp\left[-\left(\frac{n_i - n_f}{n_{th}}\right)\right] \tag{9.4.14}$$

This is an equation that determines n_f implicitly in terms of the initial value n_i and the threshold value n_{th}. Although it does not have a simple analytic solution, it is fairly simple to solve numerically (or graphically) and the result is in Fig. 9.13 in a special manner. The horizontal axis is chosen to satisfy our intuition that the development of the Q switched pulse is strongly influenced by the amount by which the initial inversion exceeds the threshold



$\frac{n_i}{n_{th}}$	$\frac{n_i - n_f}{n_i}$
1.0	0.0
1.1	0.176
1.25	0.371
1.5	0.583
1.75	0.713
2.0	0.797
2.5	0.893
3.0	0.941
3.5	0.966
4.0	0.980
4.5	0.988
5.0	0.993

FIGURE 9.13. The energy extraction efficiency for a Q switched pulse.

value at the beginning. The choice of the vertical or y axis is motivated by the practical question: How much of the initial inversion is converted to photons? The fraction of the initial inversion converted to photons is

$$\eta_{xtn} = \frac{n_i - n_f}{n_i} \tag{9.4.15}$$

Knowing this fraction, the total energy generated in the form of photons is given by this efficiency times the maximum available. Let N_p be the total number of photons that can be created at the expense of the population inversion.

$$n_i = [N_{2i} - N_{1i}]Al_g \tag{9.4.16a}$$

$$n_f = [N_{2f} - N_{1f}]Al_g \tag{9.4.16b}$$

but

$$[N_{2i} - N_{2f}]Al_g = N_p = [N_{1f} - N_{1i}]Al_g$$

If we subtract (9.4.16b) from (9.4.16a), solve for N_p, and multiply by $h\nu$, we obtain the

total energy converted into photons:

$$W_p = h\nu \frac{n_i - n_f}{2} = \eta_{xtn} \frac{h\nu n_i}{2} \tag{9.4.16c}$$

The fraction of this energy in the output is the product of (9.4.16c) and the ratio of the coupling loss to the total loss.

$$W_{\text{out}} = \eta_{\text{cpl}} \eta_{xtn} \frac{n_i h\nu}{2} \tag{9.4.16d}$$

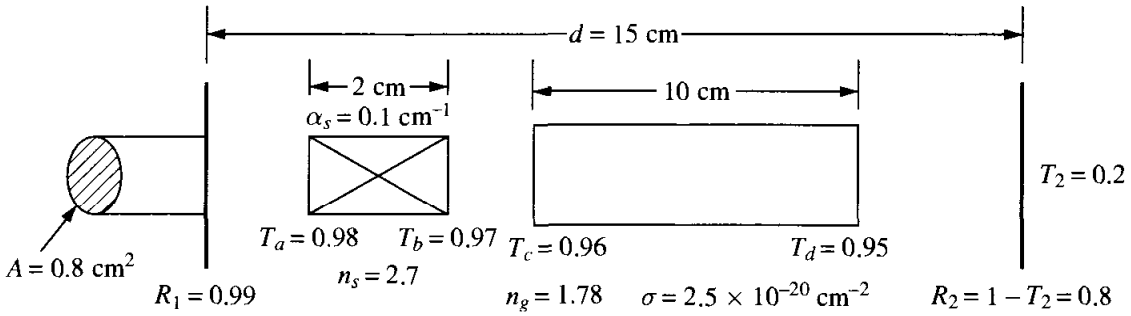
The one remaining issue to be addressed is to estimate the pulse width. For an exact answer, one needs to integrate (9.4.10a) and (9.4.10b) with respect to time, but a very reasonable estimate can be obtained by dividing the energy, (9.4.16d), by the maximum power given by (9.4.13b):

$$\Delta t \approx W_{\text{out}}/P_0(\max) \tag{9.4.17}$$

A Numerical Example

Let us apply these ideas to the laser system shown below where we presume a ruby laser ($\lambda = 6943$ Å) to be pumped to four times threshold. For simplicity, we assume equal degeneracy of the lasing states even though that is not true for ruby. (See Chapter 10 for a discussion about the details of the ruby laser.) We also assume a residual attenuation of 0.1 cm^{-1} by the switch even when it is in its high transmission state.

We start by computing the threshold gain coefficient and from that quantity the threshold inversion density or number of inverted atoms. We start with Eq. (0.1) of this book; that is, round trip gain must be greater than 1 for an oscillator and equal 1 for threshold.



$$R_1 R_2 (T_a T_b T_c T_d)^2 \, e^{-2\alpha_s l_s} \, e^{2\gamma_{th} l_g} = 1$$

or

$$\gamma_{th} = \underbrace{\frac{1}{2l_g} \ln\left[\frac{1}{R_1 (T_a T_b T_c T_d)^2}\right]}_{\text{internal cavity losses}} + \underbrace{\alpha_s \frac{l_c}{l_s}}_{\text{switch loss}} + \underbrace{\frac{1}{2l_g} \ln \frac{1}{R_2}}_{\text{external coupling loss}}$$

$$\gamma_{th} = 1.48^{-2}\ \text{cm}^{-1} + 2^{-2}\ \text{cm}^{-1} + 1.12^{-2}\ \text{cm}^{-1}$$

$$= 4.59^{-2}\ \text{cm}^{-1}$$

Thus the threshold inversion is found to be

$$(N_2 - N_1)_{th} = \frac{\gamma_{th}}{\sigma} = 1.83 \times 10^{18}\ cm^{-3}$$

The total number of inverted atoms in the cavity at threshold is

$$n_{th} = (N_2 - N_1)_{th} \cdot A \cdot l_g = 1.47 \times 10^{19} \text{ atoms}$$

and because of the specification of being pumped to four times threshold, we also know the initial inversion.

$$n_i = 4n_{th} = 5.87 \times 10^{19} \text{ atoms}$$

The other parameter of interest is the photon lifetime of the passive cavity:

$$\tau_{RT} = \frac{2(l_{\text{air}} + n_s l_s + n_g l_g)}{c} = 1.75 \text{ ns}$$

$$\frac{1}{\tau_p} = \frac{1 - R_1 R_2 (T_a T_b T_c T_d)^2 e^{-2\alpha_s l_s}}{\tau_{RT}} \qquad \tau_p = 2.91 \text{ ns}$$

Thus we first compute the maximum photon number from (9.4.12):

$$N_p(\max) = \frac{n_i - n_{th}}{2} - \frac{n_{th}}{2} \ln\left(\frac{n_i}{n_{th}}\right) = 1.18 \times 10^{19} \text{ photons}$$

Then we compute the output power at the maximum of the pulse by (9.4.13b):

$$P_0(\max) = \eta_{\text{cpl}} \cdot h\nu \cdot \frac{N_p}{\tau_p} = 282 \text{ MW}$$

which is a rather significant power. The output energy in the Q switched pulse can be found from (9.4.16d) and from Fig. 9.13, which shows that $\eta_{xtn} = 0.98$:

$$W_{\text{out}} = \eta_{\text{cpl}} \cdot \eta_{xtn} \frac{n_i h\nu}{2} = 1.98 \text{ joules}$$

Now we can estimate the pulse width by (9.4.17):

$$\Delta t \approx \frac{W_{\text{out}}}{P_0(\max)} = 7.1 \text{ ns}$$

Fig. 9.14 shows the numerical integration of the basic equations and for $n_i/n_{th} = 4$, the normalized time for the FWHM of the pulse is $\sim 3.8 - 1.6 = 2.2$ units of photon lifetime or 6.4 ns.

The basic equations of Q switching, (9.4.10a) and (9.4.10b), can be integrated by numerical techniques, given that there are a few photons within the cavity mode to start the stimulated emission process. An example is shown in Fig. 9.14. In this calculation, the initial photon number (divided by the threshold inversion) was taken to be 10^{-3} of the inversion ratio n_i/n_{th}. Although the factor 10^{-3} is arbitrary, the form follows from considerations discussed in Sec. 8.8.

Note that as the inversion ratio is increased, the time required to develop a more intense pulse decreases. This is due to two causes: (1) with a higher inversion the gain is higher and thus the rate of growth is faster and (2) a higher inversion ratio implies more atoms in the upper state, which will contribute more spontaneous power to start the oscillation.

The solid curves were all computed by using the assumption that $N_p/n_{th} = 10^{-3} \times (n_i/n_{th})$, whereas the dashed curve used a factor of 10 less for the initial starting point. It is obvious, then, that the presence or absence of the initial photon number corresponds

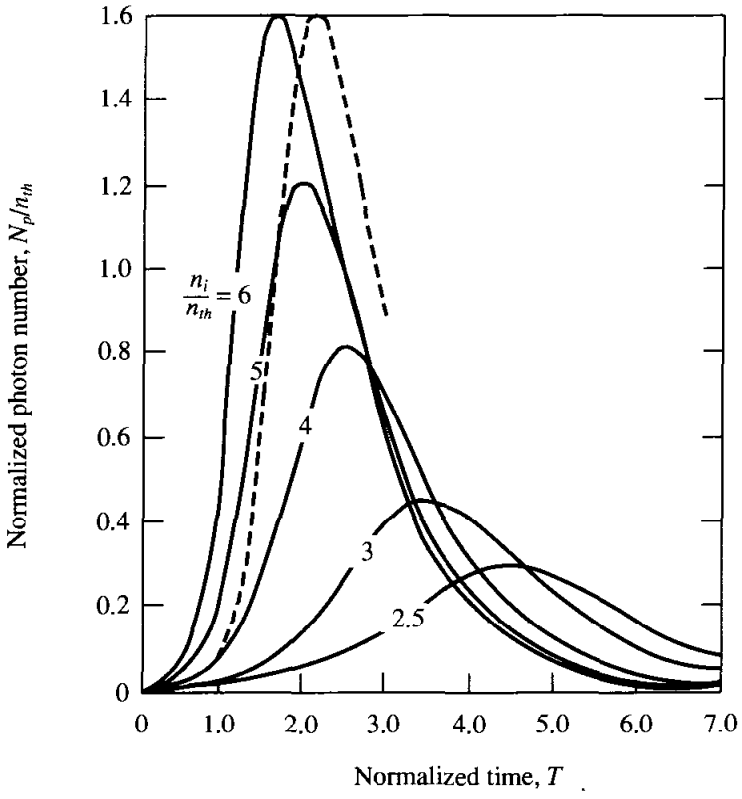


FIGURE 9.14. Time evolution of Q switched pulse for different inversions. The initial photon number was taken to be 10^{-3} times the inversion ratio for solid curves, whereas a factor of 10 less was used for the dashed curve.

to a shift in the time axis. We can readily estimate this time shift by the following line of reasoning.

If the initial photon number is decreased by a factor of 10, the medium must use a time interval ΔT to amplify the smaller number back to its original value. At these very small photon numbers, we can neglect depletion of the inversion and thus N_p grows according to

$$N_p(\Delta T) = N_p(0) \exp\left[+\left(\frac{n_i}{n_{th}} - 1\right)\Delta T\right] \tag{9.4.18}$$

Thus, for $n_i/n_{th} = 6$, the normalized time interval ΔT to compensate for the arbitrary reduction of a factor of 10 in the initial photon number is 0.461, precisely the shift shown in Fig. 9.14.

We cannot increase the initial inversion too much and obtain arbitrarily high peak powers. This is due to two causes:

1. As the initial inversion is increased, the spontaneous emission rate increases proportionately, which is then amplified by the remaining section of the gain medium. Since the single-pass gain is very high because of the high n_i, the amplified spontaneous emission may be sufficient to saturate the gain section, thereby limiting n_i (and also wasting energy). Indeed, it is this phenomenon that *limits* the amount of energy that can be stored in the population inversion in any laser.
2. The amplified spontaneous emission may be sufficient to damage the switch or at least change its characteristics. This last can be used to advantage when the switch

is composed of other atoms that absorb at the laser frequency. Then the spontaneous emission "bleaches" the switch by saturating the absorbing cell (i.e., equalizes the population).

At the very least, the switch must have enough loss or hold-off capability so that the round trip gain is less than 1. For a given switch this limits the initial inversion.

A variety of methods have been used for Q switching, ranging from the brute force technique of spinning one of the mirrors to sophisticated means using electro-optic switches. We can also use a bleachable absorber that presents a very high attenuation to small intensities but much less as the intensity grows.

9.5 MODE LOCKING

9.5.1 Preliminary considerations

In the previous sections, the generation of optical pulses whose repetition rate is determined by control elements external to the laser cavity was discussed. Such pulses can be rather intense, as the previous section has illustrated, but the repetition rate is usually limited to $\sim$ Hz or at most kHz rates.

In Sec. 9.2, we demanded that a CW amplitude reproduce itself after a round trip to obtain a self-consistent description of a laser. However, the requirement of a CW spatial distribution is not necessary—it is just the simplest case. We could also demand a self-consistent self-reproducible "pulse" or a "mode-locked" solution, whose pulses, by necessity, must repeat every τ_{RT} at any point in space. Hence, the repetition rate would be $1/\tau_{RT} \sim$ 100s of MHz to GHz depending upon the size of the cavity. Typical pulse widths range from a relatively long nanosecond (10^{-9} sec) to tens of femtoseconds (10^{-15} sec). Figure 9.15 illustrates the time history and the spectral distribution of such repetitive optical pulses. The relative intensity crossing a plane in space is sketched as a function of time in Fig. 9.15(a) and is intended to emphasize three points common to mode-locked lasers.*

1. The pulse-to-pulse spacing is τ_{RT}, and the repetition rate is $1/\tau_{RT}$.
2. The FWHM of the pulse, Δt_p, is usually very short compared to τ_{RT}. Thus, measuring the time history accurately is a significant experimental task.†
3. Usually the laser produces more or less the same time averaged power in either the CW or mode-locked state. Hence the peak power is roughly $\tau_{RT} \div \Delta t_p$ multiplied by the CW value. Thus the peak power can be quite significant even though the average power is boring.

The fact that the laser power is repetitive in the time domain implies (by virtue of the Fourier series) that the spectral distribution is as sketched in Fig. 9.15(b). The central fre-

*If we multiply by c, this is also the photon distribution in the space of an unfolded cavity.

†There is the danger of interpreting Δt_p as the photon lifetime τ_p. The two are different, and usually the distinction is clear from its use although much of the open literature uses the same letter symbol τ_p for both.

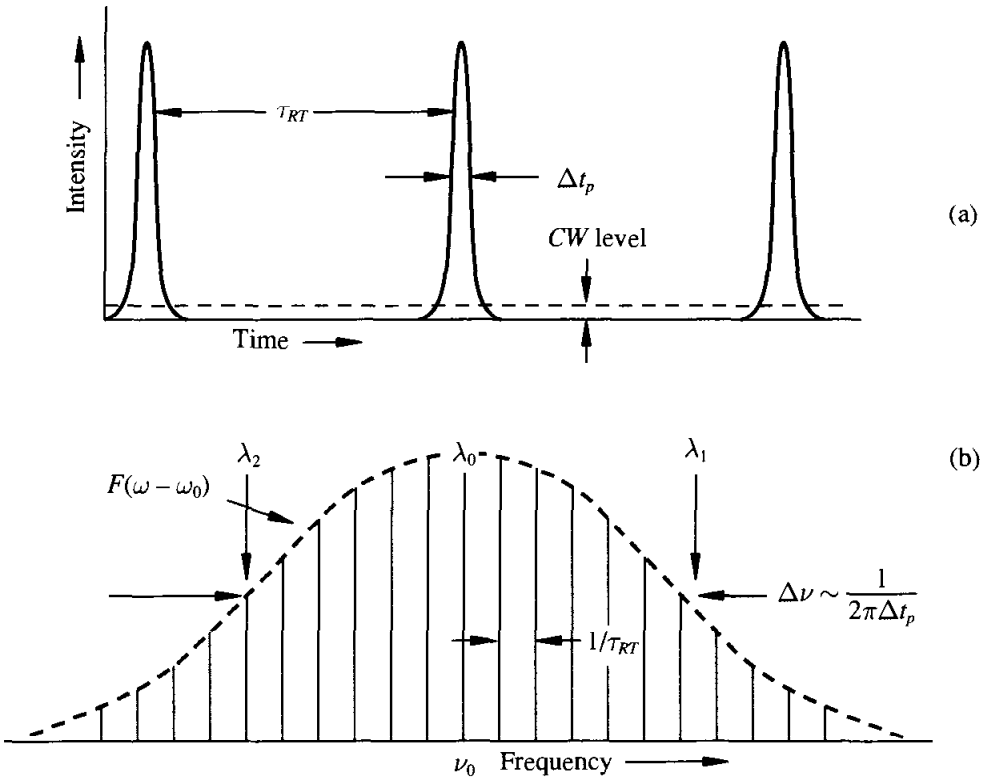


FIGURE 9.15. The (a) time and (b) frequency domain representation of a mode-locked laser.

quency/wavelength at ν_0 is surrounded by sidebands or modes spaced at $1/\tau_{RT} (= c/2d =$ FSR for a simple laser). This spacing between the sidebands or modes in frequency units (hundreds of MHz) is very small compared to the value of ν_0 (hundreds of THz), and probably could not be resolved with simple optical spectrometers, but there must be a large enough number of them to represent a very short pulse. Hence the spectral width occupied by these sidebands becomes significant compared to ν_0 if Δt_p is small.

The spectral width occupied by these modes can be quite significant. For instance, suppose $\lambda_0 = 6000$ Å and thus $\nu_0 = 5 \times 10^{14}$ Hz. If $\Delta t_p = 6$ femtoseconds $= 6 \times 10^{-15}$ sec, then $\Delta\nu$(FWHM) $\approx 1/2\pi\Delta t_p = 2.65 \times 10^{13}$ Hz. Thus, the bandwidth $\Delta\nu$ (or $\Delta\lambda$) over which the pulse interacts with the active atoms is a significant fraction ($\sim 10\%$) of the central wavelength.

If we consider only the wavelengths occupied by the FWHM, then $\lambda_1 = 5698$ Å and $\lambda_2 = 6336$ Å or $\Delta\lambda_{1/2} = 638$ Å. Significant mode amplitudes exist out to two to three times the FWHM value, and thus a wavelength band of ~ 500 Å could be required. The FWHM of such a short pulse occupies a distance of only 1.8 μm in the direction of propagation and contains only three optical wavelengths ($1.8\mu\text{m} \div 0.6\mu\text{m} = 3$). Hence, the entire pulse contains roughly six optical cycles.

Whether these modes are generated by the mode-locking process or occur naturally, as in an inhomogeneously broadened laser, will be addressed later. Historically, the latter viewpoint came first. For now we must merely recognize that they must exist or Fourier would turn over in his grave!

9.5.2 Mode locking in an inhomogeneous broadened laser

Consider the simplest case of an inhomogeneous laser line supporting multilongitudinal mode oscillation such as that discussed in Sec. 8.4. A general expression for the electric field at a particular point in space is

$$e(t) = \sum_{-(N-1)/2}^{+(N-1)/2} E_n(t) \exp\left[j(\omega_0 + n\omega_c)t + \phi_n(t)\right] \tag{9.5.1}$$

where the central (or largest) mode is presumed to be at ω_0 and there are $(N-1)/2$ sidebands or modes uniformly spaced at $\omega_c = 2\pi(1/\tau_{RT})$ on either side of ω_0 and thus there are N modes in all. For the general case, each mode amplitude E_n and phase ϕ_n could be (and usually is) changing on a time scale that is slow compared to τ_{RT} in a completely uncorrelated fashion. Usually, the *wandering* of the phases is the most serious and is the major limitation to the coherence length of such a laser. If these phases did not vary with time, but were randomly distributed over 2π, then the total intensity from all modes $e(t) \cdot e^*(t)/2\eta_0$ might have some repetitive modulation but it would not appear as in Fig. 9.15(a). If there were a large number of modes, but with phases randomly distributed over 2π, the amplitude variation would be very small.

One viewpoint of mode locking is that each phase is forced to a common value that might as well be set equal to zero since any constant value amounts to a time shift. To appreciate the significance of such a minor matter of "teaching" each mode to conform to our phase reference, consider Fig. 9.16 which shows the addition of the phasors of a seven mode laser with relative amplitudes $1/8 : 1/4 : 1/2 : 1 : 1/2 : 1/4 : 1/8$ as a function of time, with all phase terms $\phi_n = 0$. Thus the term $jn\omega_c t$ acts as a phase shift of the nth mode with respect to that at line center. At $\omega_c t = 0$, all of the phasors line up, and the total electric field amounts to 2.75 with a relative peak intensity of 7.56. At $\omega_c t = \pi$, every other mode is in opposite direction, and the relative electric field is only 0.25 and the relative minimum intensity is 0.0625. The average CW power is the sum of the squares of the relative electric fields or $2 \cdot (1/8)^2 + 2 \cdot (1/4)^2 + 2(1/2)^2 + 1 = 1.656$.

Let us advance (slowly) in the level of sophistication by considering an N mode laser with each mode of equal amplitude. This is very unrealistic but is very simple to analyze and yields results that are typical of a mode-locked laser. Equation (9.5.1) becomes

$$\frac{e(t)}{E_0} = \sum_{-(N-1)/2}^{+(N-1)/2} e^{j\omega_0 t}\, e^{jn\omega_c t} = e^{j\omega_0 t} \sum x^n \tag{9.5.2}$$

where $x = e^{j\omega_c t}$. This series can be summed to obtain*

$$e(t) = E_0\, e^{j\omega_0 t} \left\{ \frac{\sin N\omega_c t/2}{\sin \omega_c t/2} \right\} \tag{9.5.3a}$$

*Folklore has it that Gauss did this type of sum at age 6.

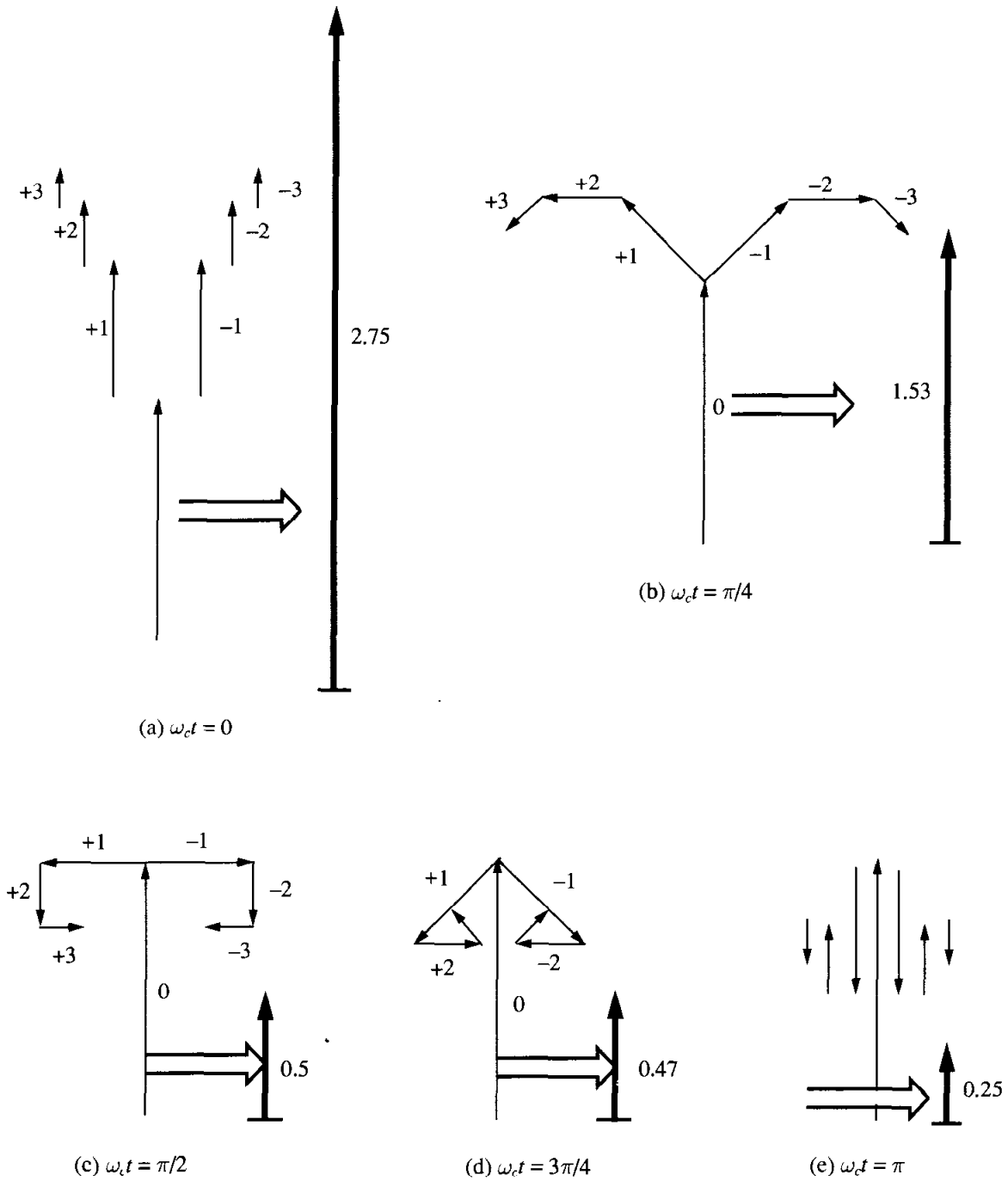


FIGURE 9.16. Phasor addition of the fields of a mode-locked laser. The mode amplitudes were chosen according to a proportional relation: 1 : 0.5 : 0.25 : 0.125.

The intensity of the laser is given by

$$I(t) = \frac{e(t)\, e^*(t)}{2\eta_0} = \frac{E_0^2}{2\eta_0}\left[\frac{\sin(N\omega_c t/2)}{\sin(\omega_c t/2)}\right]^2 \tag{9.5.3b}$$

When $\omega_c t/2$ is an integral multiple of π, the intensity is N^2 times the intensity per mode. But, of course, we have assumed N equal amplitude modes. Hence the average power of the

laser is N times the power per mode. Consequently, we obtain the (approximate) relation that the peak power is N times the average.

$$P_{\text{peak}} = N \times P_{\text{ave}} \tag{9.5.3c}$$

The numerator of (9.5.3b) is a rapidly oscillating function, going to zero when $\omega_c t/2 = \pi/N$ while the denominator slowly increases to 1 at $\omega_c t/2 = \pi/2$. Hence, the instantaneous power would appear as shown in the dashed curve of Fig. 9.17.

Rather than belabor this elementary case, the pulse width is *estimated* by using

$$P(\text{peak})\Delta t_p \approx \langle P_{\text{ave}} \rangle \tau_{RT}$$

and using (9.5.3c) for the average power. This leads to

$$\Delta t_p \approx \frac{\tau_{RT}}{N} \tag{9.5.3d}$$

where N is the number of modes *locked* to a common phase or time origin. But the number N of oscillating modes (which are locked together) is approximately the line width $\Delta\nu$ divided by the cavity mode spacing $c/2d$:

$$N \sim \Delta\nu/(c/2d) = \Delta\nu\tau_{RT}$$

or

$$\therefore \Delta t_p \sim \frac{\tau_{RT}}{N} \sim \frac{1}{\Delta\nu} \tag{9.5.3e}$$

This can be a very short pulse. Even the "narrow" line width of the He:Ne laser, $\Delta\nu = 1500$ MHz, would produce a pulse of ~ 0.6 ns. Such a pulse would occupy only 7 inches of free space. The capabilities of solid-state lasers, such as ruby and Nd-glass, stagger the imagination. Consider the Nd-glass system with $\Delta\nu = 3 \times 10^{12}$ Hz($\Delta\lambda \sim 100$ Å) around

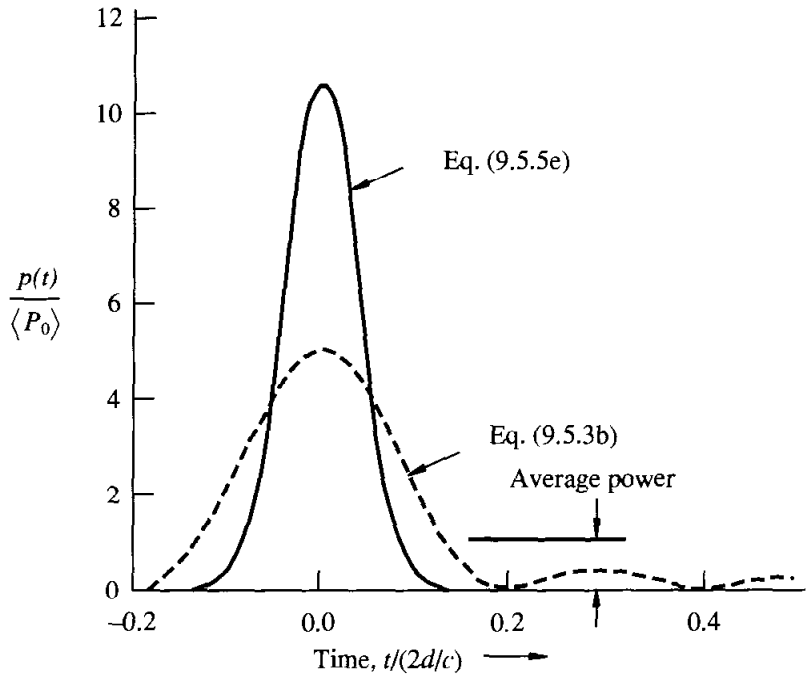


FIGURE 9.17. Comparison of the two theories of the mode-locked pulse. The two expressions were evaluated for the same average power and for $\Delta\omega/\omega_c = 5 = N$.

$\lambda = 1.06\ \mu$m. This indicates a capability of 0.3 ps with a peak power 10^4 times the average power. The pulse would only occupy 0.3 mm of space. Thus there are an enormous number of photons packed into a very small portion of space. We can readily appreciate the fact that the measurement of these ultrashort pulses is a difficult problem.

A case that illustrates the formal mathematical procedure for obtaining the time history of the mode-locked pulses from the spectral distribution and which has some practical implications is that of a Gaussian distribution of mode amplitudes:

Let the *envelope* of this intensity of the modes in Fig. 9.15 be a Gaussian such that the nth mode intensity is related to the one at line center I_0 by

$$I_n = I_0 \exp\left[-4(\ln 2)\left(\frac{n\omega_c}{\Delta\omega}\right)^2\right]$$

and hence the amplitude of the electric field of the nth mode is given by

$$E_n = E_0\left\{\exp\left[-2(\ln 2)\left(\frac{n\omega_c}{\Delta\omega}\right)^2\right]\right\}$$

where $E_n^2/2\eta_0 = I_n$. Thus (9.5.1) becomes

$$e(t) = E_0\, e^{j\omega_0 t} \sum_{n=-(N-1)/2}^{+(N-1)/2} \left\{\exp\left[-2(\ln 2)\left(\frac{n\omega_c}{\Delta\omega}\right)^2\right]\right\} \cdot (1) \cdot e^{jn\omega_c t} \tag{9.5.4}$$

We should not jump to the conclusion that $\Delta\omega$ is the Doppler width; it may be so but it may also be just a convenient scaling parameter with the Gaussian envelope being a reasonable approximation to any bell-shaped envelope.

We can sum (9.5.4) by converting it into the form of a Fourier integral and then either use the extensively tabulated transform pairs or integrate it directly. Probably the simplest approach is to modify (9.5.4) by allowing N to go to ∞ and then converting the sum to an integral by "smearing" each of the modes into a continuum of sinusoids over the interval ω_c between each mode. This last step involves the following logic. Let

$$n\omega_c = x$$

Then

$$(n + 1/2)\omega_c - (n - 1/2)\omega_c = \Delta x \rightarrow dx = \omega_c$$

$$\therefore dx/\omega_c = 1$$

which replaces the (1) in (9.5.4) and the use of dx converts the summation to an integral.

$$e(t) = E_0\, e^{j\omega_0 t}\left\{\frac{1}{\omega_c}\int_{-\infty}^{+\infty}\left\{\exp\left[-\frac{2(\ln 2)x^2}{(\Delta\omega)^2}\right]\right\}\exp[+jxt]\,dx\right\} \tag{9.5.5a}$$

$$= E_0\, e^{j\omega_0 t}\left\{\frac{2}{\omega_c}\int_{0}^{+\infty}\left\{\exp\left[-\frac{2(\ln 2)x^2}{(\Delta\omega)^2}\right]\right\}\cos xt\,dx\right\} \tag{9.5.5b}$$

This is precisely the format of Fourier transform pairs.*

$$\mathcal{F}\{f(x)\} = F(\alpha) = \int_{-\infty}^{+\infty} f(x)\, e^{-j\alpha x}\, dx$$

$$\mathcal{F}^{-1}\{F(\alpha)\} = f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} f(\alpha)\, e^{+j\alpha x}\, dx$$

and if the functions are even or odd, the Fourier cosine (and sine) transform pairs are used.

$$F_c(\alpha) = \mathcal{F}_c\{f(x)\} = \int_0^{\infty} f(x)\cos\alpha x\, dx$$

$$f(x) = \mathcal{F}_c^{-1}\{F_c(\alpha)\} = \frac{2}{\pi}\int_0^{\infty} F_c(\alpha)\cos\alpha x\, dx$$

The integral in (9.5.5b) can be evaluated by using the Fourier cosine transform pair [(formula (33.41) of *Mathematical Handbook*, p. 178)] which list the following pairs† If

$$f(x) = e^{-bx^2}$$

then

$$F_c(\alpha) = \frac{1}{2}\left(\frac{\pi}{b}\right)^{1/2} e^{-\alpha^2/4b}$$

where x of the transform table is x of (9.5.5b) $= [2(\ln 2)/\Delta\omega^2]$ and $\alpha = t$. By using these identifications, the electric field becomes

$$e(t) = E_0\, e^{j\omega_0 t}\left\{\left(\frac{\pi}{2\ln 2}\right)^{1/2}\left(\frac{\Delta\omega}{\omega_c}\right)\cdot\exp\left[-\left(\frac{\Delta\omega t}{2(\ln 2)^{1/2}}\right)^2\right]\right\} \tag{9.5.5c}$$

The function is no longer periodic anymore because the series has been replaced by an integral. But we already know that the function is truly periodic at $\tau_{RT} = 2\pi/\omega_c$ so the approximation is not serious.

The intensity is, of course, $e(t)\cdot e^*(t)/2\eta_0$, and thus the instantaneous power is given by

$$\frac{p(t)}{P_0} = \left(\frac{\pi}{2\ln 2}\right)\left(\frac{\Delta\omega}{\omega_c}\right)^2\exp\left[-\left(\frac{\Delta\omega t}{2(\ln 2)^{1/2}}\right)^2\right] \tag{9.5.5d}$$

where P_0 is power of the mode at line center. The pulse width Δt_p is found by setting the exponential term in (9.5.5d) equal to half and doubling the time interval to get the FWHM.

$$\Delta t_p = \frac{4\ln 2}{\Delta\omega} = \frac{4\ln 2}{2\pi\,\Delta\nu} = \frac{0.44}{\Delta\nu} \tag{9.5.6}$$

*It is important to consult the definition of the Fourier transform pairs if a different reference is used. Some make the pairs symmetric by dividing both integrals by $(2\pi)^{1/2}$. This definition conforms to that of the *Mathematical Handbook*, p. 175.

†Eq. (9.5.5a) can also be integrated directly by completing the square in the exponent.

This is the shortest pulse width possible given that its energy comes from a finite portion of the spectrum.* The numerical value 0.44 is valid only for a Gaussian, but the relationship to (9.5.6) is a good approximation for any reasonable "bell-shaped" curve in the frequency/time domain.

Equation (9.5.5.d) can be placed into a more convenient form by relating the power in the central mode to the total power in all modes, a quantity that is easily measured. We can add the individual powers of each mode [i.e., $\langle P \rangle = \sum P_n$ from (9.5.4a) or compute the average of (9.5.5d) with an averaging time of $2d/c$, the pulse repetition period.

$$\langle P \rangle = \frac{1}{2d/c} \int_{-d/c}^{+d/c} p(t)\, dt$$

$$\frac{\langle P \rangle}{P_0} = \frac{\omega_c}{2\pi} \int_{-\pi/\omega_c}^{+\pi/\omega_c} \left(\frac{\pi}{2 \ln 2} \right) \left(\frac{\Delta\omega}{\omega_c} \right)^2 \exp\left[- \left(\frac{\Delta\omega t}{(2 \ln 2)^{1/2}} \right)^2 \right] dt \quad (9.5.7a)$$

Now the exponential term is sharply peaked around $t = 0$ and is essentially zero at the times indicated by the limits of integration. With little error, then, the limits can be extended to $\pm\infty$, and we obtain a tabulated integral.

$$\frac{\langle P \rangle}{P_0} = \frac{1}{2} \left(\frac{\pi}{\ln 2} \right)^{1/2} \left(\frac{\Delta\omega}{\omega_c} \right) \quad (9.5.7b)$$

Thus the formula for the instantaneous power, (9.5.5d), can be rewritten in terms of the average power and the pulse width $\Delta t_{1/2}$, (9.4.6).

$$p(t) = \left(\frac{\pi}{\ln 2} \right)^{1/2} \left(\frac{\Delta\omega}{\omega_c} \right) \langle P \rangle \exp\left[-4(\ln 2) \left(\frac{t}{\Delta t_{1/2}} \right)^2 \right] \quad (9.5.5e)$$

This function is plotted as the solid curve in Fig. 9.17. This figure shows that the simplified procedure (9.5.2) → (9.5.3e) is only a fair representation of the pulse predicted by the more complicated analysis. But both theories predict an extremely short pulse, taxing the time response of the detectors and writing speeds of our best oscilloscopes.

This last case should be considered an example of the arithmetic to be followed in going from frequency domain specification to the time domain description of pulse. The first and most important step (and the one usually bypassed by the uninitiated) is to use the information on spectral distribution to establish an equation for the electric field (not the intensity) such as was done in formulating (9.5.4). Then convert the summation into an integral and evaluate.

All of the cases covered in this section presumed that some magic wand was waved to force all phases in (9.5.1) to be equal. We have focused on the natural multimode character of the inhomogeneously broadened lasers. If we think of our results for a moment, then we see that we can accomplish this task very simply. Include a modulator (or a repetitive

*The Gaussian pulse makes the Heisenberg uncertainty relationship [$\Delta\omega\Delta\tau \geq 1/2$, (1.6.1)] an equality. However, we must be careful with definitions. Equation (9.5.6) uses the FWHM, whereas the uncertainty relation deals with the rms deviation from the central peak.

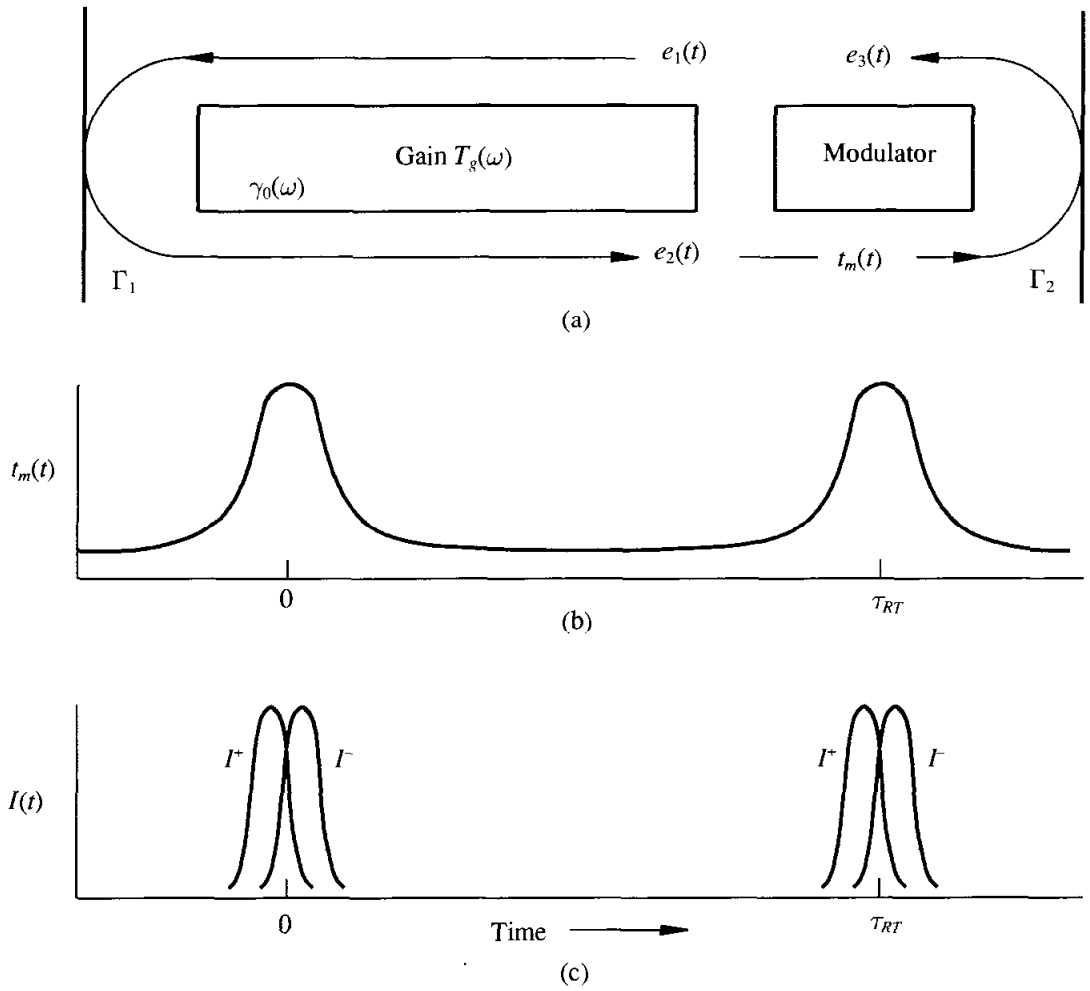


FIGURE 9.18. Mode locking of a laser. (a) Geometry showing the external modulator and the fields inside the laser cavity. (b) The transmission coefficient of the modulator. (c) The intensities arriving at the modulator.

shutter) inside the laser cavity and time the opening of the shutter to correspond to the arrival and passage of the pulse as indicated in Fig. 9.18. Such a system is not dependent upon locking the naturally occurring modes. It generates modes that naturally have a definitive phase relationship. This is the viewpoint of the next section.

9.5.3 Active mode locking

The simplicity of methods of locking the modes to a common time origin indicates that an analysis of a driven system, shown in Fig. 9.18, is in order. We will make no restriction as to the type of broadening although the functional form of a homogeneous system (Lorentzian line shape) will now be used as a convenient mathematical approximation for any bell shaped curve. The case of a strictly homogeneously broadened laser presents an interesting dilemma.

In principle, a laser operating on a homogeneously broadened transition should oscillate on a cavity mode with the highest gain-to-loss ratio to the demise of the other cavity modes. Based on previous discussion on mode locking in a multimode (inhomogeneous) laser, we would not expect similar phenomena in a homogeneously broadened system since there are no other modes! However, there is no law against inserting the time-varying modulator with either an amplitude or phase modulation, as shown in Fig. 9.18, into *any* laser cavity (the atoms have no vote on our actions). Our task is to describe the response of the total system, gain plus modulator, to the driven transfer function of the shutter.

Figure 9.18 suggests that such a ploy would be successful. If a pulse (however it gets started) arrives at the modulator when there is a high transmission coefficient, then it will pass through it with little absorption. If it arrives at the wrong time—when there is a lot of absorption—then that pulse does not survive. Thus we search for a self-consistent temporal (or pulse) solution for the fields in the laser cavity rather than a CW spectral representation. In spite of the differences in approaches, the two viewpoints lead to a common answer.

We assume an electric *field* (not intensity) entering the gain cell from the right given by Gaussian envelope function.

$$e(t) = A \exp\left[-2(\ln 2)\left(\frac{t}{\Delta t_p}\right)^2\right] e^{j\omega_0 t} = A\, e^{-at^2}\, e^{j\omega_0 t} \tag{9.5.8a}$$

with

$$\omega_0 = 2\pi c/\lambda_0 \qquad \text{and} \qquad a = 2(\ln 2)/\Delta t_p^2 \tag{9.5.8b}$$

The parameter a is just a convenient abbreviation for the factors involved with the pulse width. Follow that pulse through the gain cell and back to the entrance of the modulator whose time-varying (one-way) transmission coefficient is given by

$$t_m(t) = \exp\left[-\delta^2 \sin^2\left(\frac{\omega_m t}{2}\right)\right] \tag{9.5.9}$$

where ω_m is the (radian) modulation frequency that will turn out to be very close to the cavity mode spacing. As we will see, the choice of (9.5.9) for the transmission coefficient is for mathematical convenience with the quantity δ merely a measure of the on-to-off *field* transmission coefficient of the modulator.

The logic path for obtaining the self-consistent solution is as follows:

1. Start with $e_1(t)$ given by (9.5.8a). Find its spectral representation $E_1(\omega)$ by use of the Fourier transform.
2. Propagate $E_1(\omega)$ through the gain cell, to the mirror M_1 with a field reflection coefficient Γ_1 and back again to the exit to find $E_2(\omega)$. This requires us to find the complex transfer function (amplitude and phase) of the gain cell.
3. Go back to the time domain, via the inverse Fourier transform, to find $e_2(t)$, the field incident on the modulator.

4. Multiply $e_2(t)$ by the two-way transmission coefficient of the modulator, the square of (9.5.9), and the field reflection coefficient of M_2, Γ_2 to find $e_3(t)$, the field incident on the gain cell.
5. Compare $e_3(t)$ to $e_1(t)$. The field $e_3(t)$ must be the same as $e_1(t)$ arriving at the proper time (in the modulator period) and having the same shape in order to have obtained a self-consistent solution.

These logic steps use the following mathematics:

1. Spectral representation of $e_1(t)$. We use the Fourier transform defined by

$$E_1(\omega) = \int_{-\infty}^{+\infty} e_1(t)\, e^{-j\omega t}\, dt \tag{9.5.10a}$$

to find

$$E_1(\omega) = A\left[\left(\frac{\pi}{a}\right)^{1/2} e^{-[(\omega-\omega_0)^2/4a]}\right] \tag{9.5.10b}$$

2a. Transfer function of the gain cell. This is a homogeneously broadened transition, by assumption, and thus even the saturated line shape is a Lorentzian. However, we must use a *complex* gain coefficient to account for the amplitude *and phase* of the field. It is easy to show (see Chapters 13 and 14) that this complex gain coefficient is given by

$$\frac{\gamma(\omega)}{|\gamma(\omega_0)|} = \frac{1}{1 + j\left(\dfrac{\omega - \omega_0}{\Delta\omega/2}\right)} \tag{9.5.11a}$$

where $|\gamma(\omega_0)|$ is the usual *intensity* gain coefficient at line center and $\Delta\omega$ is the FWHM of that gain coefficient. A Taylor series expansion is used for (9.5.11a):

$$\frac{\gamma(\omega)}{|\gamma(\omega_0)|} = 1 - j\left(\frac{\omega - \omega_0}{\Delta\omega/2}\right) - \left(\frac{\omega - \omega_0}{\Delta\omega/2}\right)^2 \tag{9.5.11b}$$

Thus the single-pass transfer function through the gain cell is

$$T_g(\omega) = \exp\left\{-jkl_g + \frac{|\gamma(\omega_0)l_g|}{2}\left[1 - j\left(\frac{\omega - \omega_0}{\Delta\omega/2}\right) - \left(\frac{\omega - \omega_0}{\Delta\omega/2}\right)^2\right]\right\} \tag{9.5.12a}$$

where $|\gamma(\omega_0)|l_g/2$ must be used since we are dealing with fields rather than intensities.

2b. The field $E_2(\omega)$. The field $E_1(\omega)$ propagates to the left impinging on M_1 and thus is reduced by a factor Γ_1 and thus returns through the gain cell to the starting point of

our analysis.

$$E_2(\omega) = \Gamma_1 T_g^2(\omega) E_1(\omega) \tag{9.5.12b}$$

$$E_2(\omega) = \Gamma_1 A[\pi/a]^{1/2} \left\{\exp|\gamma(\omega_0) l_g|\right\} \times \left\{\exp\left[-\frac{(\omega-\omega_0)^2}{4a} - |\gamma(\omega_0) l_g|\left(\frac{\omega-\omega_0}{\Delta\omega/2}\right)^2\right]\right\} \times \left\{\exp\left[-j(k\cdot 2d) - j|\gamma(\omega_0) l_g|\left(\frac{\omega-\omega_0}{\Delta\omega/2}\right)\right]\right\} \tag{9.5.12c}$$

The total round-trip linear phase delay of the passive cavity, $k2d$, is used in (9.5.12c) to avoid repeating this calculation for each individual element or space. The imaginary terms in the exponent of (9.5.12c) correspond to the group delay of the pulse for a round trip. If we abbreviate $|\gamma(\omega_0)|l_g$ as g_0, rewrite k as

$$k = \frac{\omega}{c'} \equiv \frac{\omega_0}{c'} + \frac{(\omega-\omega_0)}{c'} \tag{9.5.13a}$$

and equate the imaginary terms of (9.5.12c) to a Taylor series expansion of the phase delay, we obtain the group velocity or group delay:

$$\frac{\omega_0}{c'} 2d + \left(\frac{\omega-\omega_0}{c'}\right) 2d + g_0\left(\frac{\omega-\omega_0}{\Delta\omega/2}\right) \triangleq \left[\beta_0 + \frac{\partial\beta}{\partial\omega}(\omega-\omega_0)\right] 2d \tag{9.5.13b}$$

where

$$\frac{\partial\beta}{\partial\omega} \triangleq \frac{1}{v_g} = \frac{1}{c'} + \frac{g_0}{2d}\frac{1}{(\Delta\omega/2)} \tag{9.5.13c}$$

Thus

$$\tau_{RT} = \frac{2d}{v_g} = \frac{2d}{c'} + \frac{g_0}{(\Delta\omega/2)} \tag{9.5.13d}$$

with

$$g_0 = |\gamma(\omega_0)|l_g = \text{single-pass line integrated gain} \tag{9.5.13e}$$

Thus the modulation frequency ω_m must be chosen correctly to allow passage of the pulse when the transmission coefficient is a maximum (at $\omega_m t/2 = m\pi$) in (9.5.9). This idea yields our first "design" equation:

$$\omega_m \tau_{RT} = 2m\pi$$

or

$$f_m = m\left[\frac{1}{\frac{2d}{c'} + \frac{g_0}{\Delta\omega/2}}\right] = \left[\frac{m\Delta\nu_c}{\left(1 + \frac{g_0\Delta\nu_c}{\pi\Delta\nu}\right)}\right] \tag{9.5.14}$$

Inasmuch as the correction factor in the denominator is small compared to one, (9.5.14) states that the modulation frequency should be slightly less than a multiple of the cavity intermode spacing $\Delta\nu_c = c'/2d$.

3. Find $e_2(t)$. Go back to the time domain by use of the inverse Fourier transform:

$$e_2(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} E_2(\omega)\, e^{j\omega t}\, d\omega \tag{9.5.15a}$$

and assume that the modulation frequency is chosen to obey (9.5.14) so that the phase factors of (9.5.12c) add to a multiple of 2π. Patience with the arithmetic yields

$$e_2(t) = \frac{\Gamma_1 A}{2\sqrt{qa}}(e^{g_0})(e^{-(t^2/4q)})(e^{j\omega_0 t}) \tag{9.5.15b}$$

where

$$q = \frac{1}{4a} + \frac{g_0}{(\Delta\omega/2)^2} \tag{9.5.15c}$$

4. Find $e_3(t)$. One multiplies (9.5.15b) by the square of the transmission coefficient of the modulator (9.5.9) (because of a double pass) and by the field reflection coefficient of M_2 to find $e_3(t)$. Assuming that the pulse arrives near $\omega_m t/2 = m\pi$, and expanding $\sin x \sim x$, we obtain

$$e_3(t) = \Gamma_1\Gamma_2\frac{A}{2}\frac{e^{g_0}}{\sqrt{qa}}\, e^{j\omega_0 t}\exp\left\{-\left[\frac{1}{4q} + 2\delta^2\left(\frac{\omega_m}{2}\right)^2\right]t^2\right\} \tag{9.5.16}$$

where time t now refers to one round trip later than was chosen for (9.5.8a).

5. Force self-consistency. We compare (9.5.16) to (9.5.8a) to obtain

$$A\, e^{j\omega_0 t} \Leftrightarrow \Gamma_1\Gamma_2 e^{g_0}\frac{A}{2}\frac{1}{\sqrt{qa}}\, e^{j\omega_0 t} \tag{9.5.17a}$$

and

$$at^2 \triangleq 2\ln 2\left(\frac{t}{\Delta t_p}\right)^2 \Leftrightarrow \left[\frac{1}{4q} + 2\delta^2\left(\frac{\omega_m}{2}\right)^2\right]t^2 \tag{9.5.17b}$$

where (9.5.8b) has been used to recover the definition of a. The solution for the FWHM of the pulse Δt_p from (9.5.17b) is

$$\Delta t_p = \frac{\sqrt{2\ln 2}}{\pi}\left(\frac{2g_0}{\delta^2}\right)^{1/4}\cdot\frac{1}{(f_m\Delta\nu)^{1/2}} \tag{9.5.18}$$

where it is assumed that the parameter $4g_0a/(\Delta\omega/2) = (8 \ln 2)g_0/(\Delta\omega\Delta t_p/2)^2$ is small; f_m is the modulation frequency given by (9.5.14), $\Delta\nu$ is the homogeneous line width, g_0 is the single-pass integrated gain, and δ is the measure of the on-to-off transmission of the modulator.

The other half of the self-consistency demand (9.5.17a), forces the integrated net gain, averaged over a round-trip time, to be equal to 1; that is,

$$\frac{\Gamma_1\Gamma_2}{2} e^{g_0} \frac{1}{\sqrt{qa}} = 1 \tag{9.5.19}$$

Now, by combining (9.5.8a), the definition of a, with (9.5.15c) for q, we have

$$qa = a\left[\frac{1}{4a} + \frac{g_0}{(\Delta\omega/2)^2}\right] = \frac{1}{4}\left[1 + \frac{(8 \ln 2)g_0}{(\Delta\omega\Delta t_p/2)^2}\right] \tag{9.5.20}$$

If we square both sides of (9.5.19), use this relationship for qa, and take the natural log of each side, we obtain a somewhat familiar expression:

$$2g_0 \triangleq |\gamma\omega_0)2l_g| = \ln \frac{1}{R_1R_2} + \ln\left[1 + \frac{(8 \ln 2)g_0}{(\Delta\omega\Delta t_p/2)^2}\right] \tag{9.5.21}$$

The first term on the right-hand side of (9.5.21) is merely the conventional round-trip cavity loss of a simple laser. The second term is the time-averaged loss introduced by the modulator. Some of the field passes through this modulator when $t_m < 1$, and thus there are photons absorbed, with the fraction lost per round trip equal to this last term. Recall that we assumed a Taylor series expansion for the gain coefficient [cf. (9.5.11b)], which tacitly assumed that the bandwidth of the pulse was much less than the available bandwidth of the gain. This amounts to assuming $\Delta\omega\Delta t_p/2 \gg 1$; thus the time-averaged loss in (9.5.21) is not large.

Thus we can "mode lock" a homogeneously broadened laser by generating the modes with the amplitude modulator. Although the pulse width (9.5.18) is not simply related to the amplifying bandwidth as was found in Sec. 9.5.2 and Sec. 9.5.3, a few numbers will convince us that the pulse is quite short.

We should solve for Δt_p and ω_m in terms of the parameters of the laser, linewidth, the modulator specifications, and the mirror reflectivities all of which determine the necessary line integrated gain g_0. It is much easier to assume g_0, and solve for the allowable range of mirror reflectivities that yield a self-consistent solution.

Consider the laser described by:

$\Delta\nu$ = transition linewidth = 2.5 GHz; $\Delta\omega = 1.57 \times 10^{10}$ r/s

Cavity length = 50 cm; $c' = c$; $\therefore \Delta\nu_c = c/2d = 300$ MHz

gain length $l_g = 40$ cm; line integrated gain $g_0 = 0.12$

Modulator specifications: $\delta = 1.1$; $\therefore t(\text{on}) = 1.0$ and $t(\text{off}) = 0.333$

Use (9.5.13d) to solve for the time for a round trip

$$\tau_{RT} = \frac{2d}{c} + \frac{g_0}{\Delta\omega/2} = 3.3486 \text{ ns} \qquad \therefore f_m = 298.63 \text{ MHz}$$

Use (9.5.18) to solve for the pulse width:

$$\Delta t_p = \frac{\sqrt{2 \ln 2}}{\pi} \left[\frac{2g_0}{\delta^2}\right]^{1/4} \frac{1}{(f_m \Delta\nu)^{1/2}} = 0.289 \text{ ns}$$

The loss introduced by the modulator is given by the last term in (9.5.21) and evaluates to

$$\ln\left[1 + \frac{(8 \ln 2)g_0}{(\Delta\omega\Delta t_p/2)^2}\right] = \ln(1 + 0.12916) = 0.1215$$

Thus the loss owing to the mirror reflectivities is given by

$$\ln \frac{1}{R_1 R_2} = 2g_0 - 0.1215 = 0.1185$$

or the product of the mirror reflectivities $R_1 R_2 = 0.888$.
The bandwidth required by this Gaussian pulse is related to its pulse width by (9.5.6):

$$\Delta\nu = 0.44/\Delta t_p = 1.52 \text{ GHz}$$

which fits nicely within the allotted bandwidth of the atomic transition.

Although a shutter or modulator driven by an external mechanism has been implied, that is not a necessity. If we can find a medium that can be changed by the initial portion of the pulse, the photons can operate the "toggle" switch on the shutter, changing a closed one to a transmissive medium. A very common example of this type of modulator is a saturable absorber in which the initial part of the wave saturates or bleaches the absorption (equalizes the populations) and thus the remainder of the pulse passes without loss. After passage of the main body of the pulse the populations relax back to the normal distribution so that the process can be repeated when the pulse returns after a round trip. This method of mode locking will be discussed after the dynamics of pulse amplification are discussed.

We can also use a phase modulator and obtain short pulses. We again presume a "pulse" solution for the fields within the cavity, which arrives at the proper time of the phase excursion of the modulator. We assume that the excess phase shift introduced by the modulator can be expressed as

$$t_m(t) = \exp[-j2\delta_\phi \cos \omega_m t] \tag{9.5.22a}$$

The proper time is near $t = 0$ so we expand the cosine into a Taylor series and use the following form for the transmission coefficient of the modulator:

$$t_m(t) \simeq \exp\left\{\mp j\left[2\delta_\phi - \frac{\delta\omega_m^2 t^2}{2}\right]\right\} \tag{9.5.22b}$$

This modulator impresses a time-varying phase on any field propagating through it, and thus we anticipate that the pulse may have "chirp"; that is, the frequency may be a function

of time. Therefore our assumed "pulse-like" solution should reflect this possibility.

$$e_1(t) = A\, e^{-at^2} \exp\left[j\left(\omega_0 + \frac{\Delta\omega t}{T}\right) t \right] \tag{9.5.23a}$$

where a is defined by (9.5.8b) and the parameter $b = \Delta\omega/T$ is a "chirp" parameter expressing the assumed linear variation of radian frequency with time

$$\omega(t) = \omega_0 + (\Delta\omega/T)t = \omega_0 + bt \tag{9.5.23b}$$

The same logic path is followed as was used for the amplitude modulator, but because of the chirp we encounter more involved arithmetic. A formal method to avoid repeating many of the manipulations is to replace a in the prior analysis by $a - jb$ for the present one. Thus, for instance, $E_1(\omega)$ is given by

$$E_1(\omega) = A\left\{ \left(\frac{\pi}{a - jb} \right)^{1/2} \exp\left[-\frac{(\omega - \omega_0)^2}{4(a - jb)} \right] \right\} \tag{9.5.24}$$

This field is propagated through the gain cell and back to and through the modulator in a round trip time as discussed previously. For the pulse arriving at the extremities of the phase modulator, the self-consistent requirement leads to

$$b = \pm a = \pi^2 f_m \Delta\nu \left(\frac{\delta_\phi}{2g_0} \right)^{1/2} \tag{9.5.25}$$

and the pulse width is given by

$$\Delta t_p = \frac{\sqrt{2 \ln 2}}{\pi} \left(\frac{2g_0}{\delta_\phi} \right)^{1/4} \left(\frac{1}{f_m \Delta\nu} \right)^{1/2} \tag{9.5.26}$$

Equation (9.5.26) is nearly the same as (9.5.18) except that δ appears rather than δ^2. The two possible solutions expressed by (9.5.25) represent the arrival at the two extremes of the phase modulator.

9.6 PULSE PROPAGATION IN SATURABLE AMPLIFIERS OR ABSORBERS

Mode-locking schemes produce much shorter pulses than do Q switched or gain switched lasers but at much less energy per pulse. To partially remedy the energy deficiency, we would send the pulses through an optical amplifier, whose dynamics are to be described here. Saturable absorbers are also described since they are often used to mode lock a laser and thus produce the shortest pulses on record. Absorption will be addressed by the simple

expedient of using a negative inversion (i.e., a normal population). Our attention will be limited to short duration pulses* carrying an energy per unit of area of $w(t)$ (joules/area).

Any optical amplifier is rarely operated in its linear regime since it is a waste of the very expensive pumping process to create the inversion. Why leave an inversion essentially "untouched," which is the result of operating in the linear or unsaturated regime? No, we would try to extract every bit of energy stored in this amplifier. Since the generation of one new photon reduces the upper state number ($N_2 \cdot$ volume) by 1 and adds 1 to the lower thereby changing the difference by 2, the maximum extractable energy is $h\nu(N_2 - N_1)l_g(\text{area}) \div 2$.

If we are close to extracting this maximum energy, then the time history of the pulse plays an important role. The leading edge of a pulse is always amplified more than the trailing edge (with the reverse being applicable for an absorber). Hence, we should expect significant distortion of the pulse envelope and this fact makes life a bit difficult. The intensity will change because of variations in time, space, and the inversion, which is also a function of (z, t').

$$\frac{\partial I}{\partial z'} + \frac{1}{v_g}\frac{\partial I}{\partial t'} = \Delta N(z', t')\sigma I(z, t') - \alpha_0 I(z', t') \tag{9.6.1a}$$

where t' is the time kept by a universal clock, $\Delta N(t') = N_2(t') - N_1(t')$, v_g is the group velocity, and α_0 represents an unavoidable unsaturable constant loss (dirt) distributed through the amplifier. We invoke the transformation, $z' = z$ and $t = (t' - z/v_g)$, in the manner used in Sec. 4.7 so as to measure time after the arrival of the leading edge of the pulse. Only the rate of growth with space at the local time remains.

$$\frac{\partial I(z, t)}{\partial z} = [N_2(z, t) - N_1(z, t)]\sigma I(z, t) - \alpha_0 I(z, t) \tag{9.6.1b}$$

For these short pulses, we can ignore any changes in the inversion due to pumping or due to normal relaxation processes and consider only the stimulated emission rates. The sum of the populations in states 2 and 1, $N_2 + N_1 = N$ does *not* change with time and is assumed to be independent of z.

$$\frac{\partial N_2}{\partial t} = -\frac{\sigma I}{h\nu}(N_2 - N_1) = -\frac{2\sigma I}{h\nu}N_2 + \frac{\sigma I}{h\nu}N \tag{9.6.2a}$$

or

$$\frac{\partial N_2}{\partial t} + \frac{2\sigma I}{h\nu}N_2 = \frac{\sigma I}{h\nu}N = \frac{N}{2}\frac{2\sigma I}{h\nu} \tag{9.6.2b}$$

Let us define some obvious meaningful quantities to simplify (9.6.2):

$$w(z, t) = \int_{-\infty}^{t} I(z, t)\, dt \tag{9.6.3a}$$

is energy (per unit of area) of the pulse, $I(z, t) = \dot{w}(z, t)$,

$$w_s = \frac{h\nu}{2\sigma} \tag{9.6.3b}$$

*A pulse will be classified as short if its time duration is much less than any characteristic time scales in the atomic system interacting with the pulse. Such a situation presents difficulties not encountered previously.

is a characteristic saturation energy (area^{-1}) of the medium,

and
$$u(z, t) = \frac{w(z, t)}{w_s} \tag{9.6.4}$$

is an energy normalized to the saturation value.

Equation (9.6.2a) becomes nice and compact.

$$\frac{\partial N_2}{\partial t} + \dot{u}N_2 = \frac{N}{2}\dot{u}$$

Multiply by an integrating factor:

$$\exp\left[\int_{-\infty}^{t} \dot{u}\, dt\right] = \exp\left[u(t)\right]$$

which makes both sides a perfect differential (with respect to time).

$$\frac{\partial}{\partial t}\left[N_2\, e^{u(z,t)}\right] = \frac{N}{2}\left\{\dot{u}\, e^{u} = \frac{\partial}{\partial t}\, e^{u}\right\}$$

and thus the integral is

$$N_2(z, t) = \frac{N}{2} + K\, e^{-u(z,t)} \tag{9.6.5}$$

The constant of integration, K, is evaluated by the conditions at $t = -\infty$ (i.e., before the pulse arrives) when the energy is zero: $u(z, -\infty) = 0$ and $e^{-u} = 1$. Now $N_2^0 - N_1^0 = \Delta N^0$ is the initial inversion determined by the pumping and natural relaxation processes of the system and $N_2^0 + N_1^0 = N$, a constant. Hence the sum of the two initial values leads to

$$N_2^0 = N_2(t = -\infty) = \frac{N}{2} + \frac{\Delta N^0}{2}$$

Thus, the integration constant $K = \Delta N^0/2$, and we obtain

$$N_2(t) = \frac{N}{2} + \frac{\Delta N^0}{2}\, e^{-u(z,t)} \tag{9.6.6a}$$

$$N_1(t) = \frac{N}{2} - \frac{\Delta N^0}{2}\, e^{-u(z,t)} \tag{9.6.6b}$$

and by subtraction

$$N_2(t) - N_1(t) = \Delta N(t) = \Delta N^0\, e^{-u(t)} \tag{9.6.7}$$

Now we multiply (9.6.7) by the gain cross section, substitute it into (9.6.1b), and divide by the saturation energy, which makes an important equation appear less cluttered.

$$\boxed{\frac{\partial}{\partial z}\dot{u}(z, t) = \gamma_0 \dot{u}(z, t)\, e^{-u(z,t)} - \alpha_0 \dot{u}(z, t)} \tag{9.6.8}$$

where $\gamma_0 = \Delta N^0 \sigma$ is the small signal gain coefficient (at $t = -\infty$). Equation (9.6.8) is a form of the Frantz–Nodvik equation [19] modified to account for the presence of unsaturable loss. It can be integrated with respect to time:

$$\frac{\partial}{\partial z}\int_{-\infty}^{t} \dot{u}(z,t)\,dt = \gamma_0 \int_{-\infty}^{t} \dot{u}\, e^{-u}\,du - \alpha_0 \int_{-\infty}^{t} \dot{u}\,dt$$

or

$$\boxed{\frac{\partial}{\partial z}u(z,t) = \gamma_0\left[1 - e^{-u(z,t)}\right] - \alpha_0 u} \tag{9.6.9}$$

where the value of $u(z, t = -\infty) = 0$ has been used again. The presence of the unsaturable loss impedes progress with analytic integration so let us temporarily set it equal to zero and integrate between the input (1) and output (2) planes of an amplifier l_g units long.

$$\int_{u_1}^{u_2} \frac{du}{1 - e^{-u}} = \gamma_0 \int_0^{l_g} dz = \gamma_0 l_g = \ln G_0$$

where G_0 is the small signal gain $\exp[\gamma_0 l_g]$. This is a tabulated integral [(14.515) of the *Mathematical Handbook*]

$$(u_2 - u_1) + \ln\left(\frac{1 - e^{-u_2}}{1 - e^{-u_1}}\right) = \gamma_0 l_g = \ln G_0$$

After a bit of rearrangement, we obtain

$$\boxed{e^{u_2(t)} = 1 + G_0\left(e^{u_1(t)} - 1\right)} \tag{9.6.10}$$

where the time origin for the output pulse at plane 2 is shifted by l_g/v_g with respect to the input time base so that undistorted pulses would overlap. The output intensity is found by differentiation of (9.6.10) with respect to the local time. Here, $w_s \dot{u}_2\, e^{u_2(t)} = w_s G_0 \dot{u}_1\, e^{u_1(t)}$, use (9.6.10) for $e^{u_2(t)}$,

$$w_s \dot{u}_2 = \frac{w_s G_0 \dot{u}_1\, e^{u_1(t)}}{1 + G_0\left(u^{u_1(t)} - 1\right)}$$

or

$$\boxed{I_2(t) = I_1(t)\frac{G_0\, e^{u_1(t)}}{1 + G_0\left[e^{u_1(t)} - 1\right]}} \tag{9.6.11}$$

It is worthwhile checking on the logic of (9.6.11). For instance, we can always find a time interval on the leading edge of the pulse where the normalized energy $u_1 \ll 1$, $e^{u_1(t)} \sim 1$ and the output is a replica of that part of the input but amplified by G_0. If $u_1(t) \gg 1$, then everything can be neglected except the exponential terms and (9.6.11) becomes $I_2(t) = I_1(t)$.

If the system were a saturable absorber, the definitions would be the same, the population difference would be a normal one and $G_0 < 1$, but the arithmetic would remain. However, the case of the absorber changes an important feature of the output as the following example illustrates.

Let the input intensity be given by a "smooth" bell-shaped function of the form

$$I_1(t) = \frac{w_0}{T} \sin^2\left(\frac{\pi t}{2T}\right) \qquad 0 < t < 2T$$

where w_0 is total energy (per unit of area) in the pulse.

Figure 9.19 shows the output divided by the peak input intensity (w_0/T) for various values of w_0/w_s for $G_0 = 4$ (i.e., 6 dB amplifier) and for a 6 dB attenuator. For small values of w_0/w_s, the output is just four times larger than the input for the case of the amplifier and is 25% of that value for the attenuator. For w_0/w_s equal to $\sim$ 1 to 2, the pulse shape is quite distorted with the amplified peak appearing "earlier" than that of the input delayed by l_g/v_g and thus the "risetime" of the pulse is shortened. For the case of a saturable absorber with $G_0 = 1/4$ (i.e., a 6 dB attenuator) the output bleaches the absorption and appears "later" or the "trailing edge" of the pulse is sharpened. These differences between an amplifier and an absorber play a role in mode locking with saturable absorbers.

If $u_1 \gg 1$ and thus so also is u_2 for the amplifier, then (9.6.10) degenerates to

$$u_2 = u_1 + \ln G_0 = u_1 + \gamma_0 l_g$$

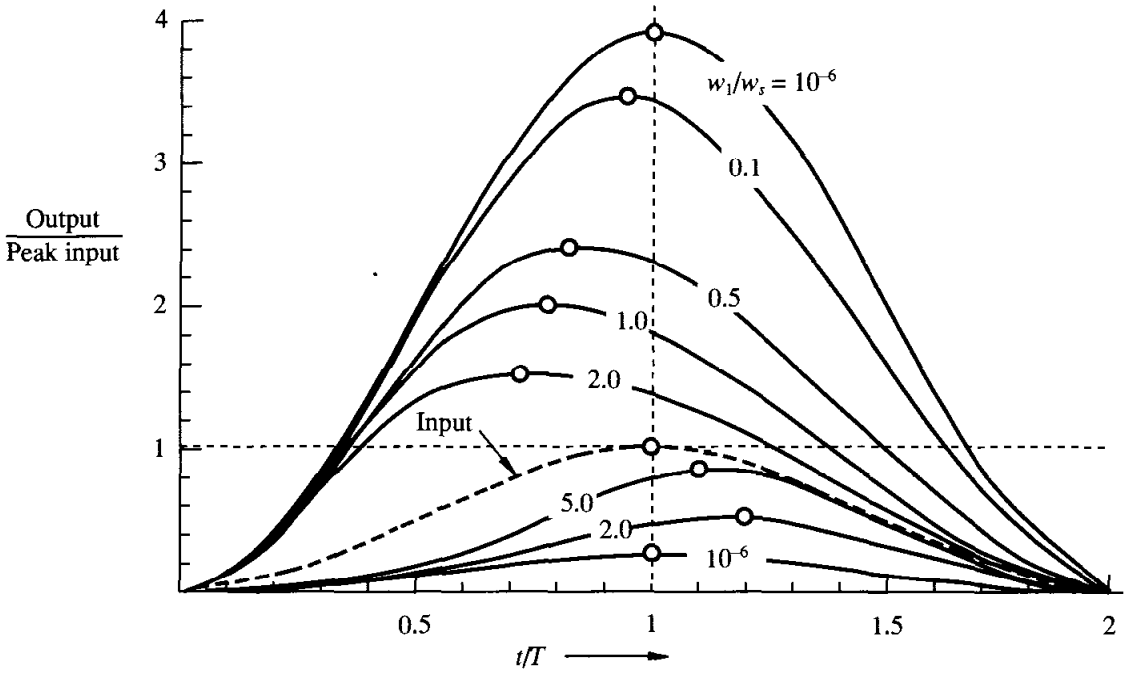


FIGURE 9.19. The relative transmission through a saturable amplifier (with values > 1) or through an absorber (values < 1) for various ratios of the input energy divided by the saturation energy of the medium. $G_0 = 4$ for the amplifier and $G_0 = 0.25$ for the absorber. Note that the peaks (denoted by the open circles) arrive earlier for the amplifier, later for the attenuator.

or

$$w_2 - w_1 = h\nu \left(\frac{\Delta N^0}{2} \right) \tag{9.6.12}$$

This is the maximum energy that can be extracted from this amplifier (or can be absorbed) and can be verified in an approximate fashion for the case labeled $w/w_s = 1$ in Fig. 9.19. We can mentally integrate the area under the curve for that case: It is ~ 2 units high with a FWHM of $\sim T$ units yielding $\sim 2T$ units of output energy, one of which was supplied by the input pulse.

The results are modified rather significantly if $\alpha_0 \neq 0$, but unfortunately we must resort to a numerical solution for arbitrary values of the parameters. There are two limiting cases that can be addressed by analytic methods.

1. If $u(z, t) \ll 1$, then (9.6.9) becomes

$$\frac{\partial u}{\partial z} - (\gamma_0 - \alpha)u = 0$$

or

$$w_2 = w_1 \exp\left[(\gamma_0 - \alpha)l_g\right] \tag{9.6.13a}$$

which is the expected small signal result.

2. If the energy is large such that $\exp[-u] \ll 1$, then (9.6.9) integrates to

$$w_2 = w_s \frac{(\gamma_0 l_g)}{(\alpha l_g)} \left(1 - \exp[-\alpha l_g]\right) + w_1 \exp[-\alpha l_g] \tag{9.6.13b}$$

For a "long" amplifier such that $\alpha l_g \gg 1$, then the limiting output is

$$w_2 = w_s \frac{(\gamma_0 l_g)}{(\alpha l_g)} \tag{9.6.13c}$$

which may—or may *not*—be bigger than the input energy. This is a rather dismal fact to contemplate: Our amplifier may become an attenuator!

A numerical integration of (9.6.9) is shown in Fig. 9.20 for the case where the loss is 10% of the gain coefficient. If $w_1/w_s = u_1 < \gamma_0/\alpha$, the energy is amplified but approaches the asymptotic limit given by (9.6.13c) even with an arbitrarily large value of the line integrated gain.* If $w_1/w_s > \gamma_0\alpha$, our expensive and complicated optical amplifier becomes an attenuator with the output approaching the same limit, but now from above.

It should be clear from this figure that the presence of unsaturable loss can make a significant difference in the performance of an optical amplifier and thus great effort should be expended to avoid it in an amplifier.

*An amplifier with $\gamma_0 l_g > 10$ would imply a small signal gain of > 43 dB, which is most impractical for a variety of reasons, not the least of which would be the difficulties with amplified spontaneous emission (see Sec. 8.7). The deleterious effects of the unsaturable loss are clear even at small values of $\gamma_0 l_g$.

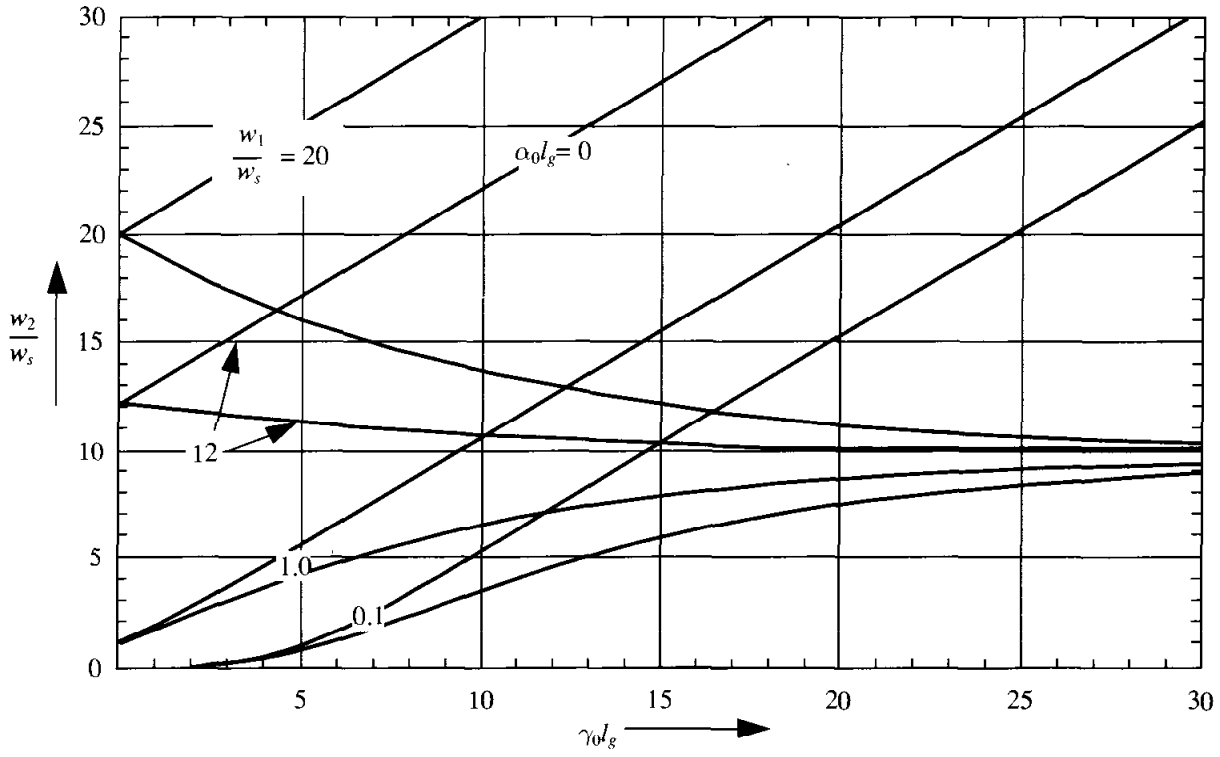


FIGURE 9.20. The normalized output energy u_2 of an amplifier as a function of the line integrated small signal gain, $\gamma_0 l_g$ for various values of u_1 equal to 0.1, 1.0, 12, and 20. The curves that are asymptotic to $u = 10$ assume a distributed loss of 10% of the gain coefficient, whereas the others assume $\alpha = 0$.

9.7 SATURABLE ABSORBER (COLLIDING PULSE) MODE LOCKING

The mode-locking scheme that has generated the shortest pulses uses a saturable absorber in the optical cavity so that the photons operate the switch. In its simplest form, a saturable absorber replaces the modulator of Fig. 9.19, but the world record for short pulses uses the "colliding pulse" mode-locking (CPM) geometry, which is shown in simplified form (used by Folk, Greene, and Shank) in Fig. 9.21a with a schematic of the path shown as a circular ring in Fig. 9.21b.

The easiest way to understand its operation is to assume the known operating conditions and then argue that a deviation from these conditions makes matters worse. The critical points are

1. The absorber is located one-fourth of the perimeter away from the gain medium as shown in Fig. 9.21(b). Both the gain and absorber lengths are very small.
2. The two oppositely directed pulses arrive at the absorber at the same time (i.e., collide) and create a grating caused by the interference of the two waves.
3. Each pulse arrives at the gain medium in sequence space by $\tau_{RT}/2$.
4. There will be some additional subtle conditions placed on the spot sizes at the absorber/gain locations, the absorber and gain cross sections, and the relaxation rates.

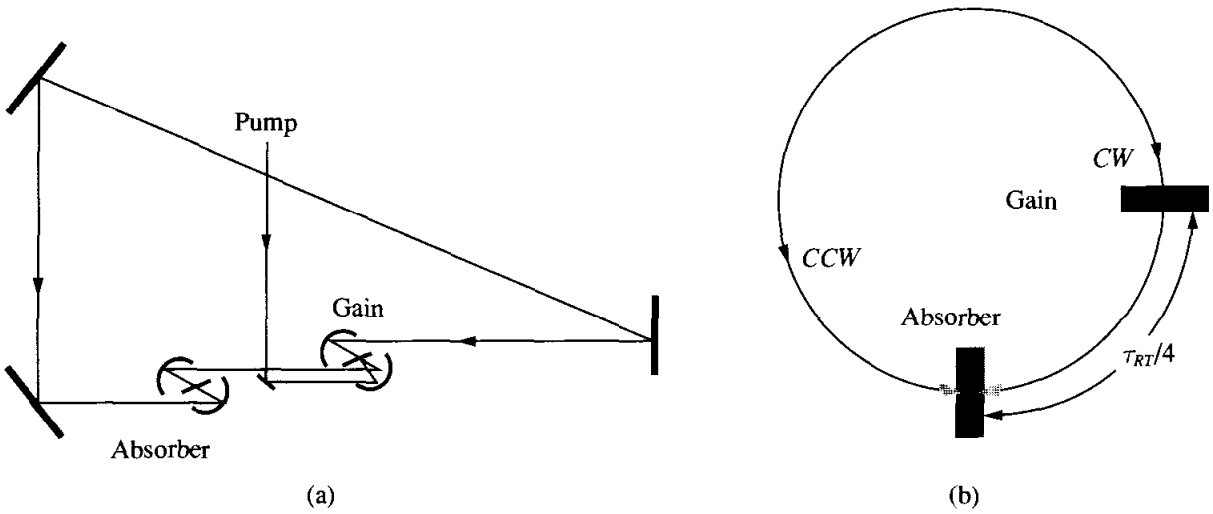


FIGURE 9.21. (a) A typical geometry of a CPM laser. (b) The circular schematic of the optical path showing the timing of the collision of the two circulating pulses in the absorber. (Adaptation of Fig. 1 of [21] and Fig. 3 of [22].

If the two pulses do "collide" in the absorber, they will effect a much higher degree of saturation (or reduction) of its loss than if the pulses were to arrive in sequence. The interference between the two pulses creates a "grating" in the saturated absorption (with minimum attenuation where the two fields add, maximum α at the destructive interference planes where the field is a minimum), which thus couples the clockwise pulse with the counterclockwise one. Since the collision of the two pulses at the absorber represents a minimum energy loss situation (since the loss is proportional to the integral of αE_T^2), the laser is self-starting in this mode and the resultant grating helps with the synchronization in the case where all dimensions are not perfectly correct.

The pulse traveling in the counterclockwise (CCW) direction arrives at the gain medium at $t = \tau_{RT}/4$, gains energy according to the theory presented in Sec. 9.6, saturates the gain medium, and then proceeds toward the absorber, arriving at τ_{RT} and meets the clockwise (CW) pulse there. Both saturate the absorber and the induced grating couples a bit of one into the other. The gain medium is continuously being pumped, and hence the line integrated gain recovers from the CCW interaction back toward the small signal gain g_0 in the manner shown in Fig. 9.22. However, the CW directed pulse arrives at $t = (3/4)\tau_{RT}$, gains energy by resaturating the gain medium and proceeds around the loop for another round trip. Each pulse encounters the same gain—the initial value encountered is not the small signal value g_0, but rather the partially recovered value g_i—and both leave it in the same state of saturation g_f.

While Fig. 9.21 implied a dye laser in a ring geometry with very short gain and absorber length, as was used for the first CPM laser, saturable absorber mode locking had a long history preceding that invention and has been accomplished in the simple standing wave type of geometry. The various analyses can be applied to the CPM geometry (although the coupling via the induced grating in the absorber is unique to the latter case). New and Haus [23,24,26] have addressed this problem and have discovered that three criteria must

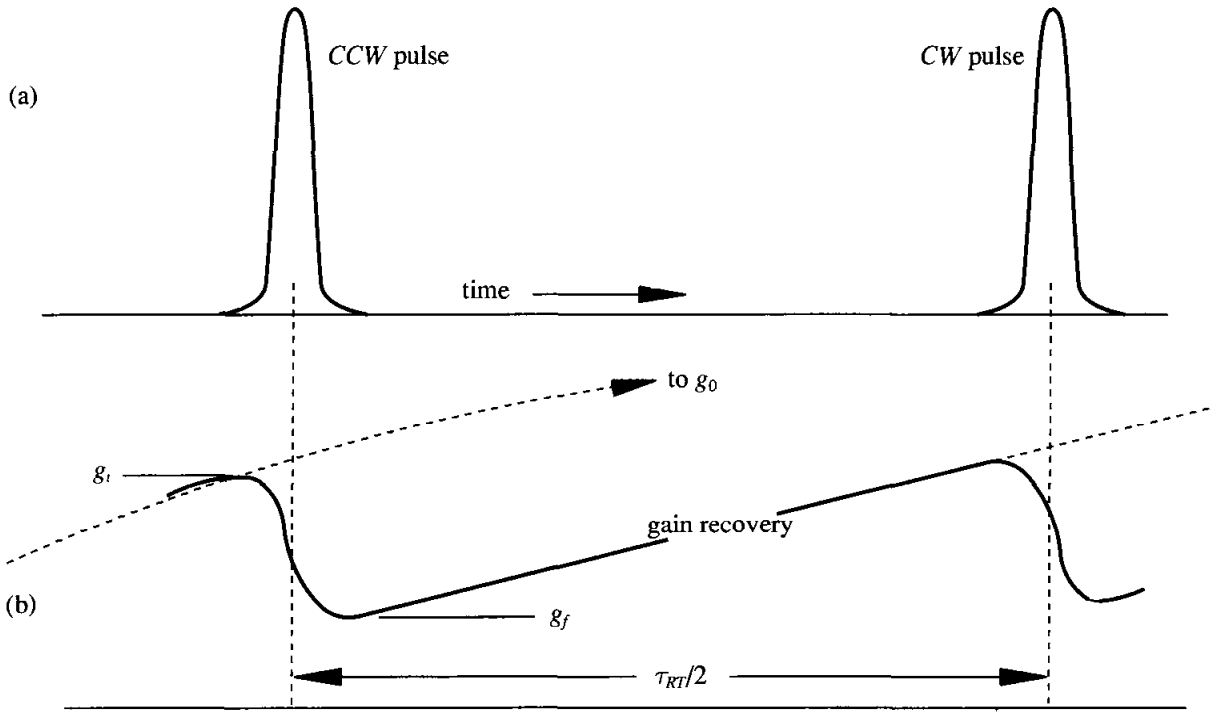


FIGURE 9.22. The interaction of the counter-propagating pulses with the gain medium. While the gain recovers between interrogations, it does not recover to the small signal value.

be satisfied to have a steady state and stable mode locked laser with a saturable absorber for the following assumed conditions.

1. The gain does recover with the interval between pulses but only to g_i, not completely to its small signal value g_0 as shown in Fig. 9.22. Thus, the time constant for gain recovery T_g is assumed to be larger than the time between pulses, but not excessively so.
2. The absorber is presumed to recover much quicker in the interval between pulses, and thus the time constant for absorber recovery T_a is less than the time between pulses.
3. The pulse width Δt_p is very short compared to τ_{RT}, T_a, or T_g.

Then the criteria for stable pulse formation is that:

1. The energy in the pulse after a round trip is a constant.
2. The leading edge of the pulse must experience a net loss.
3. The trailing edge must also experience a loss.

To *prove* these criteria is a chore best left to the literature or to more advanced texts. Here, let us agree on the implications as they apply to Fig. 9.23. The first criterion is more or less the self-consistency statement (i.e., the pulse should reproduce itself after a round trip). Since there must be net loss on the leading and trailing edges, there must be net energy gain

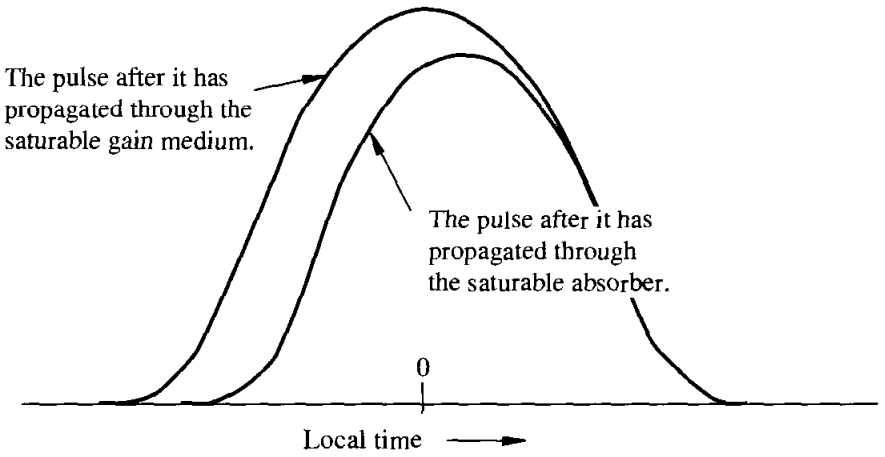


FIGURE 9.23. The transmission of a pulse through a saturable absorber or amplifying medium. The larger pulse should be considered as the input to the absorber whose output is the smaller pulse which, in turn, is the input to the amplifier.

in the central portion of the pulse to make the energy a constant. A moment of consideration also leads to the conclusion that the absorption cross section must be at least two times that of the amplifier so that the absorber is fully bleached by the peak while the gain cell can continue to amplify. These criteria tend to compress or sharpen the pulse and are explained more fully in the papers by New [23, 24].

The dynamics of an optical pulse encountering the combination are shown in Fig. 9.23 where a "bell-shaped" pulse emerging from the absorber is shown as the dotted curve and the theory of Sec. 9.5 is applied assuming that the solid curve is the input wave. The absorption is greater on the leading edge leaving the trailing edge behavior more or less equal to the incident. The action of the saturable gain is just the reverse—more amplification on the initial part of the pulse and unity transmission on the trailing side—and thus a restoration toward the original pulse.

There is no simple way of predicting the pulse width, especially when extreme cases are encountered. The three criteria mentioned tend to shorten the pulse. Eventually, other effects come into play. For instance, some of the pulses are so short in time and thus so broad in spectral content that a simple time delay of $z/(c/n)$ relating fields separated by a distance z is not adequate for any medium other than a vacuum. The index of refraction of everything composed of atoms—even air—is a function of wavelength, and thus the different wavelengths travel with different group velocities: The shorter the pulse width, the more serious is the problem, and the opportunities.

To illustrate the opportunities, Fork, Brito, Cruz, Becker, and Shank [29] used a CPM laser with the four-prism sequence of Fig. 9.24 (to be explained later) to replace the straight-line path to generate and subsequently amplify 50 fs pulses to a 1 mJ level (which corresponds to roughly 20 GW of peak power). A fraction of this energy was coupled into a fiber (0.9 cm long) where the intensity reached $\sim 10^{12}$ W/cm^2, sufficient so that the optical Kerr effect broadened the spectral content considerably by introducing a frequency chirp.* The output was sent through a grating-pair plus prism pair sequence of Fig. 9.24 to

*This is the same effect as discussed in Chapter 4 in connection with solitons.

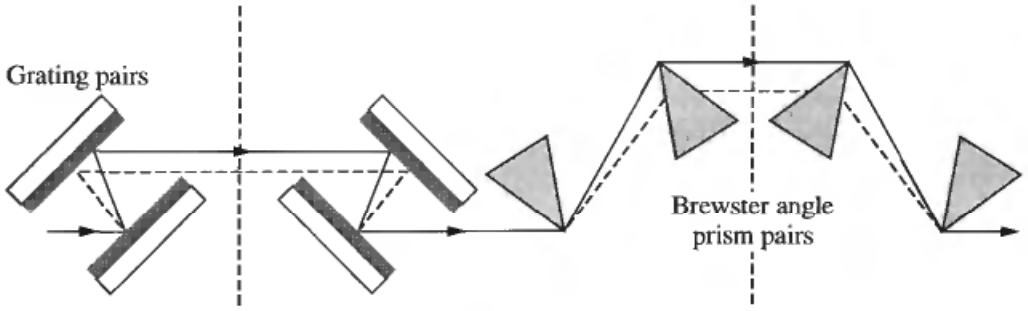


FIGURE 9.24. A sequence of gratings and prisms used by Fork, Martinez, and Gordon [32] to compensate the quadratic and cubic phase distortion. The path shown dotted is one followed by a different wavelength.

minimize the dispersion of the different spectral components and obtained a compressed pulse of only 6 fs centered at 620 nm. (Such a pulse has only about three optical cycles in the FWHM, plus a few more to make up the wings.)

The delay experienced by an electromagnetic wave is

$$t_d = \frac{d\phi(\omega)}{d\omega} \tag{9.7.1}$$

where $\phi(\omega)$ is the phase delay of each spectral component and is equal to $\omega n(\omega)z/c$. Expanding the phase in a Taylor series yields

$$\phi(\omega) = \phi(\omega_0) + \left(\frac{\partial\phi}{\partial\omega}\right)_{\omega_0}(\omega - \omega_0) + \frac{1}{2}\left(\frac{\partial^2\phi}{\partial\omega^2}\right)_{\omega_0}(\omega - \omega_0)^2$$

$$+ \frac{1}{6}\left(\frac{\partial^3\phi}{\partial\omega^3}\right)_{\omega_0}(\omega - \omega_0)^3 + \cdots \tag{9.7.2}$$

The gratings can be used to compensate for quadratic phase distortion and the combination of the grating pairs and the Brewster angle prism can equalize the delay for up to the third order deviation in ω or $d^3\phi/d\omega^3$.

The Brewster angle prism can also be used inside the CPM laser to compensate for dispersion in other components. Because all surfaces are at the Brewster angle, there is virtually no loss and the input and output paths are parallel. By translating a prism in a direction normal to its base and/or changing the spacing between them one can make a second derivative of the optical path $P = \int n(z)\,dz$.

$$\frac{d^2P}{d\lambda^2} - \frac{d^2}{d\lambda^2}\int n(z)\,dz$$

be either positive or negative to compensate the usual positive values in other components.

Mode locking by the use of saturable absorbers has progressed to the point where the fundamental physical constraint of available bandwidth is the limitation. It should also be noted that the measurement of such short pulses is a significant problem, a subject reserved for more advanced books.

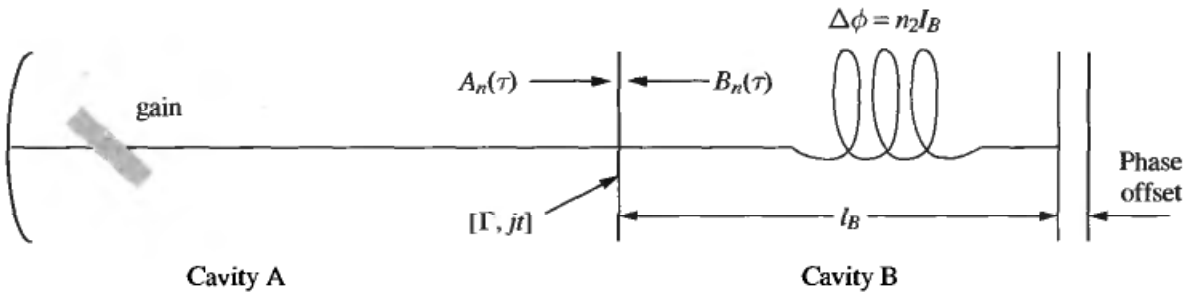


FIGURE 9.25. The geometry used by Wang [34] for the analysis of passive additive-pulse mode locking. The element common to both cavities is characterized by a field reflection coefficient Γ and a field transmission coefficient jt such that $|\Gamma|^2 + |t|^2 = 1$.

9.8 ADDITIVE-PULSE MODE LOCKING

There is another scheme for mode locking a laser that is a bit more subtle than the CPM method, but is easier to analyze. Consider Fig. 9.25, which depicts a cavity (A) with gain coupled to an external cavity (B) with a nonlinear phase modulator (usually it is a fiber) imparting an extra intensity dependent phase shift on the fields in (B). If we adjust the lengths of the two cavities properly, then the feedback from cavity B tends to add to the fields in A at the peak of a pulse but subtracts in the wings. In this case, a "pulse solution" is favored with successive additions progressively shortening the duration of the pulses. Eventually, dispersion in the components and bandwidth limitations in the gain medium counteract the pulse shortening force of the addition.

Let us consider the electric field impinging on the coupler in cavity A after n round trips to be

$$E_n(\tau) = A_n(\tau) \exp[j(\omega\tau)] \tag{9.8.1}$$

where τ is to be interpreted as the local time of arrival at the coupling mirror. That field, of course, traversed the space to the left of the coupling mirror at earlier times given by $\tau - z/v_g$. A similar notation is used for cavity B with the proviso that the round trip times of A and B are identical. The analysis consists of "adding" the (field) envelopes after successive round trips through the two cavities.

If both cavities were passive and linear, then the envelopes would simply interfere at the coupling plane. The $(n+1)^{st}$ envelope of the field on the left A_{n+1} is given by $\Gamma \cdot A_n$ (Γ is the field reflection coefficient) plus jtB_n (t is the field transmission coefficient), where B_n is the envelope of the field in cavity B.*

$$A_{n+1} = \Gamma A_n + jtB_n \tag{9.8.1a}$$

Similarly, we would relate the $(n+1)^{st}$ envelope of the field in cavity B by

$$B_{n+1} = jtA_n + \Gamma B_n \tag{9.8.1b}$$

*This is the same notation for the scattering matrix of a mirror as was used in Sec. 6.1. A brief rereading of Appendix I on the scattering matrices is appropriate for those who feel uneasy about the presence of j.

If cavity A has a *field* gain of $G^{1/2}$ per pass, then both A_n and B_n are multiplied by $(G^{1/2}) \cdot (G^{1/2}) = G$ in (9.8.1a) where G is the *single-pass* power gain or the round trip field gain. The transmitted envelope A_n and the reflected one ΓB_n (9.8.1b) experience a round trip transmission factor (T), and an excess round trip phase shift $e^{j\phi}$ associated with the external cavity, which is abbreviated by $T\,e^{j\phi} = L$. Thus the matrix form of (9.8.1) becomes

$$\begin{bmatrix} A_{n+1}(\tau) \\ B_{n+1}(\tau) \end{bmatrix} = \begin{bmatrix} \Gamma \cdot G & jt \cdot G \\ jt \cdot (T\,e^{j\phi} = L) & \Gamma \cdot (T\,e^{j\phi} = L) \end{bmatrix} \begin{bmatrix} A_n(\tau) \\ B_n(\tau) \end{bmatrix} \tag{9.8.2}$$

The pulse amplitude (represented by the envelopes) will grow with time if one of the eigenvalues of (9.8.2) is bigger than 1. These eigenvalues are found by assuming

$$A_{n+1}(\tau) = \Lambda A_n(\tau) \quad B_{n+1}(\tau) = \Lambda B_n(\tau) \quad \text{or} \quad (A_n(\tau) = \Lambda^n A_0 \text{ and } B_n = \Lambda^n B_0) \tag{9.8.3a}$$

which, when substituted into Eq. 9.8.2 yield

$$\begin{bmatrix} \Gamma G - \Lambda & jtG \\ jtL & \Gamma L - \Lambda \end{bmatrix} \begin{bmatrix} A_n(\tau) \\ B_n(\tau) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{9.8.3b}$$

The determinant of the first matrix must be zero in order to have a nontrivial solution, which leads to a quadratic equation for the eigenvalue.

$$\Lambda^2 - \Gamma(G + L)\Lambda + GL = 0$$

$$\therefore \Lambda_{2,1} = \frac{\Gamma(G + L) \pm \left\{\Gamma^2(G + L)^2 - 4GL\right\}^{1/2}}{2} \tag{9.8.4}$$

The eigenvalue Λ can be interpreted as the net field gain per round trip for a particular combination of A_n and B_n. Equation (9.8.4) is a bit deceptive because the presence of the complex number $e^{j\phi}$ and its nonlinear dependence $\phi \sim kI_B$ is hidden in the simple letter symbol L. However, some of the trends can be established fairly easily.

If the field reflection coefficient were perfect such that $|\Gamma| = 1$, then the two cavities would be decoupled with the two eigenvalues being G for cavity A, which corresponds to the plus sign and must be > 1 to even consider lasing, and L for cavity B and, since it is composed of all passive components, its magnitude < 1. The product of the two eigenvalues in any case is $G \cdot L$, a complex number.* Thus, there is a pulse solution whose amplitude grows with round trip number ($\Lambda_2 > 1$) until gain saturation occurs so that $|\Lambda_2| = 1$ with a particular excess phase shift for steady state pulse solution.

While the *possibility* of a steady state pulse solution can be argued, this does not prove that one does indeed exist. There is nothing in what we just did to discriminate between a "long" pulse solution (with $\Delta t_p > \tau_{RT}$; i.e., a CW laser) and a short pulse solution. Just as there must be a gain nonlinearity (i.e., saturation) to establish the energy in the pulse, there must be nonlinearity to enable the short pulse solution to be selected.

*We will keep the notation that Λ_2 is the largest eigenvalue, even though it may correspond to the negative sign choice.

This is the function of the fiber (or a Kerr medium) that has linear phase function and also an intensity dependent part.

$$\phi = \frac{\omega}{c}(n_0 + n_2 E^2)1_B \sim \phi_0 + kI_B \tag{9.8.5}$$

Hence, the peak of a pulse would undergo a greater phase shift than does the wings. The key trick is to choose the length of the Kerr medium (usually a single-mode fiber) and the phase offset such that the rate of change in the eigenvalue Λ_2 with respect to ϕ (and thus the intensity) is positive. A mode locking force can be defined by

$$f = \frac{\partial \Lambda_2}{\partial \phi} \tag{9.8.6}$$

so that the change in Λ moves the envelope toward a short-pulse solution by making its net gain larger.

$$\Delta\Lambda_2 = \frac{\partial \Lambda_2}{\partial \phi}\Delta\phi \approx \frac{\partial \Lambda_2}{\partial \phi}\frac{\partial \phi}{\partial I}[\Delta I = I_{\text{peak}} - I_{\text{wings}}] \tag{9.8.7}$$

The nonlinear index in (9.8.5) is positive (for $\lambda > 1.3\ \mu$m) and hence $\partial\phi/\partial I$ is also, and surely $I_{\text{peak}} > I_{\text{wings}}$ of a pulse. Thus it is only necessary to find an operating point ϕ_0 such that $\partial\Lambda/\partial\phi > 0$ so that the "net gain" on the shortest pulse is the largest. As we can anticipate, most of the calculations must be done on a computer, but the laser has no problem with our arithmetic.

This approach, which is an adaptation of Wang's [34] analysis, is the simplest method of understanding additive-pulse mode locking. Ippen, Haus, and Liu [35] have formulated a differential equation theory that includes dispersion effects so as to predict a closed form of the pulse width. A reasonable first start on the literature on this subject is given in [34] to [43].

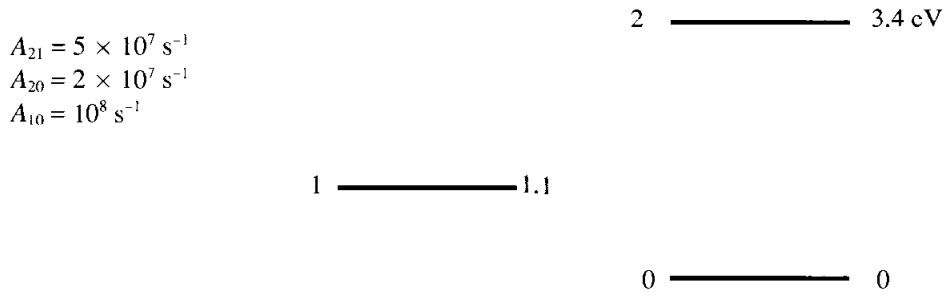
PROBLEMS

9.1. The laser transition in neon at 6328 Å is Doppler broadened by the thermal motion of the atoms at a gas temperature of $\sim 300°$C (assume pure Ne^{20}). (Volume $= 1$ cm^3.)

- **(a)** What is the full width at half maximum of the transition? (Ans.: 1.81 GHz.)
- **(b)** How many blackbody modes couple to this transition [use $\Delta\nu$ from (a)]? (Ans.: 3.79×10^8 modes.)
- **(c)** Suppose that a He:Ne laser tube is placed inside a cavity 1 m long. How many $\text{TEM}_{0,0,q}$ modes are within $\pm\Delta\nu/2$ of the line center? (Ans.: 12.)
- **(d)** What is the relative probability of a group of neon atoms radiating into one of the $\text{TEM}_{0,0,q}$ modes as opposed to all of the blackbody modes? (Ans.: $p = 3.2 \times 10^{-8}$.)
- **(e)** The energy of the upper state, $3s_2$, is 166, 658.484 cm^{-1} above the ground state. Express the upper- and lower-state energies in eV.
- **(f)** What is the quantum efficiency of this laser?
- **(g)** Find the stimulated emission cross section given that $A_{21} = 6.56 \times 10^6$ sec^{-1}.

(h) If the density of excited neon atoms in state 1 ($J_1 = 2$) is 10^{10} cm^{-3}, how many excited atoms in the state 2 ($J_2 = 1$) are required to establish a small-signal gain coefficient of 5%/m?

9.2. Consider the atomic system shown below, where the A coefficients are given.



(a) What are the wavelengths of the various transitions? Express all transitions in terms of units in common use (eV, Hz, Å, nm, cm^{-1}).

(b) What is the lifetime of state 2? (Ans.: 14.3 ns.)

(c) What is the branching ratio of the $2 \rightarrow 1$ transition? (Ans.: 0.71.)

(d) Suppose that 10^{14} atoms/cm^3 are excited from state 0 to state 2 at $t = 0$ by some external mechanism. Describe the time evolution of the various populations in state 1 and state 2.

(e) Suppose that this external agent is strong enough to keep a steady state population of 10^{14} atoms/cm^3 in state 2.

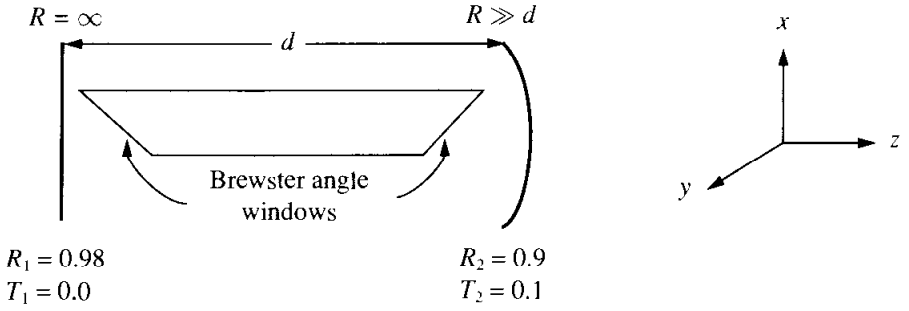
(1) How much power is required? (Ans.: 3.84 kW/cm^3.)

(2) What is the steady state population of state 1? (Ans.: 5×10^{13} cm^{-3}.)

(3) What power is radiated spontaneously in the $2 \rightarrow 1$ transition? (Ans.: 1.84 kW/cm^3.)

(4) What is the quantum efficiency of the $2 \rightarrow 1$ transition? (Ans.: 67.7%.)

9.3. An argon-ion laser at $\lambda_0 = 0.5145$ μm generates a TEM$_{0,0}$ Hermite–Gaussian beam with the laser cavity shown below.



(a) Specify the polarization of the optical electric field.

(b) If the minimum spot size is 1 mm and the *output* power is 8 W, what is the peak electric field inside the laser cavity incident on the spherical mirror?

9.4. Suppose we had a laser system with a stimulated emission cross section of 10^{-16} cm^2 and an initial inversion of 10^{14} cm^{-3} filling a cavity 50 cm long. The cavity uses

mirrors with reflectivities $R_2 = 0.9$ ($T_2 = 0.1$) and $R_1 = 0.98$ ($T_1 = 0$) and contains a nonsaturable loss of 2% per round trip. Assume homogeneous broadening.

(a) By what factor is the system above threshold (i.e., what is the ratio of $\Delta N/\Delta N_{\text{th}}$)?

(b) Estimate how long it would take for the laser intensity to increase (from noise) by a factor of 10^5. (Ignore saturation.)

(c) If the wavelength is 1 μm and the lifetime of the upper state is 1.0 μs, find the output intensity of this laser. (Assume $\tau_1 \sim 0$ and $dI/dz \sim 0$ within the laser cavity.)

9.5. The cavity shown in Fig. 9.4 with $R_1 = 0.95$ ($T_1 = 0$), $R_2 = 0.8$ ($T_2 = 0.2$), $l_g = 10$ cm, and $d = 15$ cm is used for a laser at $\lambda_0 = 7200$ Å. The stimulated emission cross section is 10^{-18} cm^2 and the saturation intensity is 20 kW/cm^2. (You may assume saturation according to the homogeneous law.)

(a) What is the photon lifetime of the passive cavity?

(b) What is the inversion density $[N_2 - (g_2/g_1)N_1]$ necessary to reach threshold for CW oscillation?

(c) Suppose that the cavity mode is "seeded" with a running wave of 10^8 photons/cm^2 (at $\lambda = 7200$ Å at $t = 0$); then an inversion of 2×10^{17} cm^{-3} is created instantaneously (by an external pump). *Estimate* the time interval required for the internal intensity of the laser to reach half of the saturation value. (NOTE: An approximate solution is desired here, but justify your approach with some physical reasoning.)

(d) Is the actual time interval longer or shorter than your estimate?

9.6. An optical amplifier of length $l_g = 10$ cm uses a homogeneously broadened transition centered at $\lambda_0 = 760$ nm, with a stimulated emission cross section of 2×10^{-20} cm^2, an upper-state lifetime of 1.54 ms, and a negligible lower-state lifetime. There is some residual loss of $\alpha = 0.01$ cm^{-1}, which is independent of intensity whereas the gain coefficient saturates according to the homogeneous law.

(a) Find the inversion necessary to obtain a net small-signal amplification of 6 dB.

(b) What is the value of the saturation intensity (in W/cm^2)?

(c) At what value of the input intensity will the net gain of this amplifier be 3 dB (i.e., $I_{\text{out}}/I_{\text{in}} = 2$)?

9.7. The following questions refer to the ring laser shown below. Assume that the active medium is an atomic gas (copper) with the energy levels as shown and that the transition is homogeneously broadened with a Lorentzian line width of 3.5 GHz. (State 1 has a long lifetime, and thus the copper vapor laser is not a CW system. However, that fact does not affect what follows.)

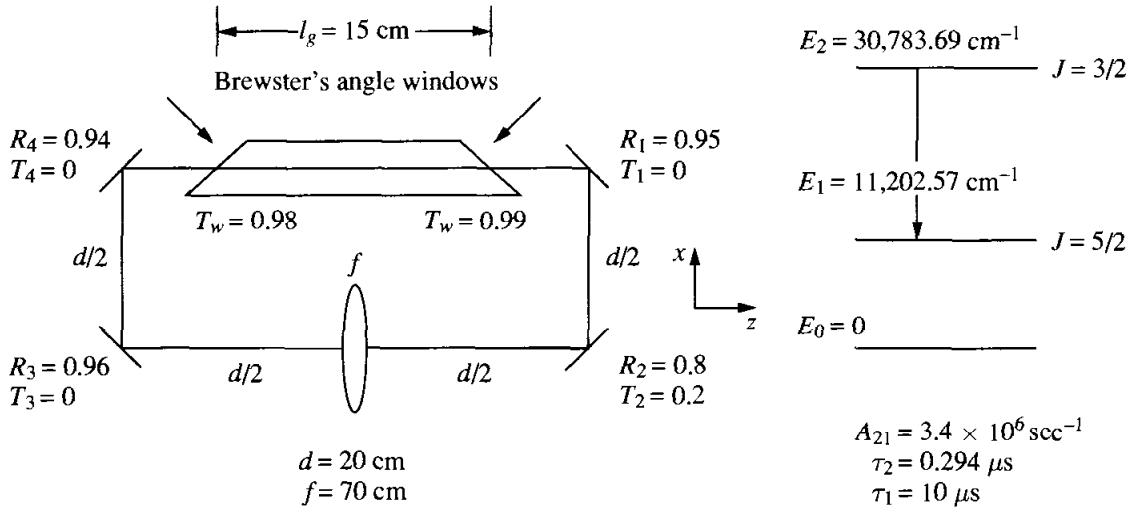
(a) Find the wavelength in nanometers. (Ans.: 510.7 nm.)

(b) What is the line width (FWHM) in angstrom units? (Ans.: 0.0304 Å.)

(c) What are the degeneracies of the states? (Ans.: $g_2 = 4$, $g_1 = 6$.)

(d) What is the stimulated emission cross section (in cm^2)? (Ans.: 6.42×10^{-14} cm^2.)

(e) What is the quantum efficiency of this laser? (Ans.: 63.6%.)



(f) What is the minimum required inversion density to reach threshold for CW oscillation? (Ans.: 4.23×10^{11} cm^{-3}.)

(g) Because of Brewster's angle windows the laser should be linearly polarized. Specify the direction of the optical electric field if the plane of the page is the xz plane and the output is along z. (Ans.: $\mathbf{E} = E_0\mathbf{a}_x$.)

(h) Is the cavity stable? Demonstrate your ability to analyze such cavities by

(1) Showing an equivalent lens waveguide starting at M_1 and proceeding counterclockwise around the cavity.

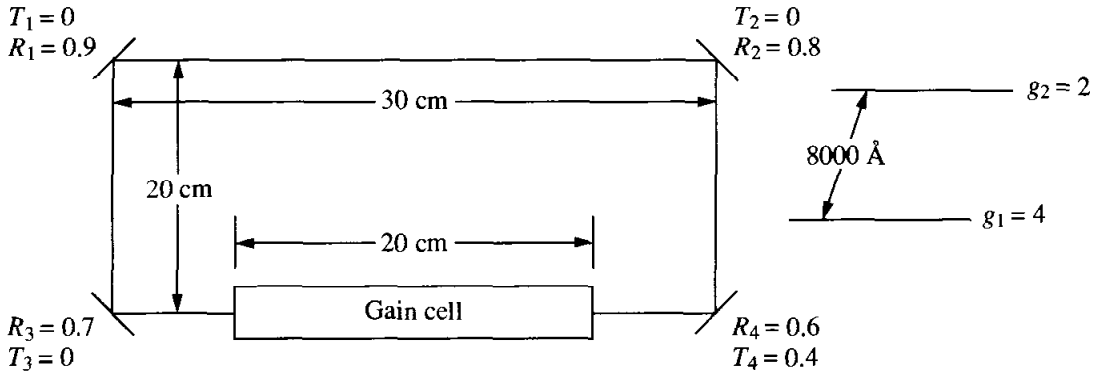
(2) Identifying a unit cell with the lens being the first element.

(3) Finding the $ABCD$ matrix for this unit cell in terms of d and f.

(4) Justifying your answer for the stability.

(i) Suppose the active medium acted as a *negative* gas lens (which it does). *Indicate* the place and order of matrix multiplication for the unit cell defined previously. (Do not expand, just indicate the operation.)

9.8. The following questions refer to the laser system shown below. Assume that only the clockwise wave is oscillating and that the transition is homogeneously broadened with $\Delta\nu_h = 2$ GHz and $A_{21} = 5 \times 10^5$ sec^{-1}.



(a) What is the stimulated emission cross section? (Ans.: 4.05×10^{-14} cm^2.)

(b) What is the population inversion necessary to reach threshold for oscillation in this cavity? (Ans.: 1.48×10^{12} cm^{-3}.)

(c) If the lifetime of the upper state (due to all spontaneous and collisional causes) is 0.9 μs and that of the lower state is 10 ns, what is the saturation intensity? (Ans.: 6.8 W/cm^2.)

(d) Express the line width in Å and in cm^{-1}. (Ans.: 0.043 Å and 0.067 cm^{-1}.)

(e) Assume that an external pumping agent is sufficient to create a small-signal gain coefficient of 0.1 cm^{-1} (which is higher than the threshold requirement). What is the output intensity? (Ans.: 3.13 W/cm^2.)

9.9. The following questions refer to a ring laser similar to that of Problem 9.8. The cavity specifications are as follows: $R_1 = 0.96$, $R_2 = 0.8$, $R_3 = 0.97$, $R_4 = 0.98$, $T_1 = T_3 = T_4 = 0$, $T_2 = 0.2$. The wavelength of the transition is 760 nm, which originates at a state at $E_2 = 3.2$ eV; the stimulated emission cross section is 2×10^{-20} cm^2; the upper-state lifetime is 1.54 ms and is pumped directly from the ground state at a rate R_{02}; and the lower-state lifetime is negligible (i.e., $N_1 = 0$ always). For the purpose of this problem, assume only the counterclockwise wave is oscillating.

(a) Find the upper-state density required to reach threshold, assuming steady-state operation.

(b) Estimate the pumping power (per unit volume) R_{02} required to reach threshold. (NOTE: Neglect stimulated emission for this part of the calculation.)

(c) Compute the output intensity (W/area) of the laser if the pumping rate was 1.5 times that required to reach threshold.

9.10. Consider the ring laser shown in Problem 9.8 with $R_1 = 0.9$, $R_2 = 0.7$, $R_3 = 0.6$, $R_4 = 0.95$, and $T_1 = T_2 = T_4 = 0$, $T_3 = 0.2$. Assume that the system is excited so that the small-signal gain coefficient is three times the threshold value in the absence of stimulated emission, that there is homogeneous broadening, and that laser oscillation is restricted to the counterclockwise wave near line center. Compute the *output* intensity if $I_s = 5$ W/cm^2 using the exact analysis and compare that with the high Q approximation method in which $dI/dz \sim 0$.

9.11. Consider an optically pumped laser system depicted in the diagram below. A strong optical pump tuned to line center of the $0 \rightarrow 2$ transition is incident on the sample from the side and promotes atoms from state 0 to 2. The atoms in state 2 can decay back to 0 by the indicated spontaneous emission ($A_{21} = 10^5$ sec^{-1}) and/or quenching processes ($k_{20} = 5 \times 10^6$ sec^{-1}), or to state 1 by spontaneous emission ($A_{20} = 10^6$ sec^{-1}) and/or stimulated emission The spacing $E_1 - E_0$ is much larger than kT, and thus we can neglect any initial population in state 1. Atoms in state 1 find their way back to 0 at a rate $\tau_1^{-1} = 10^7$ sec^{-1}. To make the problem simple, assume that the density of state 2 (or 1) is always much less than that of state 0, and thus we need not worry about the conservation of atoms (or a rate equation for state 0), that is, $N_0 =$ constant. Assume steady state and a homogeneous line shape for all transitions with $\Delta\nu_h = 10$ GHz.

(a) What is the *absorption* cross section for the pump wave at the frequency ν_{20}?

(b) What is the stimulated emission cross section for the $2 \rightarrow 1$ transition?

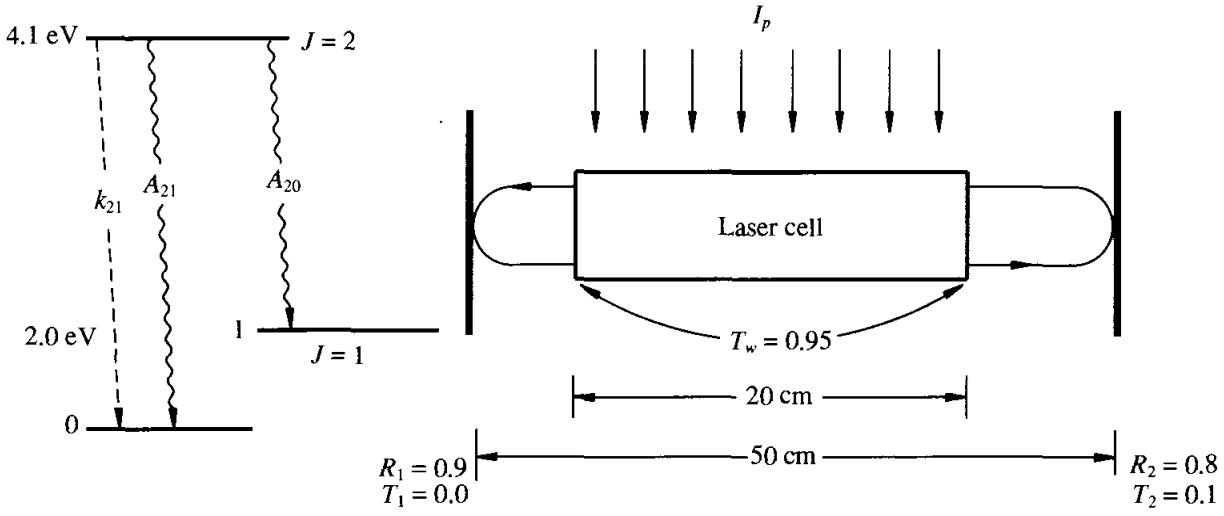
(c) What is the lifetime of state 2?

(d) If the optical pump intensity is weak, the $2 \rightarrow 1$ transition does not lase and stimulated emission at ν_{21} can be neglected. Under these circumstances, what is the ratio of the population densities N_2/N_1?

(e) What is the minimum inversion density, $N_2 - (g_2/g_1)N_1$ required to achieve oscillation on the $2 \rightarrow 1$ transition?

(f) Formulate the steady state rate equation for states 2 and 1 in terms of σ_p, I_p, σ_{21}, and I_{21} and the appropriate atomic relaxation rates. Use these equations to determine the intensity of the pump to reach threshold on the $2 \rightarrow 1$ transition.

(g) If the pump is 10 times the value required to reach threshold at ν_{21}, then the stimulated emission rate on the $2 \rightarrow 1$ transition overwhelms all other loss processes for state 2 and pumping processes for state 1. Use this fact to neglect the appropriate terms in the rate equations and thus predict the *output* intensity of this laser. (The approximations are intended to simplify matters, but if they confuse the issue, do it the long way. It is not that much more difficult.)



(h) If the pump intensity is very large, then the approximation that the ground state density does not change is not valid. The atoms are pumped to state 2, immediately stimulated to make a transition to 1, and then slowly relax back to 0. By using this scenario, predict the pump intensity at which the ground state density is depleted by 30%.

9.12. A certain laser medium is described by the following specifications: center of the transition = 2 eV; line width of the homogeneously broadened transition = 1.5 GHz; stimulated emission cross section at line center = 10^{-16} cm^2; lifetime of the upper state $\tau_2 = 1$ ms; the branching ratio of the upper state = 0.8; lifetime of lower state $\tau_1 = 10$ μs. Assume equal degeneracies of the states. The cavity is similar to that shown in Fig. 9.4 with $R_1 = 0.98 (T_1 = 0)$, $R_2 = 0.9 (T_2 = 0.1)$, a gain length $l_g = 30$ cm, and a cavity $d = 50$ cm.

(a) What is the threshold population inversion to achieve oscillation?

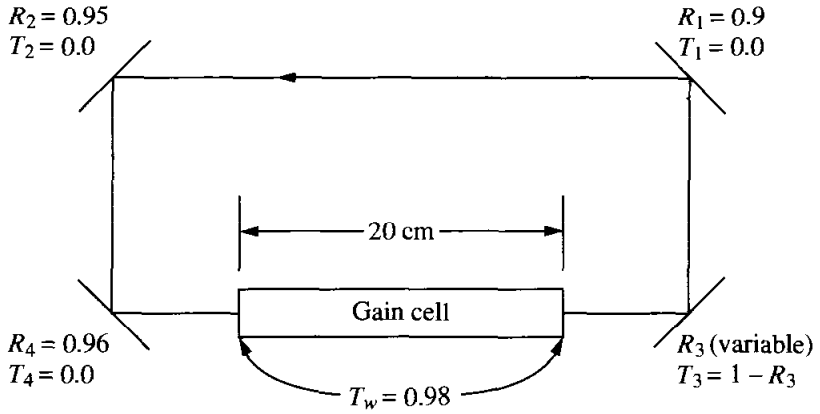
(b) If the system is pumped to 1.5 times its threshold value, explain how the laser selects the frequency (or frequencies) of oscillation.
(c) Estimate the intensity (in W/cm^2) generated by this laser when pumped to 1.5 times threshold. Use the assumption that $dI/dz \sim 0$.
(d) Use the Rigrod analysis of Sec. 9.2.3(b) to compute the output intensity and compare with (c).

9.13. Read and report on the paper "Short Pulse Amplification in the Presence of Absorption" by M. Tilleman and J. Jacob, *Appl. Phys. Lett. 50*(3), 121, 1987. Your report should include a discussion of
(a) The formulation and meaning of the various terms in Eq. 1 to 3 of the paper.
(b) The derivation of Eq. 4.
(c) If the optical bandwidth is 0.3 nm, they restrict their attention to pulses longer than 1 ps. Why?
(d) What is the asymptotic value of the energy from an amplifier in terms of the small-signal integrated gain to loss ratio γ_0/α?
(e) Recompute the information required to construct Figs. 2 and 3 for $\gamma_0/\alpha = 10$.

9.14. A homogeneously broadened ring laser with a small-signal gain coefficient of 0.2 cm^{-1} and a saturation intensity of 10 mW/cm^2 oscillates in the cavity shown in Fig. 9.2(a) with $R_1 = R_3 = R_4 = 0.98$, $T_1 = T_3 = T_4 = 0$, and $T_2 = 1 - R_2$. $l_d = 2.0$ cm and $l_g = 10$ cm. The loss coefficient of the "diode,"* which constrains oscillation to the counterclockwise direction, is 0.5 cm^{-1}.
(a) Assume that the oscillation frequency is close to the center of the transition and that the transmission of M_2 is variable. Plot the output power as a function of the transmission coefficient T_2 using the various theories of Sec. 9.2; that is,
(1) Use the self-consistent theory of Sec. 9.2.1 (use a solid curve)
(2) Assume that the losses are prorated over the gain length and that $dI/dz = 0$ within the gain cell as was done in the simplified analyses of Sec. 9.2.3(a) (use a dashed curve)
(b) If we use (9.2.9) to predict the optimum coupling, what percentage error in the output intensity results?

9.15. Assume that oscillation in the ring laser shown below is constrained to the counterclockwise direction by some unidirectional optical device. The gain coefficient of the active medium is 4.66%/cm, and there is a 2% scattering loss at the entrance and exit of the cell.
(a) What is the minimum value of the reflectivity R_3 that permits oscillation?
(b) Solve the problem self-consistently by demanding that the input to the gain cell be the feedback from the output. Plot the output intensity (normalized to the saturation value) as a function of the transmission coefficient of M_3.
(c) Solve for the output intensity by making the high Q approximation in which the saturated gain coefficient is set equal to the losses prorated over the gain length. Compare with the answer of (b).

*For information on such "diodes" see [13].



9.16. Redo the development of Sec. 9.4 on Q switching for the case $g_2 \neq g_1$; that is, show that (9.4.1), (9.4.12), and (9.4.16c) are modified to

$$\frac{dn}{dt} = -\left(1 + \frac{g_2}{g_1}\right)\frac{n}{n_{th}} N_p \tag{9.4.10b}$$

$$N_p(\max) = \frac{n_i - n_{th}}{(1 + g_2/g_1)} - \frac{n_{th}}{(1 + g_2/g_1)} \ln\left(\frac{n_i}{n_{th}}\right) \tag{9.4.12}$$

$$W_p = \frac{n_i h\nu}{(1 + g_2/g_1)} \eta_{xtn} \tag{9.4.16c}$$

where $n = [N_2 - (g_2/g_1)N_1]Al_g$.

All other equations and Fig. 9.13 remain the same.

9.17. Consider a ruby laser to be Q switched as shown in Fig. 9.11. The ruby rod has a cross-sectional area of 1 cm^2, is 15 cm in length, and is placed in a cavity 20 cm long. The Cr^{3+} doping density is 1.58×10^{19} atoms/cm^3 and the stimulated emission cross section is $\sigma = 1.27 \times 10^{-20}$ cm^2. The pumping agent creates an initial population of 10^{19} atoms/cm^3 in the upper laser state with negligible population in the pumped states (i.e., $N_3 \sim 0$). Assume that the pump band is centered at $\lambda = 4500$ Å; every atom pumped to state 3 relaxes to state 2; lifetime of state 2 is 3 ms; $g_1 = g_2$; and the mirror specifications are $R_1 = 0.95$, $T_1 = 0$, $R_2 = 0.7$, $T_2 = 1 - R_2$.

(a) How much pump power is required to maintain the population in state 2 at 10^{19}/cm^3?

(b) How much power is radiated spontaneously before the Q switch is operated?

(c) What is the peak output power of the Q switched pulse?

(d) How much energy is contained in the pulse?

(e) Estimate the pulse width by using (9.4.17) and the predictions of Fig. 9.14.

(f) Plot the peak power and energy contained in the Q switched pulse as the transmission coefficient T_2 is varied from 0 to 1.0.

9.18. Consider a ruby laser to be Q switched as shown in Fig. 9.11. The rod is 10 cm long and 1 cm^2 in area and is placed in a cavity 30 cm long. The doping density is 1.58×10^{19} atoms/cm^3, and the absorption cross section at 6943 Å is 1.27×10^{-20} cm^2. The pumping agent creates an initial population of 10^{19} atoms/cm^3 in the upper state N_2

with negligible population in pumped states ($N_3 \sim 0$). Assume that the pumping occurs by virtue of the absorption at $\lambda = 4500$ Å and that 90% of the atoms pumped to state 3 relax to state 2. The radiative lifetime of state 2 is 3 ms. The index of refraction of the ruby rod is 1.78 and that of the electro-optic switch is 2.35. The switch is 8 cm long. The mirror characteristics are $R_1 = 0.95$, $T_1 = 0$; $R_2 = 0.7$, and $T_2 = 0.3$. There is a 3% scattering loss at each air-to-ruby interface and a 5% scattering loss at each interface of the switch, and the open switch absorbs 2% of radiation passing through it. The degeneracies of the states are $g_2 = 2$, $g_1 = 4$.

(a) How much spontaneous power is radiated by the rod before the switch is opened?

(b) How much pump power is required to maintain the population of state 2 at 1×10^{19} cm^{-3}?

(c) What is the peak output power of the Q switched pulse?

(d) How much energy is contained in the output pulse?

(e) Estimate the pulse width using (9.4.12).

9.19. A laser at $\lambda = 1.0\ \mu$m, which is to be Q switched, has an initial population inversion at 10^{18} atoms, a factor of 3 above the threshold value required for CW oscillation in the high Q cavity. There are three types of photon losses in the cavity: coupling through one mirror to the outside world at a rate of 5×10^7 sec^{-1}, loss in the electro-optic switch at a rate of 10^7 sec^{-1}, and useless losses at the air-rod-switch interfaces at a rate of 2×10^7 sec^{-1}. The inversion after the Q switched pulse is over 5.9×10^{16} atoms.

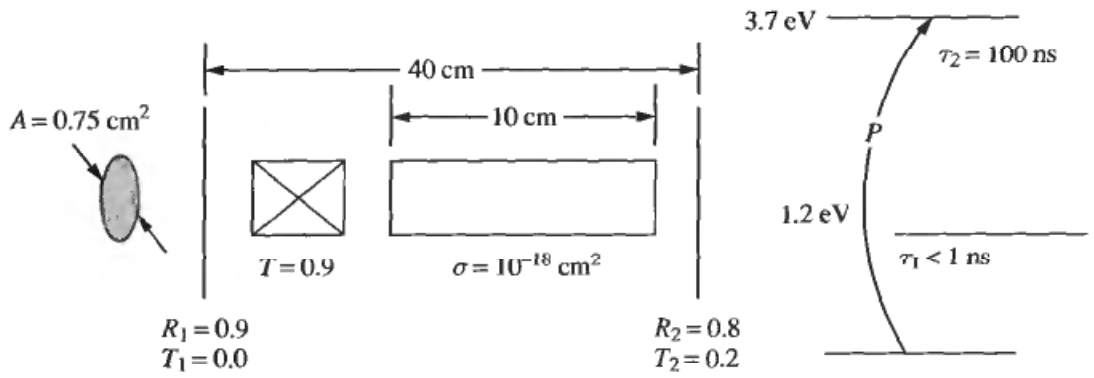
(a) Find the peak power out of this laser.

(b) Find the output energy in the Q switched pulse.

(c) Estimate the pulse width.

9.20. Let us reconsider Problem 9.17 and allow a 4% loss per pass at each air-ruby interface ($T = 0.96$). Let the Q switch element be 2 cm long, have an index of refraction of 1.5, and have a 2% residual loss per pass when open. Find the peak power in the output, the output energy, and the pulse width from Fig. 9.14 as well as (9.4.12).

9.21. The system shown in the diagram below is to be used to generate a Q switched pulse in the cavity shown there. If the system is pumped to four times the threshold value, find the output energy of the pulse. (CAUTION: This system is very different from ruby or the usual models chosen for Q switching analysis. The lower state decays very fast, even faster than a round trip transit time.)



9.22. An amplifier with a small-signal gain of 10 dB (i.e., $G_0 = 10$) is irradiated by a square pulse of 100-ns duration containing 50 mJ/cm^2 energy at the input. The saturation energy of the amplifier is 100 mJ/cm^2.

(a) Sketch the output intensity as a function of time.

(b) Identify the time at which the output intensity is half of its peak value.

(c) What is the maximum energy (per cm^2) that can be extracted from this amplifier?

9.23. This problem predicts the results of an experiment in which we irradiate an *absorption* cell with an optical pulse approximated by

$$I(t) = I_0 \sin^2\left(\frac{\pi t}{T}\right) \qquad (T = 2 \text{ ns})$$

The optical frequency is tuned to the center of the transition, which has an absorption cross section of 10^{-14} cm^2. The "small"-signal transmission through the cell is -30 dB. Graph the output intensity as a function of time for three values of energy contained in the pulse (a) $w = 2\,\mu$J/cm^2, (b) $w = 20\,\mu$J/cm^2, and (c) $w = 200\,\mu$J/cm^2. ($\lambda_0 = 5889$ Å).

9.24. A homogeneously broadened laser cell has the following specifications: a gain coefficient of $\frac{1}{2}$ %/cm (at the line center), a saturation intensity of 25 W/cm^2, a FWHM of 2 GHz, and length of 80 cm. The simple optical cavity of Sec. 9.2 is 100 cm long, with mirrors $R_1 = 0.98$ and R_2 variable; all other losses are negligible.

(a) Assume oscillation near the line center and plot the output intensity as a function of the coupling $(1 - R_2)$. Show two graphs, one using the simple theory and the other using the more exact computation.

(b) What is the coupling to achieve maximum power, and what is that power? Assume a spot size $w = 1$ cm.

(c) If we tune the mirror spacing by mounting R_1 on a translator, we can move the cavity mode across the gain profile. If this is done with the simple cavity shown in Fig. 9.2 with $R_2 = 0.7$, the output does *not* vary significantly. Explain why it does *not*.

9.25. Suppose that the amplitudes of the modes of a laser had a Lorentzian-like distribution of power such that

$$P(\omega) = P_0 \sum_{-\infty}^{+\infty} \left[\frac{(\Delta\omega/2)^2}{(n\omega_c)^2 + (\Delta\omega/2)^2}\right]^2 \delta[\omega - (\omega_0 + n\omega_c)]$$

where P_0 is the average power of the mode at the line center and $\Delta\omega$ is the FWHM of the oscillating spectrum. Compute the following quantities (as a function of $\Delta\omega$), and the power P_0.

(a) The total average power produced by this laser.

$$(\text{Ans.:} \quad \langle P \rangle = P_0(\pi/4)(\Delta\omega/\omega_c).)$$

(b) The peak power of the mode-locked pulse $[(\pi)(\Delta\omega/\omega_c)\langle P\rangle]$.

(c) The pulse width (FWHM) of the mode-locked pulse ($\Delta t = 2 \ln 2/\Delta\omega$).

9.26. Suppose that the power in the modes of a laser had a $[(\sin x)/x]^2$ distribution.

$$P(\omega) = P_0 \sum_n \frac{\sin^2(n\pi\omega_c/\Delta\omega)}{(n\pi\omega_c/\Delta\omega)^2} \delta[\omega - (\omega_0 + n\omega_c)]$$

where

$$-N < n < +N \qquad \text{and} \qquad \frac{\Delta\omega}{\omega_c} = N \qquad \omega_c = 2\pi \frac{c}{2d}$$

where P_0 is the average power of the mode *at the line center* and $\Delta\omega$ the half width of the spectrum of oscillation. Find the following quantities (as a function of average power and $\Delta\omega$).

(a) The peak power.

(b) The pulse width (FWHM) of the mode locked pulse.

(c) Sketch the time dependence of the mode locked pulse, labeling the peak power, the pulse width, and the pulse repetition frequency.

(NOTE: It is necessary to make various approximations to obtain an analytic answer to this problem. Do not hesitate to do so, but explain why the approximations are justified or not.)

9.27. Suppose the spectral distribution of power of a mode locked laser is given by

$$P(\omega) = P_0 \sum_{n=-\infty}^{+\infty} \operatorname{sech}^2\left(\frac{n\omega_c}{\Delta\omega}\right) \delta[\omega - (\omega_0 + n\omega_c)]$$

where n is the mode number measured from line center ω_0, P_0 is the power in the central mode, and $\Delta\omega$ is a characteristic width of the spectra. Assume $\Delta\omega/\omega_c = 10$, $\nu_c = \omega_c/2\pi = 100$ MHz, $P_0 \doteq 10$ mW.

(a) What is the average power?

(b) What is the peak power of the mode locked pulse?

(c) What is the width (full width at half maximum) of the mode locked pulse?

The approximate relations $P_{\text{peak}} = N\langle P\rangle$, $\Delta t = \tau_{RT}/N$ can be used as a check, but an exact answer is required here.

9.28. The envelope of the electric field in a laser cavity obeys the following differential equation (to be derived in Chapter 13):

$$\frac{d}{dt}\left(\frac{E}{E_s}\right) + \frac{1}{2\tau_p}\left[1 - (c/n)\tau_p\gamma(t)\right]\left(\frac{E}{E_s}\right) = 0 \qquad \text{(P9.28a)}$$

where spontaneous emission has been neglected, τ_p is photon lifetime; E_s is a field quantity related to the saturation intensity by $E_s^2/2\eta_0 = I_s = h\nu/\sigma\tau_2$, σ is the stimulated emission cross section, and $\gamma(t)$ is the time dependent gain coefficient for intensity. It is related to the population inversion by $\gamma(t) = N_2(t)\sigma$ if we neglect the lower state as is assumed here. The population of state 2 obeys the following (partially completed) differential equation:

$$\frac{dN_2}{dt} = P_2 - \frac{N_2}{\tau_2}[\qquad\qquad] \qquad \text{(P9.28b)}$$

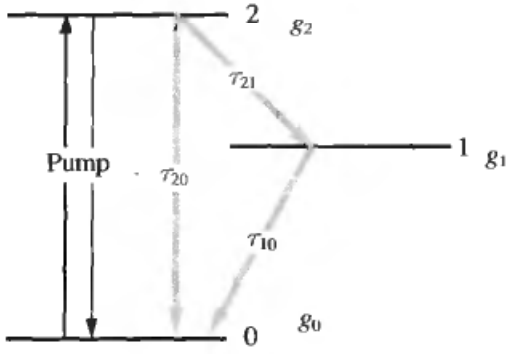
(a) What are the terms that appear in the empty bracket?

(b) Use and manipulate the above equation and the results of (a) to describe the steady state laser intensity versus pump rate.

(c) Derive a formula for the pump rate $P_2(th)$, which brings the laser to the threshold.

(d) Assume that the pump rate is m times the threshold value computed in (c). Derive a formula for the CW output *intensity* of the laser in terms of this factor m, the saturation intensity, and a transmission coefficient of the output coupling mirror.

9.29. Consider the atomic system shown below. Derive the criteria for optical gain on all transitions (i.e., $2 \rightarrow 0$, $2 \rightarrow 1$, or $1 \rightarrow 0$) presuming an optical pump tuned to $0 \rightarrow 2$. This criteria should involve the pump intensity, lifetime, degeneracy ratios, and the branching ratios. Find the limiting inversions for an infinite optical pump intensity. Neglect stimulated emission on $2 \rightarrow 1$ and $1 \rightarrow 0$ transitions.



9.30. The Rigrod analysis of Sec. 9.2.3(b) yielded an expression for the output intensity of a standing wave laser that was uniformly pumped along z, the optical axis. Now consider the case when the small signal gain coefficient varies as $\gamma_0 \exp[-\alpha z]$, which is appropriate for the case of the laser being optically pumped through mirror M_1.

(a) What is an expression for the value of γ_0 that brings the laser to threshold?

(b) What is an expression for the output intensity?

(c) Assume a pumping wavelength of 0.807 μm and lasing at 1.06 μm (see Fig. 10.5 for the energy level diagram for YAG), and find an expression for the efficiency of the laser. Assume a fraction η_p relaxes from the pumped state to ${}^4F_{3/2}$ and the output through M_2.

9.31. Consider a system in which there is an unsaturable *loss* in the gain medium, and thus the intensity obeys the following differential equation:

$$\frac{df}{dz} = \left\{ \frac{\gamma_0}{1+f} - \alpha \right\} f$$

for $0 < z < l_g$ where $f = I/I_s$.

(a) If the gain cell were used as an optical amplifier, what is the expression for the maximum extractable power (per unit of area). (If $\alpha = 0$, then the maximum is

given by (8.3.16). Your answer should have a correction factor that multiplies that value and that reduces to 1 when α vanishes.)

(b) Derive an expression for the output intensity through M_2 when this gain cell is used in a ring laser configuration similar to that shown in Fig. 9.2. Assume a perfect diode with $\alpha_{ccw} = 0$. (HINTS: (1) The differential equation can be integrated by using a partial fraction expansion. (2) It will be useful to abbreviate a few factors in terms of some easily interpreted expressions: $g_0 = (\gamma_0 - \alpha)l_g$, the net small signal integrated gain and $g_{th} = (\gamma_{th} - \alpha)l_g$, line integrated net gain at threshold, where γ_{th} is the value of γ_0, which brings the laser to threshold. (3) It will also be convenient to define an abbreviation, $\delta = \alpha/(\gamma_0 - \alpha) \ll 1$, but do not be hasty in applying the inequality. Wait until the final expression is obtained.)

9.32. Consider a ring laser constrained to oscillate in the counterclockwise direction in the manner indicated in Fig. 9.2 of the text. Instead of the diode, add a cell of length l_a containing a *saturable* absorber and thus the intensity there obeys the following differential equation:

$$\frac{1}{I} \cdot \frac{dI}{dz} = -\frac{\alpha_a}{1 + (I/I_{sa})} \qquad 0 < z < l_a$$

where I_{sa} is the saturation intensity for the absorber. Show a sketch that illustrates the solution for self-consistent intensities I_1 and I_2 (as labeled on Fig. 9.2) inside the cavity. The following is offered for a guideline.

(1) Insert the saturable absorber in place of the diode. Call its input I_3 and the output I_4.

(2) Relate I_3 and I_4 to I_2 and I_1:

(3) Derive two simultaneous (nonlinear) equations for the intensities I_1 and I_2.

(4) Construct a graphical solution for I_2 and I_1 by plotting I_2 as a function of I_1 under circumstances representing extreme cases. Start by removing the absorption cell (i.e., let $\alpha_a = 0$).

(a) If the gain cell did not saturate, what would be the I_2 versus I_1 "curve"?

(b) What is the curve that expresses the self-consistency relationship between I_1 and I_2?

(c) Show the graphical solution for the intensities I_1 and I_2. Indicate how saturation of the gain cell controls the approach to the curve I_2 versus I_1.

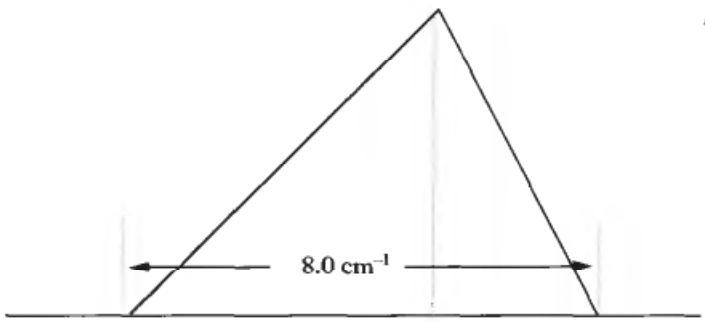
(5) Now reinsert the absorption cell.

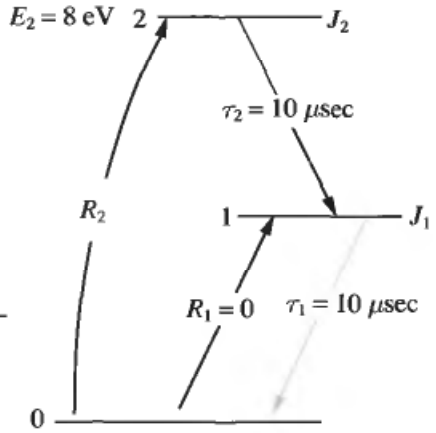
(a) Repeat part 4(a) assuming $I_{sa} = \infty$ (i.e., one with a linear absorption coefficient). What is the maximum value of $\alpha_a l_a$ permitted in terms of $\gamma_0 l_g$ and other passive losses?

(b) Repeat part 4(c) and allow for a finite value of I_{sa}. Indicate how the self-consistent solution would change as the absorber or gain saturation intensity changes.

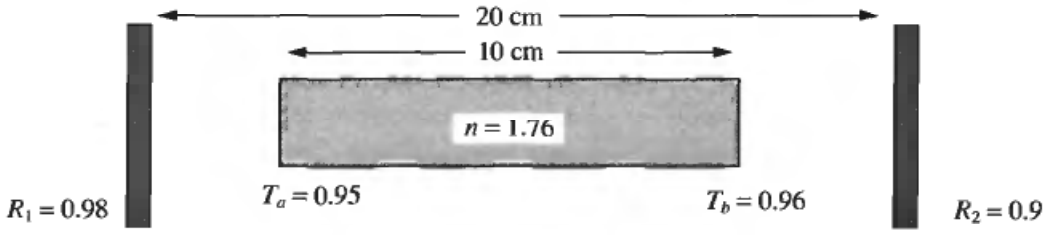
9.33. The following parameters describe the laser system shown below: The spontaneous emission profile can be approximated by the triangular shape shown, only state 2 is pumped at a rate of $R(\text{cm}^{-3}\ \text{sec}^{-1})$; state 2 decays to 1 by radiation and by collisions

with lifetime of 10 μs, which is also the lifetime for state 1; the Einstein coefficient for the $2 \rightarrow 1$ transition is 5×10^4 sec^{-1}; $J_2 = 1$ and $J_1 = 2$; and the active media has an index of refraction of 1.76. Assume steady state.

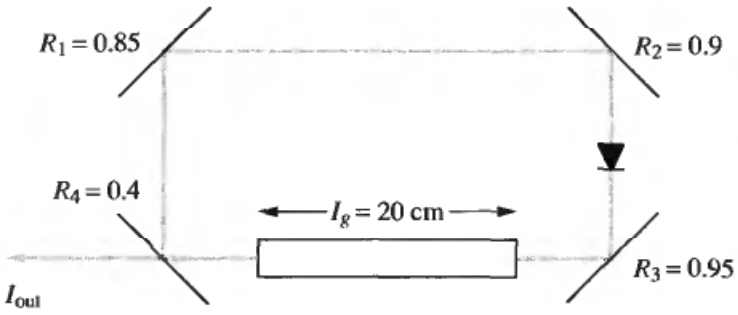




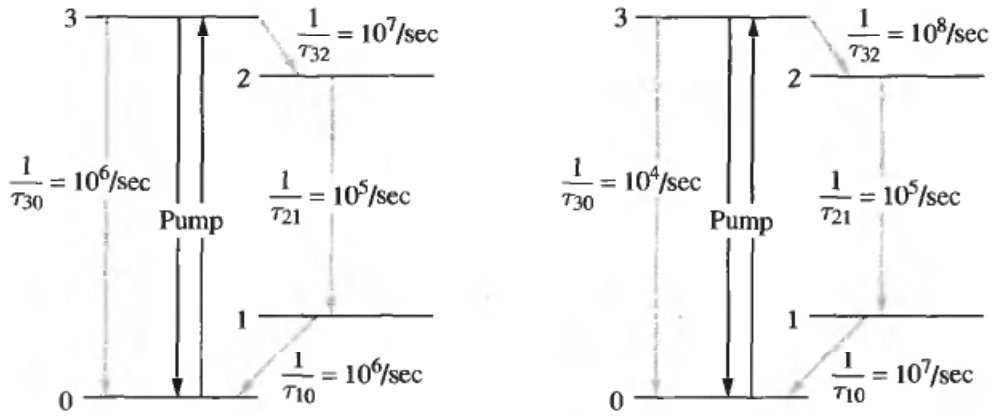
(a) What is the quantum efficiency of this laser system?
(b) What is the stimulated emission cross section (at line center)?
(c) What is the minimum pumping power (per unit of volume) required to bring the laser to threshold in the cavity shown below?
(d) Derive an expression for the gain coefficient, $[1/I_\nu][dI_\nu/dz]$, which is valid for large intensities in terms of the pumping rate R_2, the lifetimes, the degeneracies, and the stimulated emission cross section σ.



9.34. The small signal gain of the amplifying medium shown below is 6 dB (i.e., $G_0 = 4$) at 1.05 μm. The medium has a stimulated emission cross section of 10^{-17} cm^2, an upper state lifetime of 500 μs, and 10 ns for the lower. There is a 2% loss per surface at the end surfaces of the gain cell. The optical diode can be considered as perfect. Compute the output intensity.



9.35. The two atomic systems shown below have been proposed as candidates for Q switching. Both are optically pumped from the ground state with an external laser that can be considered as infinitely strong. Which one will yield the most energy in the Q switched pulse if we presume that the cavity, the Q switch element, and stimulated emission cross section of the $2 \rightarrow 1$ transition is the same? (NOTE: The answer does not depend upon mathematics. Rather it requires an essay explaining your reasoning.)



9.36. A laser generates a mode-locked spectrum whose electric field can be approximated by

$$E(t) = E_0 \exp(j\omega_0 t) \sum \frac{\sin[n\omega_c/\Delta\omega] \exp(jn\omega_c t)}{[n\omega_c/\Delta\omega]}$$

where $E_0 = 100$ V/cm, $\omega_c = 2\pi(c/2d)$, $c/2d = 200$ MHz, $\Delta\omega = 2\pi\Delta\nu$, $\Delta\nu = 5$ GHz, $\lambda_0 = 5145$ Å, and the summation goes over all positive and negative integers.

(a) How long is the cavity?

(b) What is the period T between pulses?

(c) Sketch the time behavior of the *intensity* of the pulses on a graph showing at least two periods. Label the peak intensity in watts/cm^2, the pulse width in ns, and the repetition period in ns.

(d) If this mode locking is accomplished by the use of a time dependent shutter with a power transmission coefficient given by $\cos^2(\pi t/T)$, what fraction of the pulse is lost per passage through the shutter?

9.37. The gain medium of the ring laser shown in Fig. 9.2 is characterized by the following set of parameters: homogeneously broadened with a line width of 5.0 cm^{-1}, an upper state lifetime of 255 μs (negligible lifetime for state 1), an A_{21} coefficient of 500 s^{-1}, the index of refraction of 1.81633, and a wavelength of 1.06417 μm, $R_1 = 0.97$, $R_2 = 0.5$ ($T_2 = 0.5$), $R_3 = 0.96$, $R_4 = 0.95$, and a perfect optical diode.

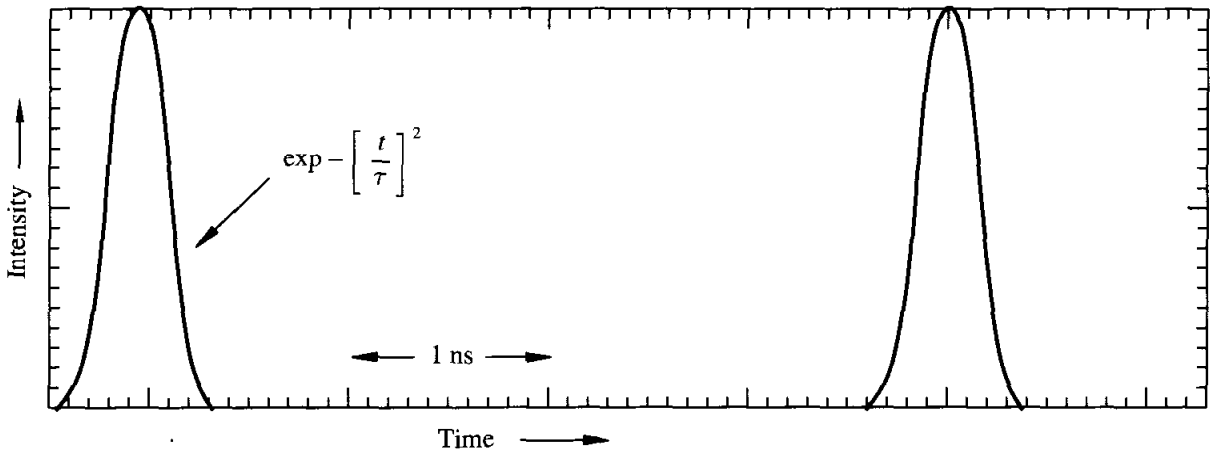
(a) What is the stimulated emission cross section and saturation intensity for this transition?

(b) What is the gain coefficient required to bring this laser to threshold?

(c) If the stimulated emission cross section and saturation intensity of this medium were 3×10^{-19} cm^2 and 2.0 kW/cm^2 [which may be different from that found in

(a)], compute the output intensity if the pumping were three times the threshold value. Assume equal degeneracies and counterclockwise oscillation.

9.38. The output of a mode locked gas laser is described on the graph shown below.



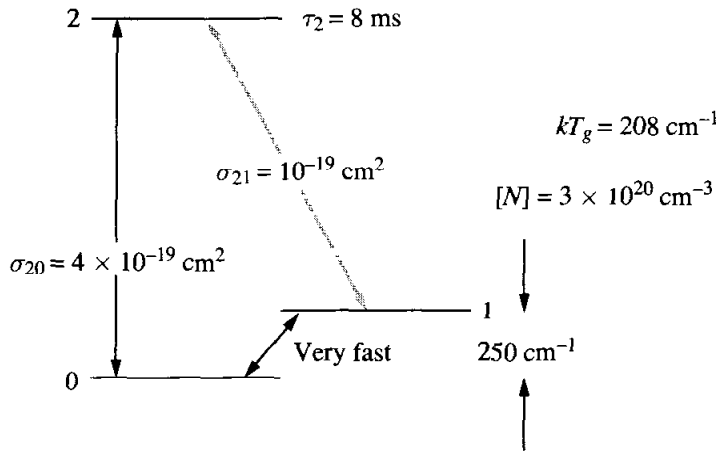
(a) What is the length of the laser used to generate this output?

(b) What is the full-width half-maximum of the laser pulse, and how does that number correlate to the fitting parameter τ? (A graphical solution is desired.)

(c) What is the ratio of the peak to average power of this laser?

(d) Give an expression for the spectral distribution of the electric field of this laser and also that of the laser power. Sketch the envelope of each and indicate the mode spacing.

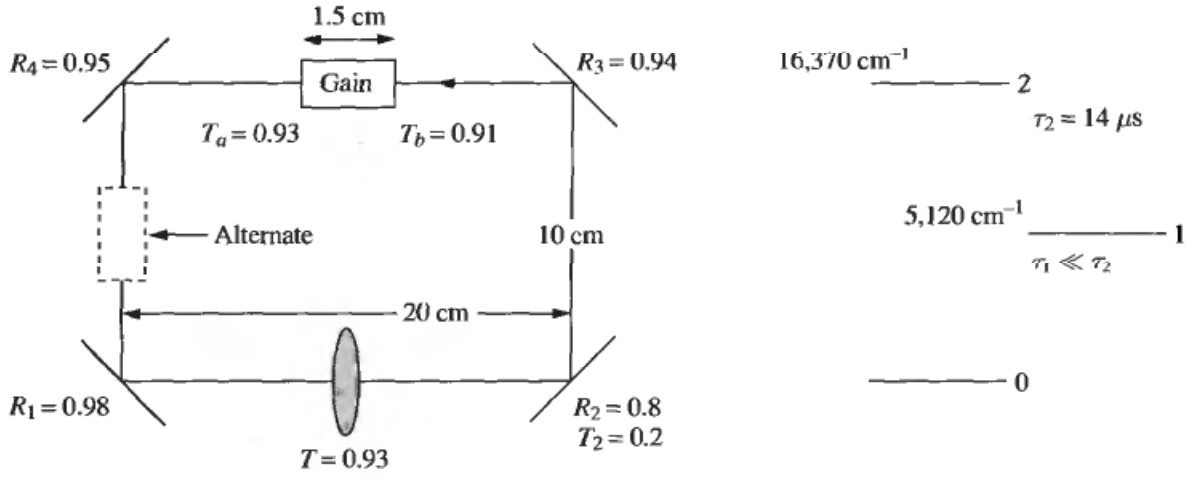
9.39. What follows is a model of a laser that is a mixture of a three- and four-level system. It is excited by absorbing optical radiation on the $0 \rightarrow 2$ transition and if the pumping is intense enough there is gain created on the $2 \rightarrow 1$ transition. Transitions between 1 and 0 are very fast, and, because $E_1 - E_0$ is comparable to kT, the populations in those states are distributed according to the Boltzmann law irrespective of what is happening elsewhere. You may assume that $E_2 \gg kT$, but in all cases the sum of the populations must equal the doping density of 3×10^{20} cm^{-3}. Assume equal degeneracies for all states.



(a) If the system were *not* excited, then what is the absorption coefficient on the $2 \rightarrow 1$ transition?

(b) If the pump intensity were infinite, then what is the maximum attainable small signal gain coefficient on the $2 \rightarrow 1$ transition?

9.40. Consider the ring laser shown below in a *stable* cavity with an active medium 1.5 cm long and a stimulated emission cross section of 10^{-18} cm^2. The upper state lifetime is 14 μs whereas that of the lower state is much shorter. The energy level diagram is shown on the side of the laser cavity.

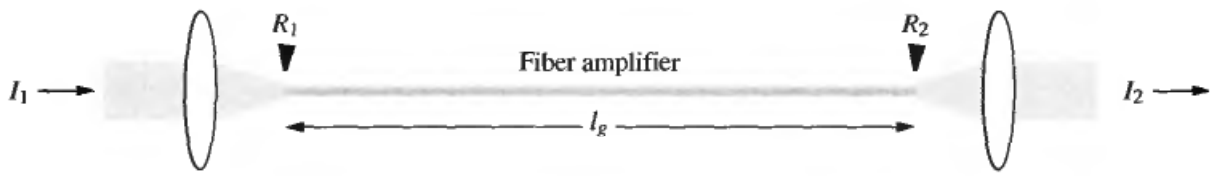


(a) What is the minimum population inversion density needed to reach threshold?

(b) What is the saturation intensity? What is the wavelength?

(c) If $\gamma_0 = 0.5$ cm^{-1}, compute the output intensity (through M_2) presuming that oscillation is constrained to the counterclockwise direction.

9.41. The Erbium-doped fiber amplifier shown below is to be used as an amplifier in an optical communication link. Unfortunately, there are always some residual reflections R_1 and R_2, and, thus, if the gain were too high, then the amplifier will oscillate and the output will not be a faithful replica of the input. If the input signal were zero, then what is an expression for the maximum small signal single-pass gain G_0 that can be built into this amplifier. Evaluate for $R_1 = R_2 = 0.025$.

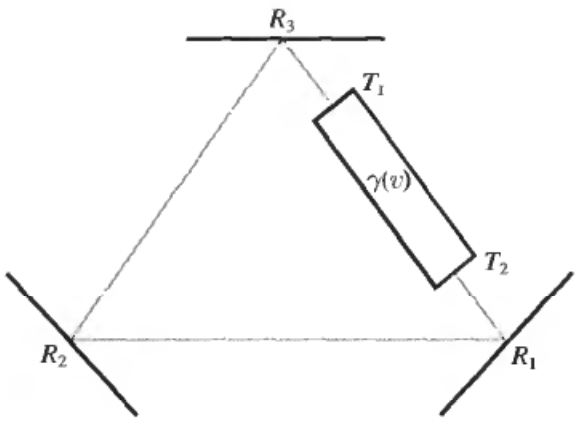


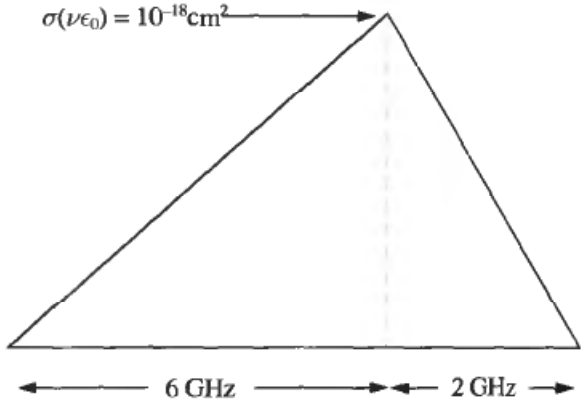
9.42. Consider the ring laser geometry shown below in which 80% of the photons survive a round trip in the passive cavity. The variation of the stimulated emission cross section with frequency is shown at the right with a peak value of 10^{-18} cm^2.

(a) What is the value of the line shape and function at $\nu = \nu_0$?

(b) What must be the population inversion to start oscillation?

(c) If the small signal gain coefficient were three times the threshold value, over what bandwidth can this system lase?



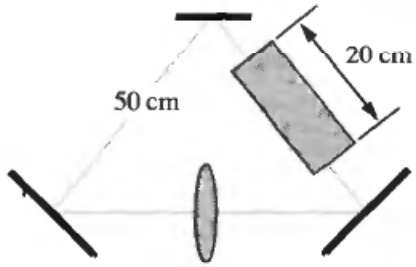


9.43. Consider the laser shown below. The medium is inhomogeneously broadened described by a Lorentzian line shape with $\Delta\nu_h = 3$ GHz with a stimulated emission cross section of 10^{-14} cm^2 at line center. In the *passive* cavity, 97.5% of the photons survive a round trip.

(a) What is the photon lifetime?

(b) What should be the population inversion to bring this laser to threshold?

(c) If the pumping established a small signal gain coefficient which was five times the threshold value, how many $\text{TEM}_{0,0}$ modes would be above threshold? (HINT: Show a diagram and the problem becomes trivial.)



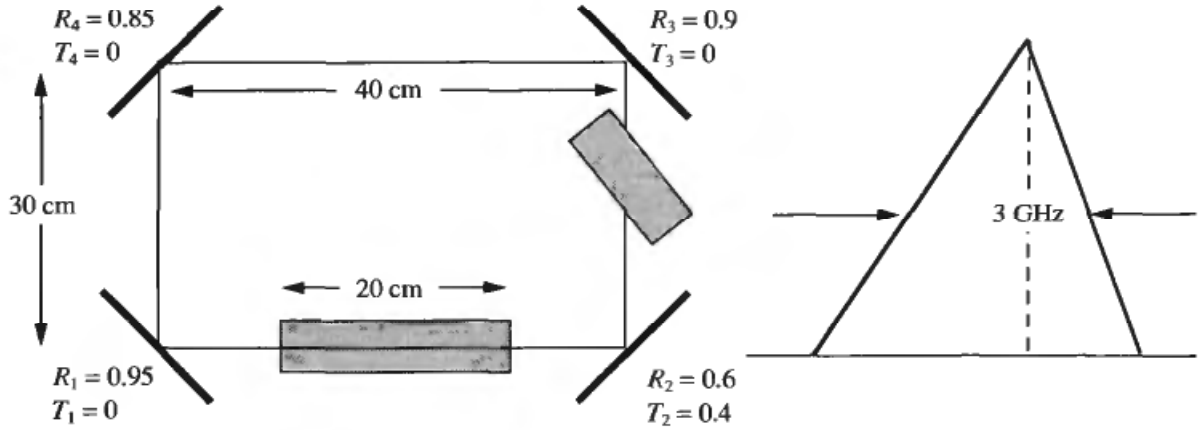
9.44. Consider the ring laser shown below operating on a homogeneous transition at $\lambda_0 = 6000$ Å with the characteristic parameters $A_{21} = 6 \times 10^4$ sec^{-1}, the line shape as shown, no loss at the air-gain interfaces, an upper state lifetime of 3.9 μs, and the *small-signal inversion* $[N_2^0 - (g_2/g_1)N_1^0] = 5 \times 10^{+13}$ cm^{-3}. Assume counterclockwise oscillation.

(a) Evaluate the stimulated emission cross section, saturation intensity I_s, and the threshold inversion density.

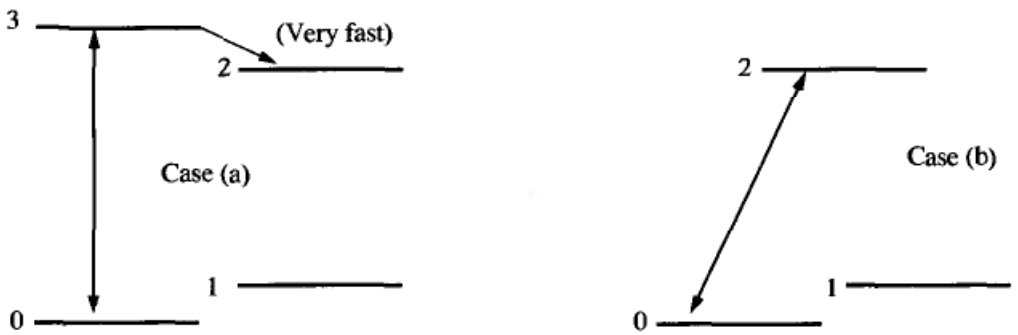
(b) If this were an inhomogeneously broadened laser, then how many $\text{TEM}_{0,0,q}$ modes would be above threshold?

(c) If the intensity were $3 \times I_s$, what would be the inversion?

(d) Evaluate the CW output power through M_2.



9.45. The purpose of this problem is to evaluate the advantages and disadvantages of the two systems shown below being optically pumped from 0. The only difference between the two is that the pumping to state 2 is via the route from $0 \rightarrow 3$ with a very fast relaxation to 2. Assume that the stimulated emission cross section for $2 \rightarrow 1$ are the same, τ_2 is the same, state 1 remains in Boltzmann equilibrium with state 0 under any and all circumstances, $g = 1$ for all states, equal absorption cross sections for the pumping route, steady state, and a common density $[N]$ of active atoms. If $I_p = 0$, then the densities in states 2 or 3 can be considered to be zero. Obviously, it takes a higher energy photon to create the upper state in case (a) than in case (b) and hence the quantum efficiency of case (a) is less than that of case (b).



But the ease of creating an inversion is usually much more important than the quantum efficiency. The task is to compare inversion for the two cases in a graphical format, where the independent variable (x axis) should be measured in units of I_p/I_{sp} : $I_{sp} = h\nu_{30}/\sigma_p\tau_2$, a convenient abbreviation for the many constants; I_p is pump intensity, and $h\nu_{30} = E_3 - E_0$.

(a) If $I_p = 0$, then there is absorption on the $2 \rightarrow 1$ transition. Indicate its value on the graph.

(b) Indicate the ratio I_p/I_{sp} for optical transparency for the two cases by using the following numerical values. $E_3 = 2.1$ eV, $E_2 = 1.9$ eV, $E_1 = 315$ cm^{-1}, $kT = 208$ cm$^{-1} = 0.0259$ eV.

(c) If $I_p/I_{sp} \gg 1$, then each case approaches a different asymptotic limit for $\gamma/(\sigma[N])$. Show those limits by dashed lines, and indicate their numerical values.

9.46. Some mode locked lasers produce a "chirped" pulse whose electric field is given by

$$e(t) = E_0 \exp\left[j(\omega_0 + \Delta\omega t/T)t\right] \exp\left[-(t^2/2\tau_p^2)\right] + \text{c.c.} \qquad -d/c < t < d/c$$

$e(t + 2d/c) = e(t)$ (i.e., the pulse repeats every $T = 2d/c$ second) where τ_p is characteristic of Gaussian pulse shape $\ll d/c$ and $\Delta\omega$ is a "chirp" parameter. The instantaneous frequency of any wave is the time derivative of its phase

$$\frac{d\left[\phi + \omega_0 t + \Delta\omega t^2/T\right]}{dt} = \omega_0 + 2(\Delta\omega/T)t$$

indicating that the lower frequencies (longer wavelengths) appear on the leading edge of the pulse ($t < 0$), whereas the higher frequencies (shorter wavelengths) appear on the trailing edge ($t > 0$) for $\Delta\omega$ positive.

(a) Plot the relative intensity (as a function of time) represented by this field. Does the "chirp" affect the FWHM?

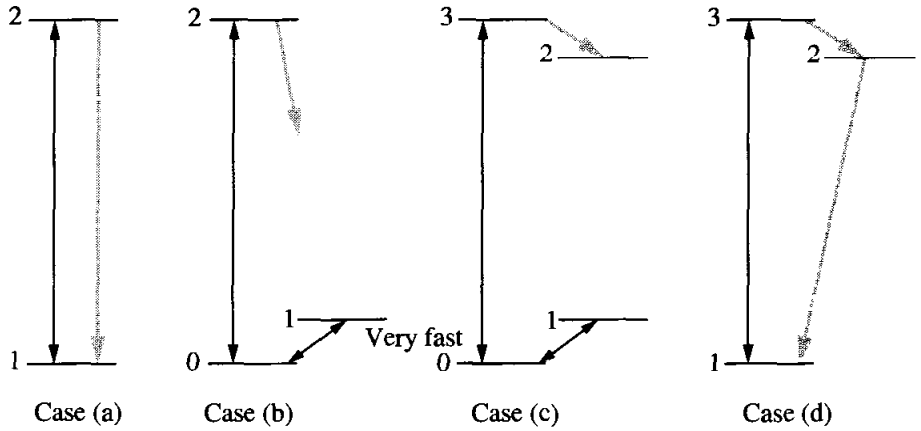
(b) What is the envelope of the spectral distribution of the electric field (i.e., in the angular frequency domain). Express your answer as a magnitude times an exponential with a phase (i.e., $|E(\omega)| \exp[-j\theta(\omega)]$).

(c) The "chirp" affects the FWHM in the frequency domain. What is the ratio of the spectral width to that without the "chirp"?

(d) Plot the relative distribution of intensity in the frequency domain, and label the FWHM in the angular frequency domain. Do this for two cases: $\Delta\omega = 0$ and $\Delta\omega\tau_p = 1$.

(e) What is an expression for the time-frequency product $(\text{FWHM})_\nu \cdot (\text{FWHM})_t$?

9.47. Four possible "laser" systems are shown below using *optical pumping* between the lowest (ground state) and the highest upper state with the object of obtaining a popu-



lation inversion on the $2 \rightarrow 1$ transition. The problem is to derive a formula relating the pump intensity [normalized to the saturation value $I_{sp} = h\nu_p/\sigma_p\tau_u$, $u = 2$ for

cases (a) and (b), 3 for cases (c) and (d)], which makes the laser transition $2 \rightarrow 1$ optically transparent (i.e., $N_2 - g_2/g_1 N_1 = 0$). In cases (b) and (c), assume that there is some internal mechanism that maintains a Boltzmann distribution between states 1 and 0; that is,

$$\frac{N_1}{N_0} = \frac{g_1}{g_0} \exp - \frac{\Delta E_{10}}{kT}$$

with a similar relation between 3 and 2. Assume $\Delta E_{32} \gg kT$ for cases (b) and (c). Assume steady state [i.e., $d(\ )/dt = 0$], separate degeneracies for each state and lifetimes as appropriate. The double-headed arrows represent the pump, and the dotted lines represent relaxation.

For case (a), compute the ratio N_2/N_1 in the limit of $I_p/I_{sp} \rightarrow \infty$. What should be the ratio I_p/I_{sp} to obtain 75% of this limit? For cases (b)–(d), show the limit for $\Delta E_{10}/kT \gg 1$ and/or any favorable lifetime ratios.

9.48. Modify (9.6.1) to (9.6.7) to account for the degeneracies (g_2, g_1) of the two states. [Ans.: (9.6.3b) becomes $w_s = h\nu/[\sigma(1 + g_2/g_1)]$, and some of the 2 are changed (Eq. 9.6.8) and the rest of Sec. 9.6 are independent of (g_2, g_1) provided γ is interpreted as $[N_2 - (g_2, g_1)N_1\sigma]$ per usual.]

REFERENCES AND SUGGESTED READINGS

1. A. G. Fox and T. Li, "Resonant Modes in a Maser Interferometer," *Bell Syst. Tech. J. 40*, 453–488, Mar. 1961.
2. H. Kogelnik and T. Li, "Laser Beams and Resonators," *Appl. Opt. 5*, 1550–1567, Oct. 1966.
3. A. E. Siegman, "Unstable Optical Resonator for Laser Applications," *Proc. IEEE 53*, 277–287, Mar. 1965.
4. A. E. Siegman, *Introduction to Lasers and Masers* (New York: McGraw-Hill Company, 1971).
5. W. R. Bennett, Jr., "Gaseous Optical Masers," *Appl. Opt., Suppl. Opt. Masers*, 24–61, 1962.
6. W. R. Bennett, Jr., "Inversion Mechanisms in Gas Lasers," *Appl. Opt., Suppl. 2 Chem. Lasers*, 3–33, 1965.
7. W. S. C. Chang, *Principles of Quantum Electronics* (Reading, Mass.: Addison–Wesley, 1969). This book has an extensive bibliography.
8. W. W. Rigrod, "Saturation Effects in High-Gain Lasers," *J. Appl. Phys. 36*, 2487, 1965.
9. W. G. Wagner and B. A. Lengyel, "Evolution of the Giant Pulse in a Laser," *J. Appl. Phys. 34*, 2042, 1963.
10. R. W. Hellworth, "Theory of the Pulsation of Fluorescent Light from Ruby," *Phys. Rev. Lett. 6*, 9, 1961.
11. E. O. Schultz–DuBois, "Pulse Sharpening and Gain Saturation in Traveling-Wave Masers," *Bell Syst. Tech. J. 43*, 625, 1964.
12. "FM and AM Modelocking of the Homogeneous Laser Part 1, Theory; Part 2, Experiment," *IEEE J. Quant. Electron. QE-6*, 694, 1970. See also Siegman and Kuezenga, *Appl. Phys. Lett. 14*, 181, 1969.

13. T. F. Johnston and W. Proffitt, "Design and Performance of a Broad-band Optical Diode to Enforce One-direction Traveling Wave Operation of a Ring Laser," *IEEE J. Quant. Electron. QE-16*, 483, 1980.
14. D. L. Huestis, see the 5 volume set, *Applied and Atomic Collision Physics*, Eds. Massey, McDaniel, and Beterson (New York: Academic Press, 1982). Parts of the article by D. L. Huestis (Vol. 3, Chapter 1, p. 1, Eds. McDaniel and Nighan) entitled *Introduction and Overview*, were the basis for sec. 9.2.4.
15. A. E. Siegman, *Lasers*, (Mill Valley, Calif.: University Science Books, 1986), Chaps. 5–6.
16. A. Yariv, *Quantum Electronics* (New York: John Wiley & Sons, 1975).
17. Peter W. Smith, "Mode Locking of Lasers," *Proc. of IEEE 58*, 1342–1356, 1970. This contains over 150 additional references.
18. R. Spiegel, *Mathematical Handbook of Formulas and Tables*, Schaum's Outline Series (New York: McGraw–Hill, 1968).
19. L. M. Frantz and J. S. Nodvik, "Theory of Pulse Propagation in a Laser Amplifier," *J. Appl. Phys. 34*, 2346–2349, 1963.
20. M. M. Tilleman and J. H. Jacob, "Short Pulse Amplification in the Presence of Absorption," *Appl. Phys. Lett. 50*, 121–123, 1987.
21. R. L. Fork, I. Greene, and C. V. Shank, "Generation of Optical Pulses Shorter than 0.1 ps by Colliding Pulse Mode Locking," *Appl. Phys. Lett. 38*, 671–672, 1991.
22. R. L. Fork, C. V. Shank, R. Yen, and C. A. Hirlimann, "Femtosecond Optical Pulses," *IEEE J. Quant. Electron., QE-19*, 500–505, 1983.
23. G. H. C. New, "Mode-Locking of Quasi-Continuous Lasers," *Opt. Comm. 6*, 188–192, 1972.
24. G. H. C. New, "Pulse Evolution in Mode-Locked Quasi-Continuous Lasers," *IEEE J. Quant. Electron. QE-10*, 115–124, 1974.
25. E. M. Garmire and A. Yariv, "Laser Mode-Locking with Saturable Absorbers," *IEEE J. Quant. Electron. QE-3*, 222–226, 1967.
26. H. A. Haus, "A Theory of Forced Mode Locking," *IEEE J. Quant. Electron. QE-11*, 323–330, 1975.
27. C. V. Shank, R. L. Fork, R. Yen, and R. H. Stolen, "Compression of Femtosecond Optical Pulses," *Appl. Phys. Lett. 40*, 761–763, 1982.
28. N. J. Deifo, "Ultrashort Pulse Propagation in Saturable Media: A Simple Physical Model," *IEEE J. Quant. Electron. QE-19*, 511–519, 1983.
29. R. L Fork, C. H. Brito Cruz, P. C. Becker, and C. V. Shank. "Compression of Optical Pulses to Six Femtoseconds by Using Cubic Phase Compensation," *Opt. Lett. 12*, 483–485, 1987.
30. H. A. Haus, "Theory of Mode Locking with a Slow Saturable Absorber," *IEEE J. Quant. Electron. QE-11*, 736–746, 1975.
31. M. S. Stix and E. P. Ippen, "Pulse Shaping in Passively Mode-Locked Ring Dye Lasers," *IEEE J. Quant. Electron. QE-19*, 520–525, 1983.
32. R. L. Fork, O. E. Martinez, and J. P. Gordon, "Negative Dispersion Using Pairs of Prisms," *Opt. Lett. 9*, 150–152, 1984.
33. W. H. Knox, "Femtosecond Optical Pulse Amplification," *IEEE J. Quant. Electron. 24*, 388–397, 1988.
34. J. Wang, "Analysis of Passive Additive-Pulse Mode Locking with Eigenmode Theory," *IEEE J. Quant. Electron. 28*, 562–568, 1992.

35. E. P. Ippen, H. A. Haus, and L. Y. Liu, "Additive Pulse Mode Locking," *J. Opt. Soc. Am. B 6*, 1736–1745, 1989.

36. D. E. Spence, P. N. Kean, and W. Sibbett, "60-fsec Pulse Generation from a Self-Mode-Locked Ti:Sapphire Laser," *Opt. Lett. 16*, 42–44, 1991.

37. P. M. W. French, J. A. R. Williams, and J. R. Taylor, "Femtosecond Pulse Generation from a Titanium-Doped Sapphire Laser Using Nonlinear External Cavity Feedback," *Opt. Lett. 14*, 686–688, 1989.

38. J. Mark, L. Y. Liu, K. L. Hall, H. A. Haus, and E. P. Ippen, "Femtosecond Pulse Generation in a Laser with a Nonlinear External Resonator," *Opt. Lett. 14*, 48–50, 1989.

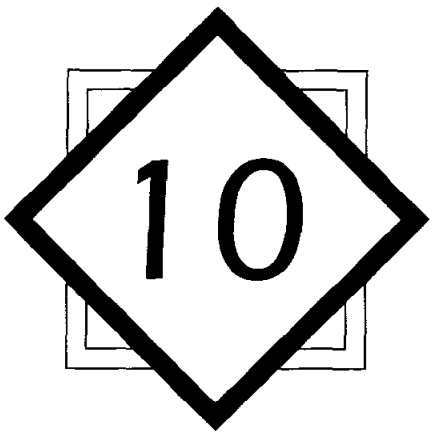
39. J. Goodberlet, J. Wang, J. H. Fujimoto, and P. A. Schulz, "Femtosecond Passively Mode-locked Ti:Al_2O_3 Laser with a Nonlinear External Cavity," *Opt. Lett. 14*, 1125–1127, 1989.

40. G. P. A. Malcolm, P. F. Curley, and A. I. Ferguson, "Additive-Pulse Mode Locking of a Diode-Pumped Nd:YLF Laser," *Opt. Lett. 15*, 1303–1305, 1990.

41. L. Y. Liu, J. M Huxley, E. P. Ippen, and H. A. Haus, "Self-Starting Additive-Pulse Mode Locking of a ND:YAG Laser," *Opt. Lett. 15*, 553–555, 1990.

42. J. Goodberlet, J. Jacobson, and J. G. Fujimoto, "Self-Starting Additive-Pulse Mode-Locked Diode-Pumped Nd:YAG Laser," *Opt. Lett. 15*, 504–506, 1990.

43. E. P. Ippen, L. Y. Liu, and H. A. Haus, "Self-Starting Condition for Additive-Pulse Mode-Locked Lasers," *Opt. Lett. 15*, 183–185, 1990.



Laser Excitation

10.1 INTRODUCTION

This chapter introduces the "facts of life" of real lasers and discusses the physical processes that can lead to a population inversion. The pumping process is the most crucial issue facing laser research. It is well and good to talk about such esoteric topics as cavity modes, energy levels, and stimulated emission, but without adequate pumping, the system will not lase.

Fortunately, it is not as difficult as it might appear at first glance. Some state emphatically that "something in everything will lase if hit hard enough." The fact that ordinary Jello (the dessert) will lase gives considerable substance to the statement [and also becomes the first edible laser [1]]. Thus it is not news for another laser to be found, but it is news to understand the excitation route of any laser.

Almost any energy source can be used, even another laser. The following is a partial listing of the pumping agents that have been used successfully:

1. Optical
   - **a.** Incoherent—a flash lamp
   - **b.** Coherent—another laser
2. Electrons
   - **a.** A swarm—a gas discharge (DC, RF, or pulsed)

   b. An energetic electron beam (> 100 kV)
3. A thermal oven
4. A chemical reaction
   a. A chemical burn—a flame
   b. A rapid burn—an explosion
5. Heavy particles
   a. Ion beams
   b. Fission products from a reactor-like environment
6. Ionizing radiation
   a. A nuclear bomb (the ultimate in a one-shot experiment; i.e., singular)
   b. An x-ray source

Although considerable prejudice toward the gas laser has been used in compiling this list, many of the same techniques can be used for other types, such as the semiconductor laser. In that case, the most attractive means is to inject the electron-hole pairs into the appropriate cavity for recombination; but semiconductors have been pumped with electron beams, other lasers (gas, solid state, or semiconductor), and with incoherent sources. Semiconductor lasers have become so important that they will be covered in a separate chapter (see Chapter 11).

In view of this wide diversity of techniques and materials, we will have to restrict our attention to a few examples chosen to illustrate the principles involved, rather than to make the reader an expert in laser excitation. We restrict our attention to the first four categories, because most efficient lasers are excited in this manner.

Let us make a final general observation: We start talking about *real* systems, and real states in real systems have spectroscopic names. As interesting as the quantum rules and regulations are for assigning these names, we will not cover these issues here. Rather, we assume that you have some familiarity with spectroscopic notation and get on with the business at hand.*

10.2 THREE- AND FOUR-LEVEL LASERS

As was mentioned in Sec. 9.1.2, most lasers can be considered to have three or four *important* levels. This is, of course, a gross simplification since practical lasers have a very large number of levels. Nevertheless, the approximation is often used to establish the boundaries on the performance of various lasers.

Figure 10.1 describes four arrangements of energy levels that represent a very large number of lasers to be discussed in this chapter.

The first important issue to be addressed for any laser is the ease of establishing a population inversion and thus optical gain on the $2 \rightarrow 1$ transition. We start our analysis

*Some of the elementary ideas and formulas associated with atomic and molecular notation are given in Chapter 15. The section on vibrational-rotational transitions should be read before attempting to read the sections beyond Sec. 10.7.4.

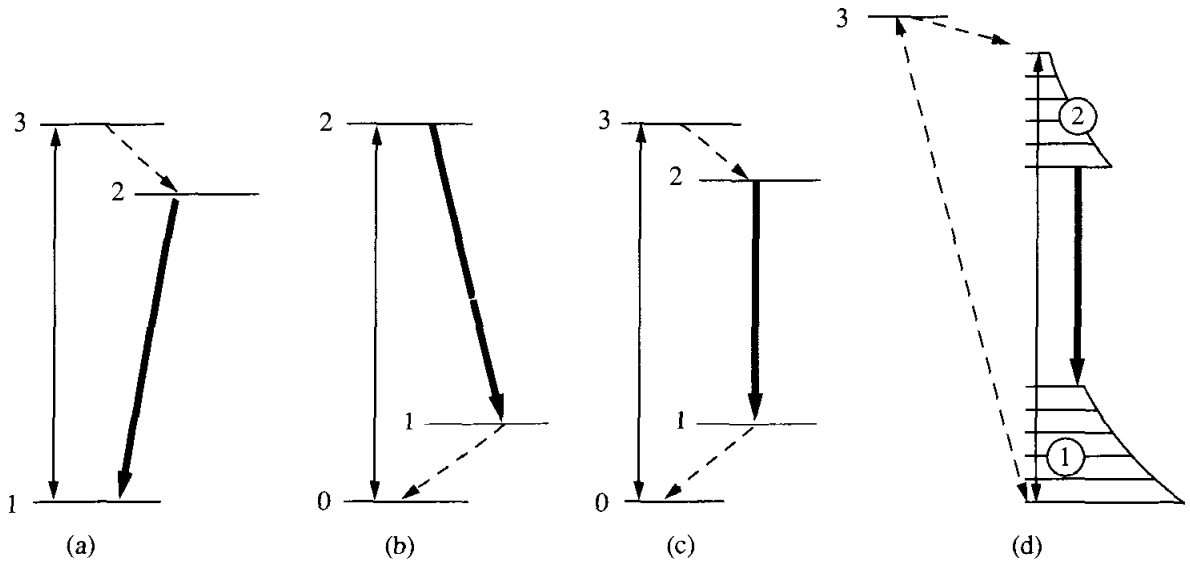


FIGURE 10.1. Arrangements of the important energy levels in various lasers. (a) and (b) are three-level lasers; (c) is a four-level one, and (d) is a two manifold case. In all cases, the double headed arrows represent the pumping route, the shaded arrow is the relaxation, and the heavy downward one represents the laser. The envelope of the levels in (d) represents the Boltzmann distribution within a manifold.

by first stating the obvious; namely, that the *sum* of the populations in all of the states must be equal to the original number of active atoms.

$$N_0 + N_1 + N_2 + N_3 + \cdots = [N] \tag{10.2.1}$$

Then we establish the initial conditions or the normal state of affairs that we wish to invert (subvert?) by the pump. If the external pump rate were equal to zero, then the system would be in thermodynamic equilibrium described by an ambient temperature T. This implies that

$$\frac{N_j}{N_0} = \frac{g_j}{g_0} \exp\left[\frac{E_j - E_0}{kT}\right] \tag{10.2.2}$$

where $g_{j,0}$ are the degeneracies and $E_{j,0}$ are the energies of the states $(j, 0)$.

If the energy gap between the lowest state and the next one up the ladder is large compared to kT, then it is a "safe" approximation to say that *all* of the $[N]$ atoms reside in the lowest state in the absence of pumping. If this gap is not large compared to kT, then we must account for the initial populations.

Consider Fig. 10.1(a) to appreciate one of the consequences of this statement. In order to obtain gain on the $2 \rightarrow 1$ transition, we must have more atoms in state 2 than in 1. Any population in state 3 has the possibility of being returned back to 1 and contributing nothing (absolutely zero!) to the laser. Hence we "hope" for a fast relaxation from 3 to 2 such that $N_3 \sim 0$ with nothing returning back to 3.

Our hopes must be tempered by reality. If there is a thermally generated physical process that converts state 3 to 2, there is also an inverse process that takes 2 back to 3. Quite often, we have incomplete knowledge about the details of this "process," and we have to resort to some general thermodynamic principles.

If r_{32} (sec^{-1}) is the rate of converting state 3 to 2, then $r_{23} = r_{32} \cdot \exp[-(E_3 - E_2)/kT]$ is the rate of reconverting 2 back to 3. If $(E_3 - E_2) \gg kT$, then we can neglect the back reaction. Thus we arrive at the conclusion that roughly half of the total atom population must be pumped to state 2 before one reaches optical transparency on $2 \rightarrow 1$ and can begin to think about a laser. Thus a three-level laser requires a significant pumping effort. If, for instance, state 2 decayed by radiation only, then we must surely need to supply a *critical fluorescence* power given by

$$\frac{P_f}{\text{vol}} = \frac{h\nu_{21}}{\tau_{\text{sp}}} \{N_2 = [N]/2\} \tag{10.2.3}$$

In Fig. 10.1(b), the starting population in state 1 is down by the Boltzmann factor. Thus,

$$N_0 + N_1 = [N] = N_0(1 + N_1/N_0) = N_0\{1 + \exp[-(E_1 - E_0)/kT]\}$$

or

$$N_0 = \frac{[N]}{1 + \exp[-(E_1 - E_0)/kT]} \quad \text{and} \quad N_1 = \frac{[N]\exp[-(E_1 - E_0)/kT]}{1 + \exp[-(E_1 - E_0)/kT]} \tag{10.2.4}$$

If $(E_1 - E_0)$ is also $\gg kT$, then it does not take a strong pump to establish a population in state 2 that exceeds that remaining in state 1. If the thermally driven interchange between 2 and 1 is very fast, then the population in state 2 must be

$$N_2 > \frac{[N]\exp[-(E_1 - E_0)/kT]}{1 + 2\exp[-(E_1 - E_0)kT]} \tag{10.2.5}$$

and thus the fluorescence power is less than that given by (10.2.3).

If $E_1 - E_0 \gg kT$, then $N_2 \ll [N]$ and most of the atoms continue to reside in the ground state. Unfortunately, if the pump can create atoms in 2 from 0, then it can also return atoms back from 2 to 0 as can be immediately verified by assuming optical pumping by another laser tuned to the $2 \rightarrow 0$ transition as a pump. Then the reverse reaction is stimulated emission from 2 back to 0.

The four-level scheme shown in Fig. 10.1(c) is the most desirable one. One pumps from 0 to 3, which relaxes to 2, lasing takes place between $2 \rightarrow 1$ and state 1 relaxes back to 0. If $E_3 - E_2 \gg kT$, then there is very little reaction sending atoms from 2 back to 3; and if $E_1 - E_0 \gg kT$, then there is little population in 1 to be overcome. Most of the CW lasers fall into this category.

A case that has become very important for fiberoptic amplifiers is shown in Fig. 10.1(d) and is very similar to that appearing in the vibrational-rotational band of molecular gas lasers covered later. There are two *manifolds* containing multiple levels, which interact within each manifold at a rate that far exceeds the interaction rate between them.* Under such circumstances, the total population in each manifold will be apportioned to the levels according to the Boltzmann relation. If there is sufficient total population in 2, then an inversion can exist between the lower states of it and the upper ones of 1 even though the

*The semiconductor laser falls into this category, but, because of its technological importance, it will be covered separately in Chapter 11.

total population in 2 may be less than in 1. One can pump (optically) on an absorptive transition (shown dashed) or to an intermediate state 3 provided that those relax back to 2.

All of these considerations can be found in practical lasers.

10.3 RUBY LASERS

Probably the most straightforward technique for creating a population inversion is to use another photon source as a pump. Historically, this approach to lasers grew out of earlier successes in the microwave portion of the spectrum (i.e., maser) and is still widely used. We use a coherent or an incoherent source of radiation at the frequency $1 \rightarrow 3$ (or $0 \rightarrow 3$) in Fig. 10.1 to create the population inversion.

The ruby laser was the first laser [2] and operated in the visible portion of the spectrum at $\lambda_0 = 6943$ Å. The chemical composition of ruby consists of a crystal of sapphire (Al_2O_3) in which a small amount of the aluminum is replaced by chromium by adding Cr_2O_3 to the melt in the growth process. The pure host crystal possesses a rhombohedral unit cell shown in Fig. 10.2(a) where the axis of symmetry is the so-called c axis. Because of the arrangement of the atoms, the crystal is uniaxial with different indices of refraction depending on whether the electric field E is perpendicular to c (an ordinary wave with $n_0 = 1.763$) or parallel to c (an extraordinary wave with $n_e = 1.755$).

The chromium atoms are active in the lattice as triply ionized ones, Cr^{3+}, and give rise to the energy-level diagram shown in Fig. 10.2(b). The states participating in the laser transition are labeled by 2 and 1 to correspond to our previous notation. Note that the upper state is split into two levels, denoted by $2\bar{A}$ and $\bar{E}$, which are separated by 29 cm^{-1}. Transitions originating from the $2\bar{A}$ are denoted by R_2, and those from $\bar{E}$ are called R_1. The lifetime of the E states because of spontaneous emission is 3 ms, which is very long by atomic process standards, and thus these levels are sometimes classified as *metastable*. Because of the anistropy, the spontaneous emission (and thus stimulated emission) depends on the orientation of the optical electric field with respect to the axis (more about this later). The degeneracy of each of the upper states is 2, whereas that of the ground state is 4. Hence, we need to have the population in the $\bar{E}$ level to be half of that in the $^4A_2$ state to obtain optical transparency, $N_2 - (g_2/g_1)N_1 = 0$, on the R_1 transition. Note that the $2\bar{A}$ level will have an almost equal population to that of the $\bar{E}$ state, and thus roughly half of the doped chromium atoms must be prompted to the ^2E manifold to obtain transparency on the R_1 line. Further pumping yields gain and oscillation if proper feedback is supplied.

One of the very fortunate and desirable features of ruby is the presence of very strong absorption bands corresponding to the $^4A_2 \rightarrow {}^4F_2, {}^4F_1$ transitions shown. These bands are nearly 1000 Å wide and are located in the green (18,000 cm^{-1}) and violet (25,000 cm^{-1}) portions of the spectrum. Because of their width and strength of optical attenuation they can absorb a significant fraction of light emitted by an incoherent flash lamp that radiates more or less as a blackbody source at an elevated temperature of about 7000 K to 9000 K. Fortunately the lifetimes of the pumped states, $^4F_2$ and $^4F_1$, are very short, with most of the atoms returning to the ^2E levels rather than back to the ground state. The mean quantum efficiency of the pumping process for the two bands is about 70%; that is, 70% of the atoms

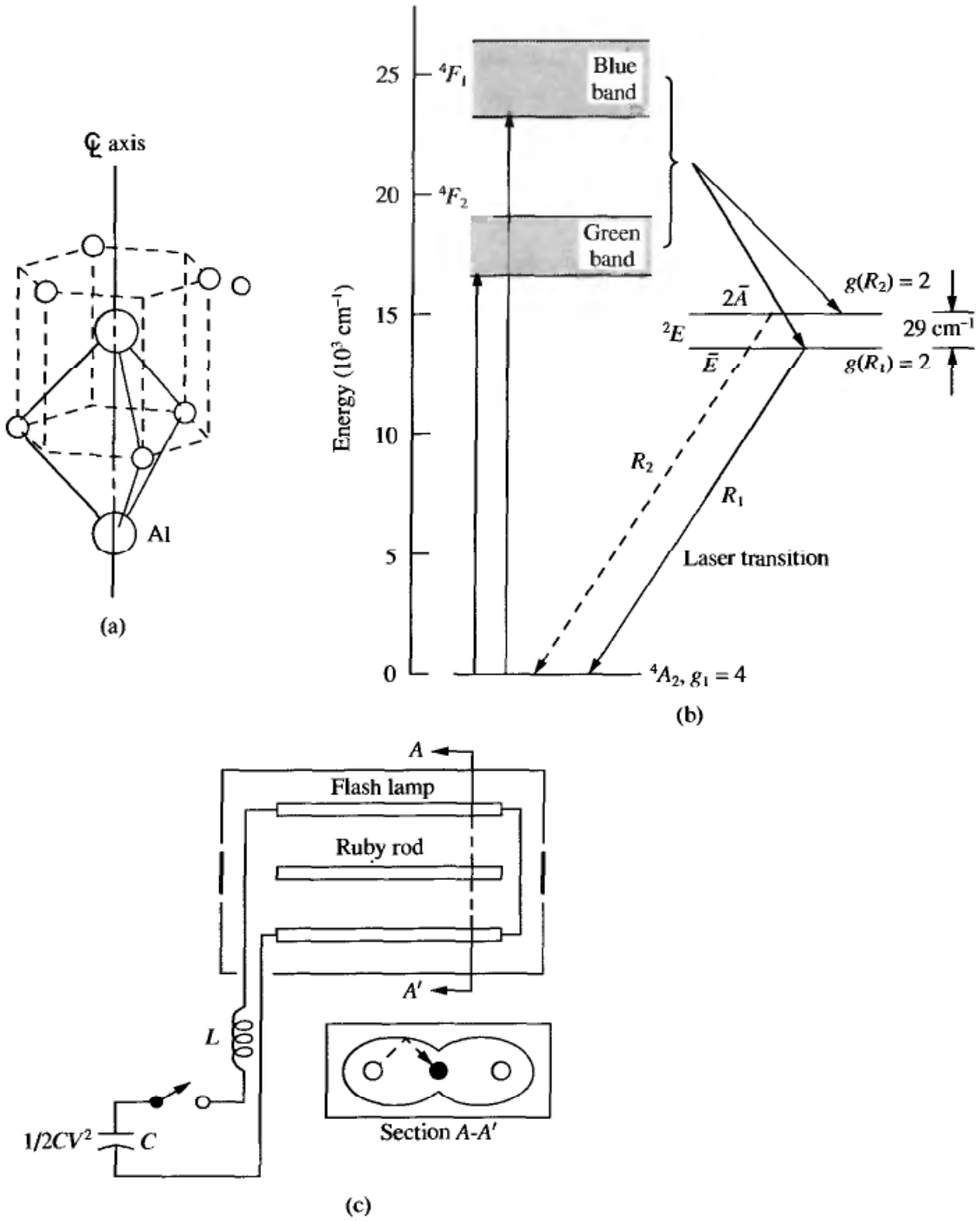


FIGURE 10.2. Ruby laser. (a) Crystalline structure (chromium enters at an Al site.). (b) Energy-level diagram for Cr^{3+} in Al_2O_3. (c) Typical pumping scheme using a dual elliptical cavity.

promoted to the ^4F levels appear in the ^2E levels. A summary of data for a typical ruby rod is given in Table 10.1, and some is presented graphically in Fig. 10.3.

Note that the absorption cross section for the normal (i.e., unpumped) rod shown in Fig. 10.3 is very strong for the green and violet bands, much stronger than for the laser transition. Indeed the absorption at 6943 Å is hardly "on the map" of Fig. 10.3(a). The details of that absorption are shown with a greatly expanded scale in Fig. 10.3(b). It is important to realize that the data presented in Fig. 10.3 were taken with a low-level signal, small enough to neglect any significant depletion of the ground state population. When the rod is pumped, the $^4A_2$ state density decreases, as does the absorption on all of the bands originating there, hopefully converting what was absorption at 6943 Å to gain.

The stimulated emission cross section at the R_1 and R_2 lines can be found directly from the data shown in Fig. 10.3(b) with due attention paid to the degeneracies shown in Fig. 10.2(b). The stimulated emission cross section is related to the absorption cross section by

$$\sigma_{SE} = \left[\frac{g_1}{g_2}\right]\sigma_{abs} \tag{7.7.3}$$

Hence, $\sigma_{SE} = 2.5 \times 10^{-20}$ cm^2 for the R_1 line with **E** perpendicular to the c axis.

Obviously, ruby is a three-level laser with all the attendant difficulties. The pump-to-laser route is as follows:

1. The optical radiation from the flash lamps in the wavelength region between 400 and 600 nm is absorbed by the atoms in the ground state ($^4A_2$), promoting them to the $^4F_2$ state. This same radiation attempts to stimulate these atoms to return to the ground

TABLE 10.1 Physical Data on a Typical Ruby System*

Item	Value
Cr_2O_3 doping (% by weight)	0.05
Cr^{3+} concentration	1.58×10^{19} cm^{-3}
Output wavelength (at 25°C)	R_1 : 14403 cm^{-1} $\leftrightarrow$ 6943 Å
	R_2 : 14432 cm^{-1} $\leftrightarrow$ 6929 Å
Spectral line width (300°K)	11 cm^{-1} or 5.3 Å
Quantum efficiency (of pumping)	0.7
Absorption cross section of R_1 laser line	1.22×10^{-20} cm^2
Stimulated emission cross section	2.5×10^{-20} cm^2
Residual scatter losses in crystal at R_1	$\sim$ 0.001 cm^{-1}
Major pump bands	
Blue (404 nm)	$\alpha_\parallel = 2.8$ cm^{-1} $\alpha_\perp = 3.2$ cm^{-1}
Green (554 nm)	$\alpha_\parallel = 2.8$ cm^{-1} $\alpha_\perp = 1.4$ cm^{-1}
Refractive indices	
Ordinary ray ($E \perp c$)	1.763
Extraordinary ray ($E \parallel c$)	1.755

*Data from Koechner (Ref. 24).

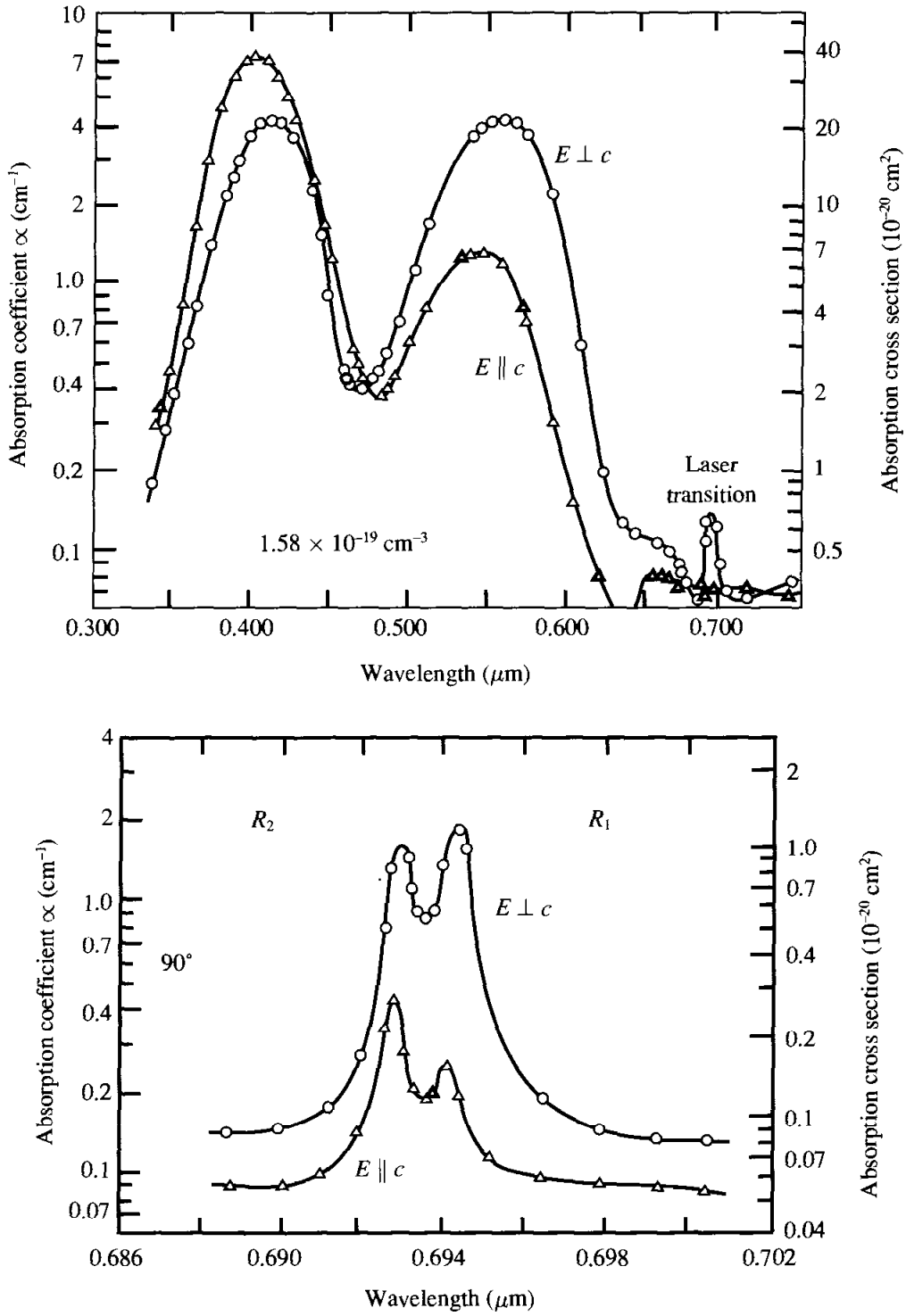


FIGURE 10.3. Absorption coefficient and cross section for ruby with $Cr^{3+} = 1.58 \times 10^{19}$ cm^{-3}.

level. In competition with this stimulated return is a fast relaxation of level 3 to the upper laser level.

2. Because of this fast relaxation and very slow spontaneous emission from the upper laser level back to the ground, a considerable number of atoms can accumulate in the ^2E state. If the $1 \rightarrow 3 \rightarrow 2$ process is fast compared to (τ_{sp}^{-1}), an inversion of population can be achieved on the $2 \rightarrow 1$ transition.
3. We have a laser if proper feedback is provided.

There are several questions that could (or should) be asked concerning this scenario:

1. Where did the energy go in the fast relaxation between levels 3 and 2? Answer: Most, if not all, goes into heating the crystal lattice.
2. Why is the pump radiation not absorbed at the R_1 and R_2 transitions? Answer: It is, but observe the magnitude of the absorption cross section in the vicinity of 694.3 nm: $\sigma \leq 1.2 \times 10^{-20}$ cm^2 [see Fig. 10.3(b)]. This is 20 times lower than the absorption cross section at 550 nm.
3. Why is the pump radiation not amplified on the R_1 and R_2 wavelengths? Answer: It is, but the amplification path is only the diameter of the rod, the stimulated cross section is small, and the pump radiation is comparatively weak compared to the much more intense laser radiation (when it starts).

A mathematical analysis can be quite involved even for such a "clean" optical pumping scheme. First, we need to determine the spectral distribution of the flash lamp. Obviously, if most of its radiation is concentrated outside the absorption bands, few atoms will be promoted to state 3 and thus to 2. On the other hand, if the power is concentrated in a spectral region where the absorption is too high, most of the pumping occurs on the outer radius of the rod. Thus matching the spectral characteristic of the lamp to the active medium to optimize the pumping is not a simple engineering job.

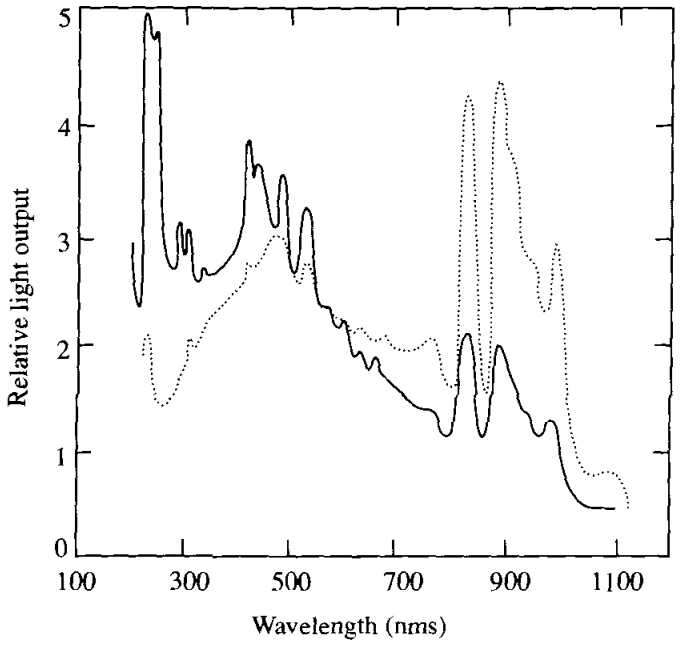
Typical emission characteristics are shown in Fig. 10.4 for flash lamps filled with different gases and excited in a variety of ways. According to Ref. 4, 55% of the electrical energy into the lamp of Fig. 10.4(b) appears as radiant energy* in the wavelength interval 300 to 1000 nm; a similar percentage applies for the lamp of Fig. 10.4(a) . This figure illustrates that the spectral characteristics depend on the fill gas—xenon or krypton—and as Fig. 10.4(a) shows, the current density in the lamp.

Given this information, we could in principle do a fair job of modeling the ruby laser. However, it is quite tedious and requires a computer to do an accurate job. Some of the computation difficulties and some "obvious" physical limits to the pumping sequence are given next.

1. We must conserve atoms; that is,

$$N_1 + N_2 + N_3 = N_0 \tag{10.3.1}$$

*This efficiency is far better than that of the best visible laser.



Spectral data curves show the effect of operating the FX 38C-3 (4 mm I.D., 3 inch arc length, xenon filled–400 to flash tube at 50 joules input at low and high voltage to enhance either the IR or the UV output. Solid line curve shows data at 50 joules input or 1400 VDC, 51 μF, 0 inductance. Dotted line curve shows data at 50 joules inp or 700 VDC, 205 μF, 0 inductance. Shifts in spectral dat are primarily due to changes in current density J, A/cm^2. Higher current density J shifts spectral output toward the shorter wavelengths or blue region, UV. Conversely low current density J shifts spectral output toward the longer wavelength or red region, IR. Data was taken with EG&(580/585 Spectroradiometer system. Spectral resolution 1 nm in the 200–700 nm region; and 20 nm in the 700–120 nm region.

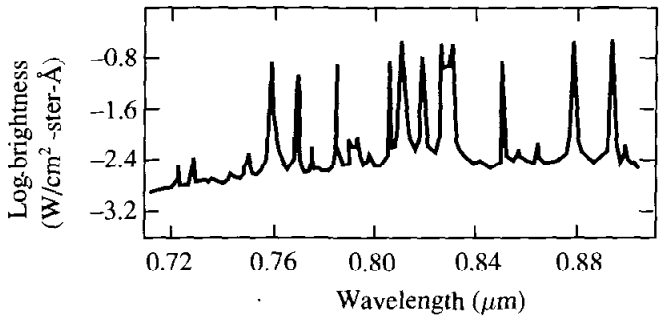


FIGURE 10.4. Spectral characteristic of typical lamps: (a) data on EG & G lamp (Courtesy of EG & G Electro-Optics.); (b) brightness of a CW-pumped krypton arc lamp. (From Koechner et al. [4].)

or the sum of the atom densities in $^4A_2$ (state 1), the sum of those in $\bar{E} + 2\bar{A}$ (state 2), and the sum of those in $^4F_2 + ^4F_1$ (state 3) must equal the original doping density N_0. Fortunately the lifetime of ^4F levels is very short, and hence their density can be neglected.

2. The time scale for the interchange of population between the $\bar{E}$ and $2\bar{A}$ states is short ($\sim$ 1 ns) but not instantaneous. For pulses short compared to this time, only the population in a particular state—say the $\bar{E}$ level—can be extracted by stimulated emission by the R_1 radiation. For pulses longer than this time scale, the ^2E manifold can be considered as one level, with, however, a stimulated emission cross section of the R_1 line (2.5×10^{-20} cm^2).

3. Because of the short time for interchange of populations between $2\bar{A}$ and $\bar{E}$ we can consider those states to be in local thermodynamic equilibrium at the lattice

temperature, the ratio of populations being given by the Boltzmann factor.

$$\frac{N(2\bar{A})}{N(\bar{E})} = \frac{2}{2} \cdot \exp\left[-\frac{\Delta E}{kT}\right] \tag{10.3.2a}$$

$$= 0.87 = K \tag{10.3.2b}$$

for $\Delta E = 29\ \text{cm}^{-1}$ and $kT = 208\ \text{cm}^{-1}$ (300 K). Thus if we name the sum $\bar{E} + 2\bar{A}$ as state 2, there are

$$N(2\bar{A}) = \frac{K}{1+K} \cdot N_2 \tag{10.3.2c}$$

$$N(\bar{E}) = \frac{1}{1+K} \cdot N_2 \tag{10.3.2d}$$

atoms in the respective levels of the ^2E manifold. To obtain optical transparency at 6943 Å, we need $N(\bar{E}) = 2/4N(^4A_2) = 1/2N_1$. Thus

$$\frac{N_2}{1+K} - \frac{1}{2}N_1 = 0 \qquad \text{(optical transparency)} \tag{10.3.3a}$$

$$N_2 + N_1 = N_0 \qquad \text{(conservation of atoms)} \tag{10.3.3b}$$

or

$$N_2 = \frac{1+K}{3+K}N_0 = 0.483N_0 \tag{10.3.3c}$$

$$N_1 = \frac{2}{3+K}N_0 = 0.517N_0 \tag{10.3.3d}$$

Note those values are just about the same as we would obtain by assuming equal degeneracies between N_2 and N_1 and ignoring the splitting in the upper state as was done in the section for Q switching.

4. There is considerable spontaneous fluorescence emitted even at the condition of zero optical gain.

$$P_{\text{spont.}} = h\nu \cdot \frac{N_{2c}}{\tau_{sp}} = \frac{1+\text{K}}{3+\text{K}} \cdot \frac{h\nu N_0}{\tau_{sp}} = 727\ \text{W/cm}^3 \tag{10.3.3e}$$

for $N_0 = 1.58 \times 10^{19}\ \text{cm}^{-3}$ and $\tau_{sp} = 3$ ms. (This is called the critical fluorescence power.) Any further pumping yields gain.

5. If we assume a quantum efficiency of the pumping process of 0.7, then we must absorb

$$P_{ab} = \frac{P_{sp}}{\eta} = 1.04\ \text{kW/cm}^3 \tag{10.3.4}$$

from the flash lamp in order to just reach the condition of optical transparency.

6. If we assume a conversion efficiency of electrical energy stored in the capacitor to emitted radiation of the flash lamp into all wavelengths of 55%, with 20% of that radiated energy being absorbed in the two bands, and a time scale for the pumping

of 3 ms (it cannot be much longer—otherwise the atoms decay as fast as they are produced), then we require a stored energy in the capacitor ($\frac{1}{2}$ CV^2) of

$$W = \frac{1.04 \times 10^3 \text{ W/cm}^3 \times (3 \times 10^{-3} \text{ sec})}{0.55 \times 0.2} = 28.3 \frac{\text{joules}}{\text{cm}^3 \text{ of rod}} \tag{10.3.5}$$

Obviously, more energy is required to obtain a laser.

7. A better computation is frustrated by the fact that the pump radiation is propagating across the chord of a cylindrical rod and the number of atoms absorbing that radiation is changing with time. It is safe to say that the calculation is tedious.

10.4 RARE EARTH LASERS AND AMPLIFIERS

10.4.1 General Considerations

A common solid state laser material consists of an optically transparent host such as glass (mostly SiO_2), YAG ($Y_3Al_5O_{12}$), sapphire (Al_2O_3), etc., doped with a small amount of a rare earth oxide.* The active atom is the rare earth and usually acts as if it were triply ionized since the outermost three electrons are involved in the bonding (the solid state material is charge neutral).

Some of the unique laser properties arise because of a subtle point appearing in nature. In the rare earths, the active electron participating in the optical transition is one of the $4f$ electrons that is shielded by electrons in the larger $n = 5$ and 6 shells (or orbits). One of the consequences is that the energy levels of rare earths are only weakly dependent on the host lattice. Thus, while there are small changes in the wavelengths of the RE laser depending on the host lattice, those changes appear in the third or fourth significant figure.

Per usual, we depend upon experiment to identify the character of the quantum states, the energy levels, any nonradiative quenching rates, branching ratios, the Einstein coefficients, and the transition probabilities.

The characteristics of the quantum states are usually specified by the spectroscopic name according to the following scheme.

$$\boxed{{}^{\text{superscript = number}}[\text{Letter}]_{\text{subscript = number}}}$$

where the letter symbol indicates the orbital angular momentum quantum number according to the following prescription:

Letter	S	P	D	F	G	H	I	etc.
$L =$	0	1	2	3	4	5	6	etc.

(10.4.1)

the numerical value for the superscript $= 2S + 1$ (10.4.2)

*The rare earths are those elements with atomic numbers from 57 to 71 with two of the most important being Nd (60) and Er (68).

where S is the spin angular momentum (not to be confused with the letter symbol S indicating an orbital momentum of 0).

$$\text{the numerical value for the subscript} = J \tag{10.4.3}$$

where J is total angular momentum quantum number.

A level such as $^4I_{9/2}$ (the ground state of the neodymium) is referred to as "quartet-I-nine-halves" and from its "name", we immediately know that the degeneracy of that level is $g = 2J + 1 = 10$. In other words, there are 10 different orthogonal wavefunctions describing this state.

However, these active atoms experience the local electric field generated by the host lattice and thus these levels are shifted by the electric field at the site where the rare earth resides (the Stark effect). Since the Stark effect is proportional to $|E|$, not its direction, a positive and negative field yield the same shift. For Nd:YAG, the one manifold splits into $(2J + 1)/2$ distinct doubly degenerate levels.* Hence, we can ignore all degeneracy factors in our prior equations or let all g's be equal to 2 as one sees fit.

10.4.2 Nd:YAG lasers: Data

A common solid-state material is made by doping a rare earth, neodymium, into a variety of host materials, with the most common ones being amorphous glass and crystalline yttrium aluminum garnet ($Y_3Al_5O_{12}$), or YAG. In each case, the active atoms participate as if they were triply ionized, Nd^{3+}, with energy levels and broadening of the states dependent on the host lattice. (The crystal is, of course, charge neutral, the three electrons being bonded to the neighboring atoms of the host.) The dopant (1% to 2%) goes into the amorphous glass at random sites, and thus each Nd^{3+} ion "sees" a slightly different environment. On the other hand, the Nd^{3+} substitutes for Y^{3+} in the cubic crystal of YAG, and thus each of these dopant atoms sees more or less identical environments. It should come as no surprise then that the glass laser transition is inhomogeneously broadened (but not by the Doppler phenomenon) with a comparatively wide line, whereas the YAG transition line widths are much smaller (although still inhomogeneously broadened by lattice vibrations and site-to-site differences). While the YAG and the glass laser resemble each other, both lasing at $\lambda_0 \sim 1.06\ \mu$m, they are sufficiently different to warrant separate discussions.

Pure YAG, $Y_3Al_5O_{12}$, is an optically isotropic crystal with a cubic structure characteristic of garnets. Because of the difference in size of the Nd^{3+} that is substituted for Y^{3+} (about 3%), one is limited in the amount of neodymium that can be included to about 1%; otherwise, the crystal becomes severely strained. Because the active atoms are in a well-defined environment, the energy levels are well defined and narrow and are shown in Fig. 10.5(a) with a greater detail in Fig. 10.5(b) where the dominant laser transition at $\lambda_0 = 1.064$ is shown as the transition between the upper state of the $^4F_{3/2}$ (at 11507 cm^{-1}) and the lower state $^4I_{11/2}$ at 2110 cm^{-1}. In Fig. 10.5(c) the energy levels are shown presuming that the host lattice was cooled to liquid nitrogen temperatures. Specific data on the various

*[The exact number depends upon whether J is even or odd and the symmetry presented by the host lattice (see Ref. 103)].

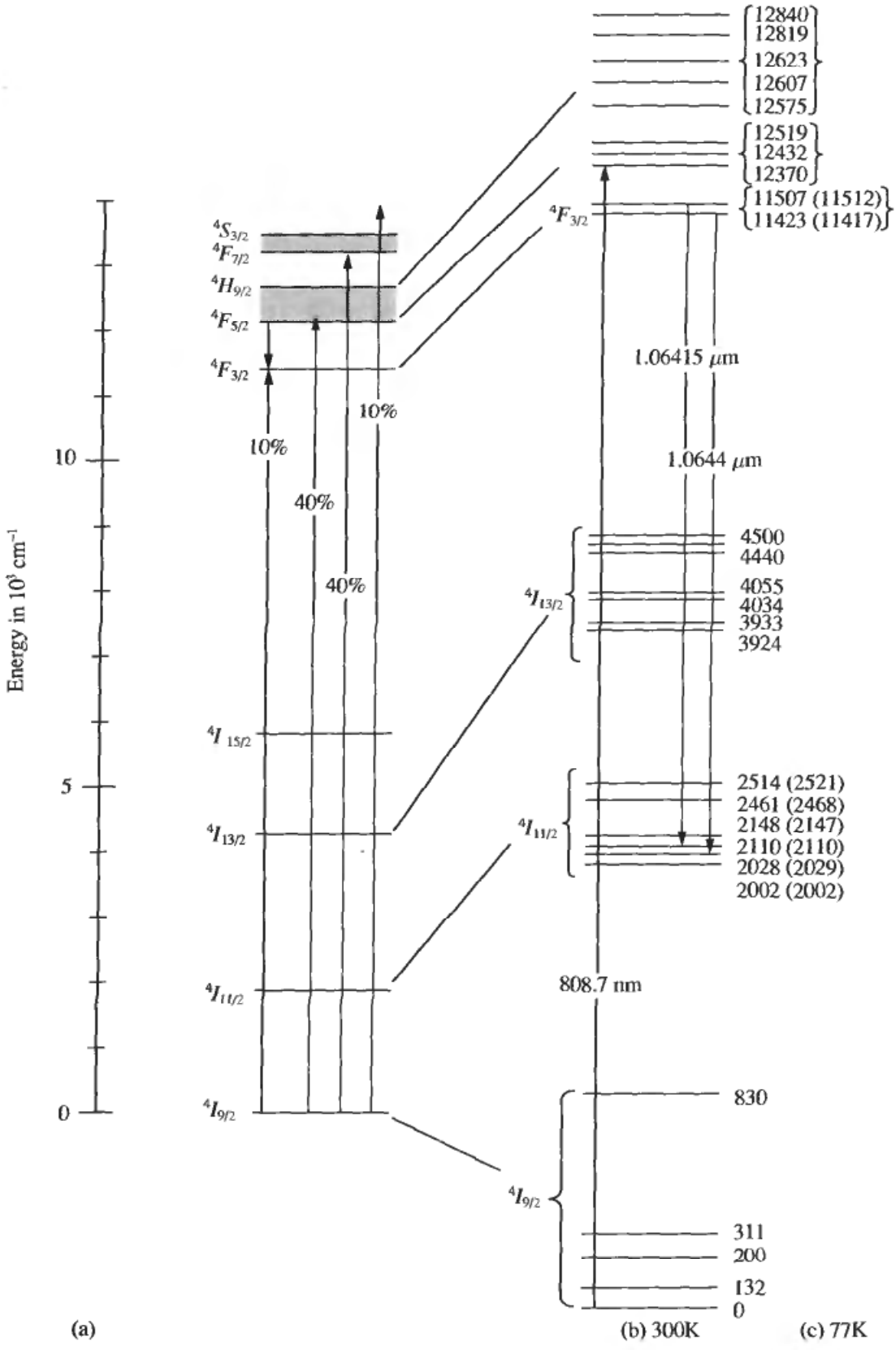


FIGURE 10.5. Energy level for neodymium in YAG. (a) Structure of YAG showing the pumping routes with the percentages referring to a pump with a broad spectral output. (b) Details of the manifold at 300 K showing the dominant transitions, the semiconductor laser pumping route is also shown. (c) Energy levels at 77 K. [Data from Kaminski [25]. See also Koechner [24].]

TABLE 10.2 Characteristics of a Typical Neodymium-YAG Laser Rod

Bulk parameters	Value
Chemical formula	Nd: $Y_3Al_5O_{12}$
Weight % (Nd)	0.725
Atomic % (Nd)	1.0
Density of Nd atoms	$1.38 \times 10^{+20}$ cm^{-3}
Index of refraction	1.81633 (1.06 μm, 1% doping)
Scattering losses	0.002 cm^{-1}
Dominant transition	(11, 507 $\rightarrow$ 2111 cm^{-1}) 1.064 μm
Fluorescent lifetime of $^4F_{3/2}$	255 μs
Fluorescent efficiency of $^4F_{3/2}$	99.5%
Stimulated emission cross section	2.7–8.8 $\times 10^{-19}$ cm^2 (doping and temperature dependent)
Lower state lifetime ($^4I_{11/2} \rightarrow {}^4I_{9/2}$	$\sim$ 30 ns

From Koechner [24].

TABLE 10.3 Detailed Data on $^4F_{3/2} \rightarrow {}^4I_{13/2}, {}^4I_{11/2}, {}^4I_{9/2}$ Transitions

Transition	Wavelength (μm)	Branching ratio	Line width (FWHM in cm^{-1})	σ (10^{-19} cm^2)
$^4F_{3/2} \rightarrow {}^4I_{9/2}$	0.86–0.95 μm	0.3		
$^4F_{3/2} \rightarrow {}^4I_{11/2}$	1.0521 C	0.0383	4.5	
	1.0549	0.0023	4.5	
	1.0615 B	0.0799	3.6	
	1.06415 A	0.1275	5.0	2.9
	1.0644 A′	0.0533	4.2	*
	1.0682	0.0340	6.5	
	1.0737	0.0657	4.6	
	1.0779	0.0463	7.0	
	1.1055	0.0145	11.0	
	1.1119	0.0297	10.2	
	1.1158	0.0356	10.6	
	1.1225	0.0328	9.9	
		$\Sigma = 0.56$		
$^4F_{3/2} \rightarrow {}^4I_{13/2}$	1.3–1.4	0.14		
$^4F_{3/2} \rightarrow {}^4I_{15/2}$	1.74–2.13	~0.01		

*Because of their proximity the A and A′ transitions contribute to each other. Thus the effective stimulated emission coefficient at 1.06415 μm increases when the shift, line width, and Boltzmann factor are included in the calculation.

From Koechner [24].

transitions, arranged in order of wavelength, are given in Tables 10.2 and 10.3. The last column of Table 10.3 is "saved" as a problem at the end of the chapter. The other transitions will lase given a suitable wavelength selective cavity.

The typical pumping route, such as from a flash lamp, is also indicated in Fig. 10.5(a) with approximately 10% of the absorbed energy going directly to the upper laser manifold $^4F_{3/2}$ and the remainder going to the higher states. Also shown is the route for the case of a semiconductor laser pumping the YAG using radiation near 808 nm. For the most part, the ions promoted to $^4F_{5/2}$, $^4H_{9/2}$ and so on return to the $^4F_{3/2}$ level for participation in the laser action. The lower laser level, $^4I_{11/2}$, decays with a lifetime of $\sim$ 30 ns, but that radiation is strongly absorbed by the host lattice and thus that energy shows up as heat. Fortunately, YAG has a high thermal conductivity and this unwanted energy can be removed by conduction. The system can be operated either CW or pulsed, mode locked or Q switched (or combinations thereof).

The fact that the states are well defined is a double-edged sword. The system is obviously a four-level laser, and because of the narrow line widths the stimulated emission cross section is large and the pumping threshold is low. However, the absorption bands are also narrow, and thus we do not make good use of the radiation emitted by an efficient Xe flash lamp. For that reason, we attempt to use other gases in the pumping lamp, such as krypton, which matches the pumping bands.

Recently, there has been a significant breakthrough for low power Nd-YAG systems by using the output of semiconductor lasers to pump the upper levels directly. Inasmuch as the efficiency of semiconductor lasers to convert electrical energy to photons is quite high ($> 40\%$), this enables us to pump the YAG laser with simple power supplies (i.e., flashlight batteries as a last resort). By using frequency doubling techniques, the output wavelength can be converted to the "green" ($\lambda_0 = 1.064/2 = 532$ nm) with a diffraction-limited $TEM_{0,0}$ mode as the output. The entire system can be contained in a "cigar box" or smaller. The analysis of such a system is tractable and serves as a good review of the basic laser theory.

10.4.3 Nd:YAG Pumped by a Semiconductor Laser

Let us use the data of Tables 10.2 and 10.3 and the energy levels for Nd:YAG specified by Fig. 10.5 to predict the performance of the axially pumped ring laser shown in Fig. 10.6. The following approximations are made to make the problem tractable but keeping it realistic:

1. Oscillation is constrained to the counterclockwise direction at a wavelength of $\sim$ 1.064 μm. Mirror M_1 is presumed to have a very high transmission ($T \sim 1$) at the pump wavelength of 808 nm and a high reflectivity at 1.06 μm.
2. CW operation and hence $d(\ )/dt = 0$.
3. We ignore the significant problem of optimizing the overlap between the pump beam (at $\sim$ 808 nm) and the laser (at 1.06 μm).
4. We presume that the YAG crystal is long enough so that pump intensity at $z = l_g$ is very small compared to that at $z = 0$.

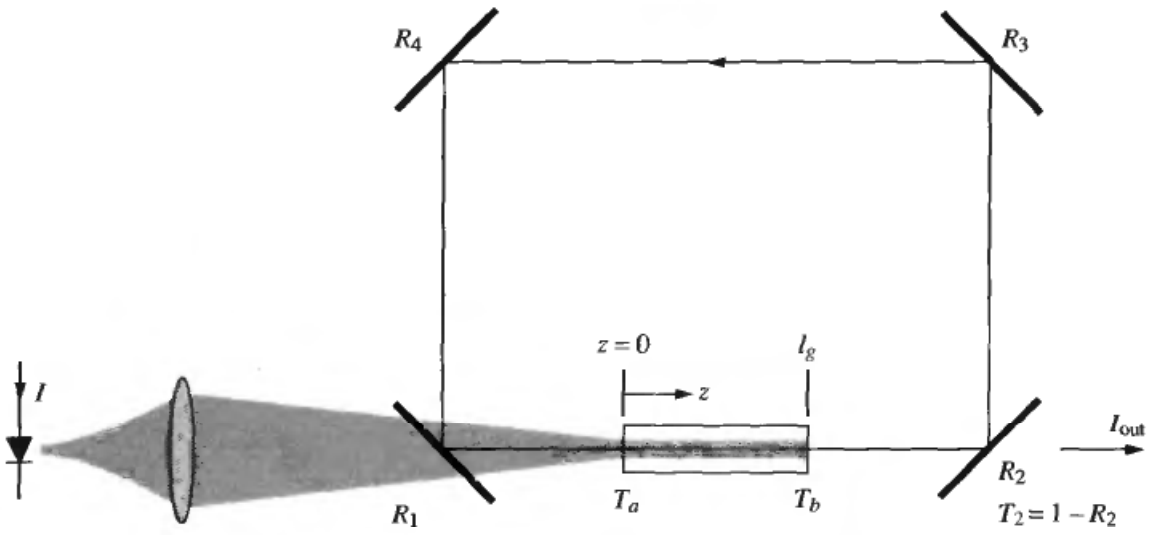


FIGURE 10.6. Schematic of the YAG ring pumped by a semiconductor laser.

5. The pumping will be to one or more levels of the $^4F_{5/2}$ or $^4H_{9/2}$ manifold, from which a fast relaxation to the $^4F_{3/2}$ takes place. To avoid a separate rate equation, we define η_p, the pumping efficiency, also called the quantum yield, as being the fraction (< 1) of the atoms promoted to the $^4F_{5/2}$ or $^4H_{9/2}$ manifolds that relax to the $^4F_{3/2}$ group.
6. A Boltzmann distribution of the populations among the states of $^4F_{3/2}$ and $^4I_{9/2}$ manifolds is maintained by some very fast thermally driven process. Thus, at room temperature ($kT = 208\ \text{cm}^{-1}$) the state at 11,507 cm^{-1} has 40.04% of that in the $^4F_{3/2}$ manifold with the remainder in the state at 11,423 cm^{-1}. The distribution in the bottom $^4I_{9/2}$ manifold is 46.4% in 0, 24.6% at 132 cm^{-1}, 17.7% at 200 cm^{-1}, 10.4% at 311 cm^{-1}, and 0.9% at 830 cm^{-1}. All of these percentages should be verified by the student.
7. The lifetime of the $^4I_{11/2}$ manifold is so short (30 ns) compared to the $^4F_{3/2}$ (255 μs) that we set any changes in $N_1 = [^4I_{11/2}] = 0$. We also ignore the thermal population in $^4I_{11/2}$ since its density amounts to only 0.02% of the total doping.
8. Note that there are two laser intensities, frequencies, wavelengths, stimulated emission cross sections, and saturation intensities. We will use the letters P and L to distinguish between them.
9. Assume all states have the same degeneracy ($g = 2$). Ignore the small losses associated with scattering from crystal imperfections.

Our goal is to predict the intensity and the laser wavelength in the vicinity of 1.06 μm of the output.

The specification of the wavelength equal to 1.064 μm illustrates a new issue arising in a practical system. Table 10.3 indicates that there are two closely spaced transitions which fit this specification.* They are so close that the gain on one will contribute to the other.

*The wavenumber for the B transition of Table 10.3 is 9420.6 cm^{-1}, which is far removed from the A and A' values.

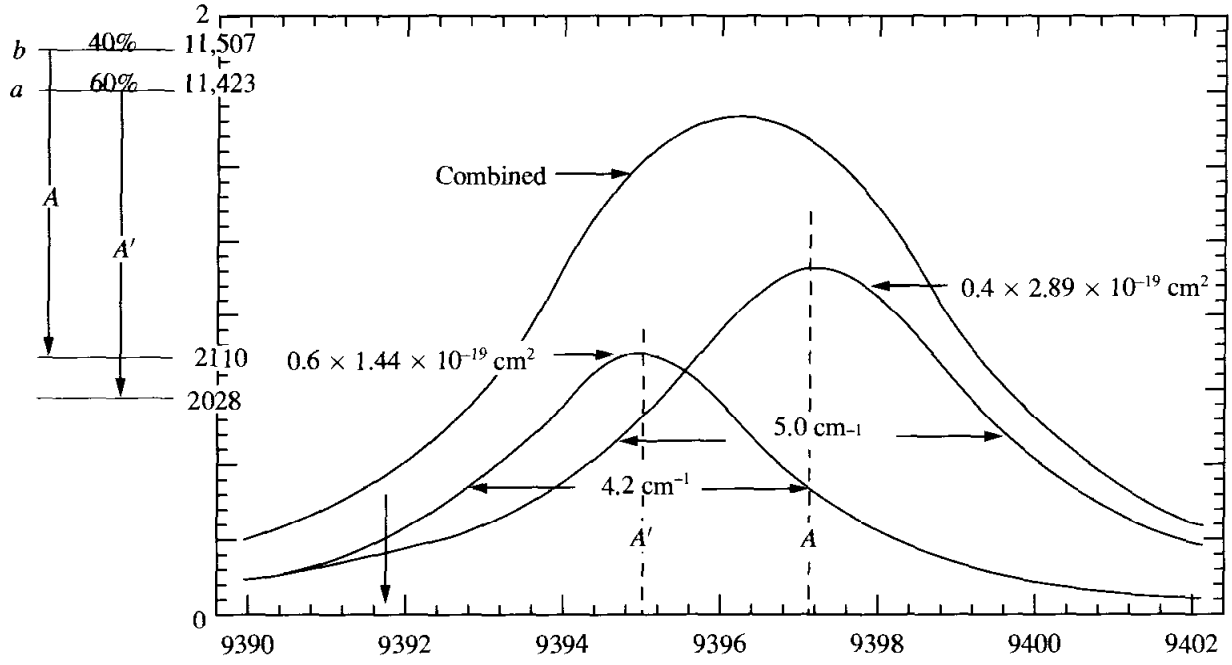


FIGURE 10.7. The states contributing to the gain at 1.064 μm and the overlapping line shapes. The composite line shape includes the fraction of the density in the $^4F_{3/2}$ manifold in each state and maximizes at $\sim$ 1.23 cm^{-1} from the center of the A' transition and 0.94 cm^{-1} from that of the A.

This is shown in Fig. 10.7, where the energy levels associated with the A and A' transitions are identified. At each peak the stimulated emission cross section is evaluated and weighted by the fraction of the $^4F_{3/2}$ population in the upper state, and the variation of σ with wave number for each transition is shown along with the sum using the data of Table 10.2.

If we assume that the population of $^4I_{11/2}$ (the lower state manifold) is negligible (consistent with its very short lifetime), then the gain coefficient is given by

$$\gamma(\nu) = N_{2a}\sigma_{A'}\bar{g}_{A'}(\nu) + N_{2b}\sigma_A\bar{g}_A(\nu) \tag{10.4.4}$$

where $N_{2a} + N_{2b} = [^4F_{3/2}]$, the density in upper state manifold.

Assumption 6 specifies that there is a fast thermal process maintaining a Boltzmann distribution in the $^4F_{3/2}$ manifold.

$$\frac{N_{2b}}{N_{2a}} = \exp[-\Delta E/kT] \tag{10.4.5}$$

with the ratio being 0.668 at room temperature ($kT = 208$ cm^{-1}) and $\Delta E = 11,507 - 11,423 = 84$ cm^{-1}. Thus, if we wish to express the gain coefficient in terms of the $[^4F_{3/2}]$ density, (10.4.4) becomes:

$$\gamma(\nu) = [^4F_{3/2}]\left\{f_a\sigma_{A'}\bar{g}_{A'}(\nu) + f_b\sigma_A\bar{g}_A(\nu)\right\} = [^4F_{3/2}]\sigma_{\text{eff}}\bar{g}_c(\nu) \tag{10.4.6}$$

where $f_{a,b}$ is fraction of $[^4F_{3/2}]$ in a or b.

$$f_a = \frac{1}{1 + \exp[-\Delta E/kT]} = 0.6 \qquad f_b = 1 - f_a = 0.4$$

for

$$kT = 208\ \text{cm}^{-1}$$

As Fig. 10.7 illustrates, the composite cross section peaks between the centers of the A and A' transitions. Thus, we must add the normalized lineshape functions weighted by the fraction in the upper states to obtain an effective stimulated emission cross section at $\lambda = 1.06426\ \mu\text{m}$ of $\sigma_{\text{eff}} = 1.67 \times 10^{-19}\ \text{cm}^2$ and a composite line width of about $6.8\ \text{cm}^{-1}$ (a fit of a Lorentzian to the composite line shape).*

Another "new" issue arising in the YAG system is the use of $^4F_{3/2}$ population by stimulated emission. The rate equations for the (a, b) levels are

$$\frac{dN_{2b}}{dt} = R_{2b} - \frac{N_{2b}}{\tau_2} - N_{2b}\sigma_A \bar{g}_A(\nu)\frac{I_\nu}{h(\nu)} - r_{ba}N_{2b} + r_{ab}N_{2a} \qquad (10.4.7a)$$

$$\frac{dN_{2a}}{dt} = R_{2a} - \frac{N_{2a}}{\tau_2} - N_{2a}\sigma_{A'} \bar{g}_{A'}(\nu)\frac{I_\nu}{h(\nu)} + r_{ba}N_{2b} - r_{ab}N_{2a} \qquad (10.4.7b)$$

where r_{ba}, r_{ab} represent the fast intramanifold relaxation process and are related by detailed balancing arguments (see Appendix I).

$$r_{ab} = r_{ba}\exp[-\Delta E/kT]$$

If these rates are fast, much faster than the natural decay rate $(1/\tau_2)$, or the stimulated rate $(\sigma_{\text{eff}} I_\nu/h\nu)$ or the sum, then the population ratio $N_b/N_a = \exp[-\Delta E/kT]$ even when lasing, and the stimulated emission can use the population in both states. If we add (10.4.7a) and (10.4.7b), the sum $(N_{2a} + N_{2b} = [^4F_{3/2}])$ becomes a rate equation for the entire manifold.

$$\frac{d[^4F_{3/2}]}{dt} = R_2 - \frac{[^4F_{3/2}]}{\tau_2} - \frac{I_\nu}{h\nu}[N_{2b}\sigma_A \bar{g}_A(\nu) + N_{2a}\sigma_{A'}\bar{g}_{A'}(\nu)]$$

$$= R_2 - \frac{[^4F_{3/2}]}{\tau_2}\left(1 + \frac{I_\nu}{I_s}\right) \triangleq R_2 - \frac{[^4F_{3/2}]}{\tau_2}(1 + L) \qquad (10.4.8)$$

where

$$I_s = \frac{h\nu_L}{\sigma_{\text{eff}}\tau_2}\ [(\sigma_{\text{eff}} \text{ defined by (10.4.6)}] \qquad (10.4.9)$$

$$= \text{saturation intensity for the laser at } 1.0643\ \mu m$$

and

$$L = I_\nu/I_s, \text{ the laser intensity divided by } I_s.$$

The $[^4F_{3/2}]$ manifold is pumped by absorption from one of the states in $^4I_{9/2}$ to one or more in the $^4F_{5/2}$ and $^4H_{9/2}$ manifolds. As indicated in Fig. 10.5, we assume pumping to the state at $12{,}370\ \text{cm}^{-1}$. Once the Nd atoms are promoted to there, a fraction η_p (the pumping efficiency) relax to the $[^4F_{3/2}]$ level losing at least $863\ \text{cm}^{-1}$ to the crystalline lattice. Even though the $12{,}370\ \text{cm}^{-1}$ state is directly pumped, we can neglect its population compared

*The cross sections of Table 10.3 yield the gain coefficients in terms of the upper state populations whereas σ_{eff} refers to the total population in $^4F_{3/2}$ manifold.

to that in the lowest level of the $^4I_{9/2}$ manifold if we assume that the relaxation to $^4F_{3/2}$ is fast and also neglect the rate sending atoms back from $^4F_{3/2}$. This upward rate is a small fraction, exp[−863/208] = 0.016, of the downward rate and also indicates that the density in $^4F_{5/2}$ will be no more than ~ 2% of that in $^4F_{3/2}$. Thus R_2 in (10.4.8) is equal to $\eta_p \alpha_p(\lambda) I_p / h\nu_p$ where $\alpha_p(\lambda)$ is the absorption coefficient of the pump.

For these calculations, the absorption data from Fan and Byer [38] which is reproduced in Fig. 10.8 is used for a Nd:YAG sample with the density [Nd] = $1.52 \times 10^{+20}$ cm^{-3}. The attenuation values shown there must be reduced to conform to the density of 1.38×10^{20} cm^{-3} of Table 10.2. The "sticks" labeled with the two numbers represent the centers of transitions from the states in $^4I_{9/2}$ (the first number starting at 0, 1, 2, 3, 4), to one of the states in $^4F_{5/2}$ or $^4H_{9/2}$ (the second number, starting at 1 in $^4F_{5/2}$ and proceeding upward). Notice that the absorption is significant in the vicinity of 808.4 nm corresponding to a transition from 0 to 12,370 cm^{-1}. Thus conversion of the pump intensity to population can be effected in a very short distance of roughly $l_g = 5/\alpha_p$. Using the data shown in Fig. 10.8, we compute the attenuation coefficient for the pump at 808.5 nm for our density to be $\alpha_p(808.5) = 8.1 \text{ cm}^{-1} \times (1.38/1.52) = 7.4 \text{ cm}^{-1}$. Thus most of the pump is absorbed in $l_g \sim 0.68$ cm. Let us choose $l_g = 1.0$ cm to allow for the fact that the pump wavelength may drift off of the peak of the absorption.

For instance, suppose the pump wavelength drifted by 0.5 nm. Then $\alpha_p(809 \text{ nm}) = 3.7 \text{ cm}^{-1}$ and $\exp[-\alpha_p l_g] = 0.025$ or 97.5% of the pump would still be absorbed. If we chose $l_g = 0.68$ cm, then only 92% of the pump would be utilized. Quite often, the temperature of the semiconductor laser and the YAG are controlled to ensure against such a drift.

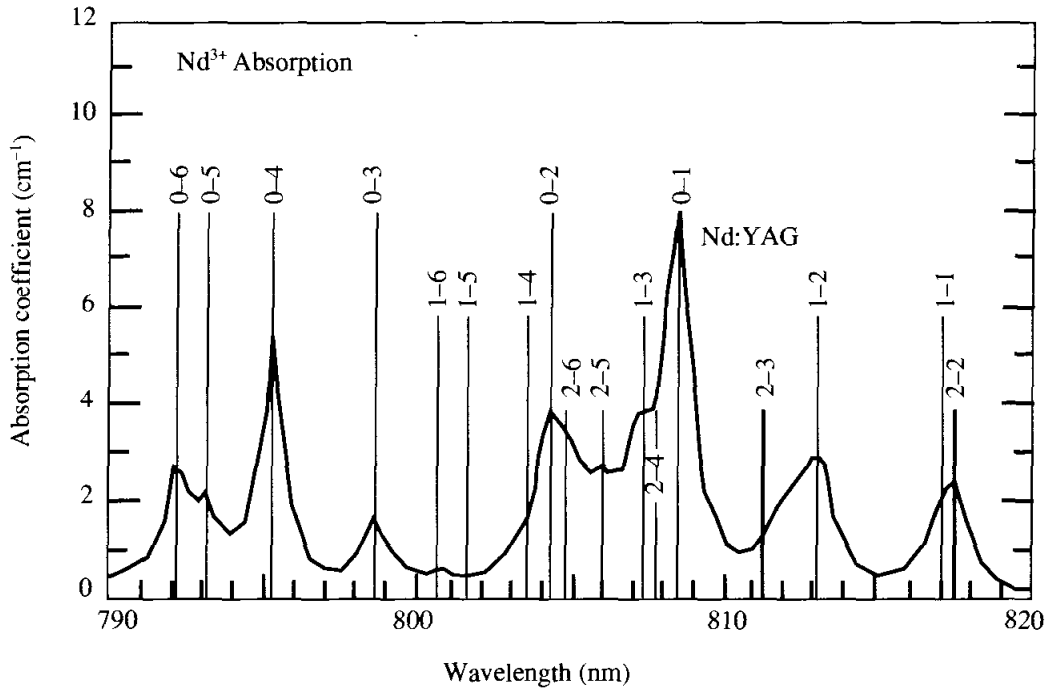


FIGURE 10.8. The absorption of Nd:YAG in the near IR for $[Nd]^0 = 1.52 \times 10^{20}$ cm^{-3}. Data from Fan and Byer [38].

The fact that the pump is attenuated with z indicates that the $[^4F_{3/2}]$ density and thus the small signal gain coefficient will also be a function of z.

$$\frac{d[^4F_{3/2}]}{dt} = \eta_p \frac{\alpha_p I_p(z)}{h\nu_p} - \frac{[^4F_{3/2}]}{\tau_2}[1 + L(z)] \tag{10.4.10}$$

Before proceeding too far, let us try to compress the notation by some obvious identifications, conservation laws, and abbreviations.

$$\alpha_p \triangleq f_0 N_0 \sigma_p(\lambda);$$

$$\sigma_p = \text{absorption cross section} = 1.16 \times 10^{-19}\ \text{cm}^2 \text{ at } 808.5\ \text{nm} \tag{10.4.11a}$$

$$f_0 = \text{fraction of } N_0 = [^4I_{9/2}] \text{ manifold in the lowest state (0.464 at 300 K)}$$

$$N_2 \triangleq \text{total population in } [^4F_{3/2}] \text{ manifold; upper state density} \tag{10.4.11b}$$

$$N_0 + [^4F_{3/2}] = [\text{Nd}]^0;\ \text{the doping density of neodymium in YAG}^* \tag{10.4.11c}$$

$$I_{sp} = \frac{1}{f_0 \eta_p} \frac{h\nu_p}{\sigma_p \tau_2};\ \text{the pump saturation intensity}$$

$$= 18\ \text{kW/cm}^2 \text{ for } 808.5\ \text{nm} (\eta_p \sim 1) \tag{10.4.11d}$$

$$P(z) = I_p / I_{sp};\ \text{the normalized pump intensity} \tag{10.4.11e}$$

$$L(z) = I_L / I_s = \text{the } 1.06\ \mu\text{m intensity normalized to its saturation value}$$

$$I_s = h\nu / \sigma_{\text{eff}} \tau_2;\ \text{the saturation intensity at } 1.064\ \mu\text{m}. \tag{10.4.11f}$$

and the superscripts ^{0} indicates a value in the absence of a strong intensity.

Equation (10.4.10) can be rewritten in terms of those abbreviations:

$$\frac{dN_2}{dt} = \frac{P(z)N_0}{\tau_2} - \frac{N_2}{\tau_2}[1 + L(z)]$$

and thus the CW value for N_2 is given by

$$N_2 = \frac{P(z)N_0}{1 + L(z)} \tag{10.4.12}$$

The sum of the densities in $^4I_{9/2}$ and $^4F_{3/2}$ must be equal to the doping density $[\text{Nd}]^0$

$$N_2 + N_0 = [Nd]^0$$

which leads to

$$N_0 = \frac{[\text{Nd}]^0}{1 + P(z)/[1 + L(z)]} \tag{10.4.13a}$$

*Our assumptions indicate that the active atoms are either in the ground or upper laser state. Thus if $[^4F_{3/2}]$ becomes significant, then the ground state density decreases.

and

$$N_2 = \frac{P(z)[\mathrm{Nd}]^0}{1 + P(z)/[1 + L(z)]} \cdot \frac{1}{1 + L(z)} \tag{10.4.13b}$$

The variation of the pump with z is described by

$$\frac{dP(z)}{dt} = -\{f_0 N_0 \sigma_p\} P(z) = -\frac{f_0[\mathrm{Nd}]^0 \sigma_p P(z)}{1 + P(z)/[1 + L(z)]} = -\frac{\alpha_p^0 \cdot P(z)}{1 + P(z)/[1 + L(z)]} \tag{10.4.14}$$

Combine (10.4.12) with (10.4.13b) to obtain the variation of the laser intensity with z.

$$\frac{dL(z)}{dz} = N_2 \sigma_{\mathrm{eff}} L(z) = [\mathrm{Nd}]^0 \sigma_{\mathrm{eff}} \frac{P(z)}{1 + P(z)/[1 + L(z)]} \cdot \frac{1}{1 + L(z)} \cdot L(z)$$

or

$$\frac{dL(z)}{dz} = \gamma_m \frac{P(z)}{1 + P(z)/[1 + L(z)]} \cdot \frac{1}{1 + L(z)} \cdot L(z) \tag{10.4.15}$$

where $\gamma_m = [\mathrm{Nd}]^0 \sigma_{\mathrm{eff}}$, the maximum possible gain coefficient with all of the Nd atoms in the $^4F_{3/2}$ manifold. If we divide (10.4.15) by (10.4.14), cancel common factors, arrange all of the laser terms on the left, the pump on the right, and integrate from $z = 0$ to $z = l_g$, then considerable simplification is possible.

$$\int_{L_1}^{L_2} \frac{[1 + L(z)]}{L(z)}\, dL = -\frac{\gamma_m}{\alpha_p^0} \int_{P_1}^{P_2} dP \tag{10.4.16}$$

where the subscript 1 refers to the parameters at $z = 0$, the entrance to the gain medium, and the subscript 2 to the exit plane at $z = l_g$. We fervently hope the pump intensity at $z = l_g$ is small compared to the input (good pump utilization) but, in any case, we define

$$\eta_a = \text{absorption efficiency} = \frac{P_1 - P_2}{P_1} = \frac{I_p(1) - I_p(2)}{I_p(1)} \tag{10.4.17}$$

Equation (10.4.16) can be easily integrated to yield

$$\ln\left(\frac{I_2}{I_1}\right) + \frac{I_2}{I_s}\left(1 - \frac{I_1}{I_2}\right) = \frac{\gamma_m}{\alpha_p^0}[P_1 - P_2] = \frac{\gamma_m}{\alpha_p^0} \eta_a \frac{I_p(1)}{I_{sp}} \tag{10.4.18}$$

For self-consistency, we require that

$$I_1 = SI_2, \text{ where } S \text{ is photon survival factor in the passive cavity.} \tag{10.4.19}$$

Now we proceed exactly as was done in Sec. 9.2: identify a threshold pump intensity (where $I_2/I_s \to 0$) in (10.4.18).

$$\frac{\gamma_m}{\alpha_p^0} \cdot \eta_a \frac{I_p(th)}{I_{sp}} = \ln\frac{1}{S} \quad \text{or} \quad I_p(th) = \frac{\alpha_p^0}{\gamma_m} \frac{I_{sp}}{\eta_a} \ln\left(\frac{1}{S}\right) \tag{10.4.20}$$

We use the fact that $I_1/I_2 = S$ (for a CW laser) and recognize that the difference $[I_p(1) - I_p(th)]$ must be greater than zero for the intensity I_2 at 1.06 μm to become significant. Thus

the intensity impinging on mirror M_2 is given by

$$I_2 = \frac{\gamma_m}{\alpha_p^0}\eta_a \frac{I_s}{I_{sp}}\left[\frac{I_p - I_p(th)}{1-S}\right] \tag{10.4.21}$$

The output intensity is just $(1 - R_2)I_2$. The remaining task is to use the definition of the saturation intensity of the pump (10.4.11d) and of the laser (10.4.9) to find that most of the cross sections disappear from the expression.

$$\begin{aligned} I_{\text{out}} &= \eta_p \cdot \eta_a \cdot \eta_{qe}\left[\frac{1-R_2}{1-S}\right]\cdot\{I_p - I_p(th)\} \\ &= \eta_p \cdot \eta_a \cdot \eta_{qe} \cdot \eta_c \cdot \{I_p - I_p(th)\} \end{aligned} \tag{10.4.22}$$

where $$\eta_c = \frac{1-R_2}{1-S} = \frac{\text{coupling loss}}{\text{total loss}} = \text{coupling efficiency}$$

$$\eta_{qe} = \frac{\nu_L}{\nu_p} = \frac{\lambda_p}{\lambda_L} = 0.762, \text{ the quantum efficiency,}$$

and

$$I_p(th) = \frac{\alpha_p^0}{\gamma_m}\frac{I_{sp}}{\eta_a}\ln\left(\frac{1}{S}\right) = \frac{I_s}{\eta_{qe}\eta_p\eta_a}\ln\left(\frac{1}{S}\right) \tag{10.4.23}$$

If, for instance, we assume a pumping efficiency of $\eta_p = 90\%$, a pump utilization efficiency $\eta_a = 100\%$, $I_s = 4.4$ kW/cm^2, and if the survival factor for the passive cavity $S = 0.9$, then the threshold pump intensity would be 676 W/cm^2. For a mode size of 500 μm diameter, we find a threshold pump power of 1.33 W, quite reasonable and attainable with laser diodes. Once threshold is exceeded, then the excess pumping appears as output reduced by the products of the various efficiencies as indicated by (10.4.22).

10.4.4 Neodymium-Glass Lasers

Glass, being a technology that has been studied since ancient times, can be made very uniform, doped with a variety of donor substances, polished to fantastic accuracy, and cast into large ingots. The details for Nd^{3+} in a typical glass host are shown in Fig. 10.9 with the energy-level diagram shown in Fig. 10.9(a), a typical spontaneous emission profile for the dominant laser transition in various types of glass in Fig. 10.9(b), and the absorption of ED-2 glass shown in Fig. 10.9(c).

Some of the more important physical details of the fluorescence from the $^4F_{3/2}$ manifold (necessary to solve the problems) are given in Table 10.4. Because of the amorphous character of glass, the absorption bands are much broader than those of YAG, which is good. However, this same characteristic leads to wide fluorescent lines and thus to a smaller stimulated emission cross section that, to some, may appear undesirable.

However, gain is not the only criterion for the "goodness" of the medium. If one is dealing with an optical amplifier, for instance, (which is the main use of glass), then energy storage in the population inversion is of primary concern. Because of the high doping

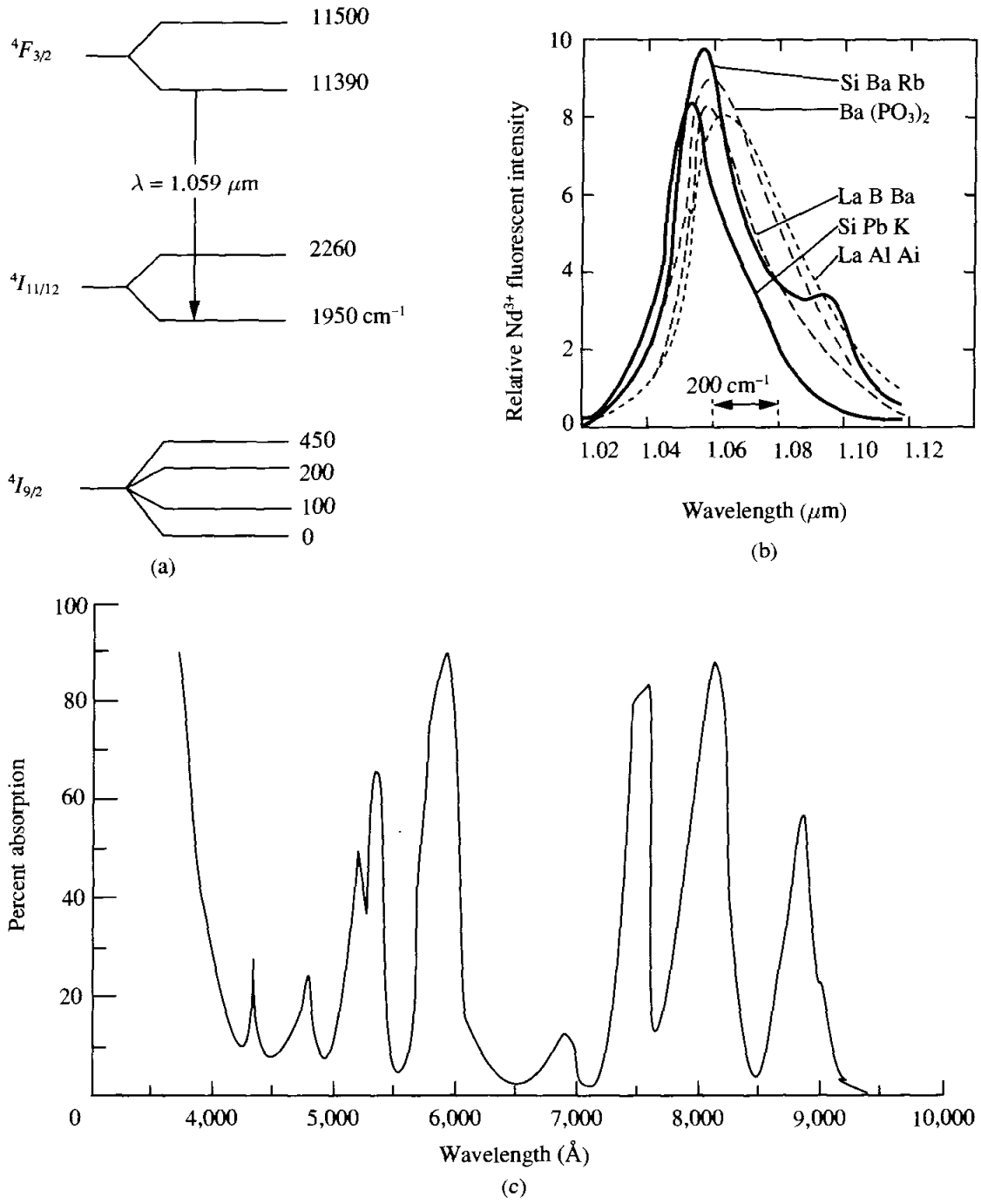


FIGURE 10.9. Neodymium-glass laser. (a) Diagram of the manifold showing the laser levels. (b) Spontaneous emission profile of Nd^{3+} in various glasses. (c) Absorption of a 6.3 mm slab of ED-2 glass. (From Koechner. Ref. 24.)

TABLE 10.4 Properties of Common Neodymium-Glass Systems

Glass Designation	ED-2	LG-630	LCG-11	LGH-5	EV-2
Type	Silicate	Silicate	Silicate	Phosphate	
$[Nd^{3+}]$ (10^{20} cm^{-3})	2.8	2.8	3.8	3.17	3.1
Index	1.555	1.509	1.546	1.531	1.504
λ (μm)	1.0623	1.0623	1.0624	1.0560	1.054
$A_T(^4F_{3/2}$ in 10^3 sec $^{-1})$	3.33	1.56	1.74	3.45	1.77
$\Delta\lambda$ (FWHM in Å)	260	220	290	186	170
σ (10^{-20} cm^2)	3.03	2.1	2.0	3.9	4.7
Lifetime ($^4I_{11/2}$)*	~ 11 ns	(50–100 ns)			

*Data from Koechner [24] and from Owens-Illinois. See also Brown [26, p.68], which contains data on many more glasses.

densities permitted in glass, with each active atom storing $h\nu/2$ joules of energy, the glass amplifiers have a very high figure of merit. Furthermore, the wide band width over which gain is significant permits the amplification of very short pulses, with tens of picoseconds being common.

One last comment should be made about glass: It is a very poor thermal conductor, about a factor of 40 less than that of YAG. It is difficult to remove the waste heat, and consequently the repetition rate of glass lasers is severely limited.

10.4.6 Erbium-Doped-Fiber-Amplifiers

The addition of a small amount of erbium to an optical fiber results in a system capable of amplifying over significant bandwidths in the vicinity of the important 1.55 μm spectral region. Thus the insertion of such an amplifier in a fiberoptic link can compensate for the residual loss introduced by the much longer nonamplifying link.

To appreciate how important this is to long-haul (many km) telecommunications, consider the alternatives shown in Fig. 10.10, where it is presumed that the net attenuation between terminals A and B is very large, so large that something has to be done at intermediate distances. Consider a specific case: Suppose the net power sent by A towards B at 1.55 μm were 10 mW (+10 dBm) and that the distance between the stations were 300 km (~ 187 miles). While the loss in the fiber might be only ~ 0.2 dB/km (see Fig. 4.6), the total attenuation is 60 dB and thus the received power would be only −50 dBm (10 nW) if we were to go directly from A to B. While some detectors can detect such small power, there are definite limitations appearing. For instance, in 1 ns, there are only 78 photons arriving (which indicates a bit-rate limitation) and even if each photon generates an electrical carrier, the detected current would be only 12.5 nA.

A better "system" would be to detect the signal before it gets too small, say at the repeater stations RS1 and RS2 spaced at the 100 km points as shown in Fig. 10.10(b), where the digital optical signal is detected, converted to high speed electronic logic, and the optical signal is regenerated and sent on its way. The signal would only be down by 20 dB (to −10 dBm) and the current generated by a perfect detector is $P_{opt}/(h\nu/e) = 125$ microamps.

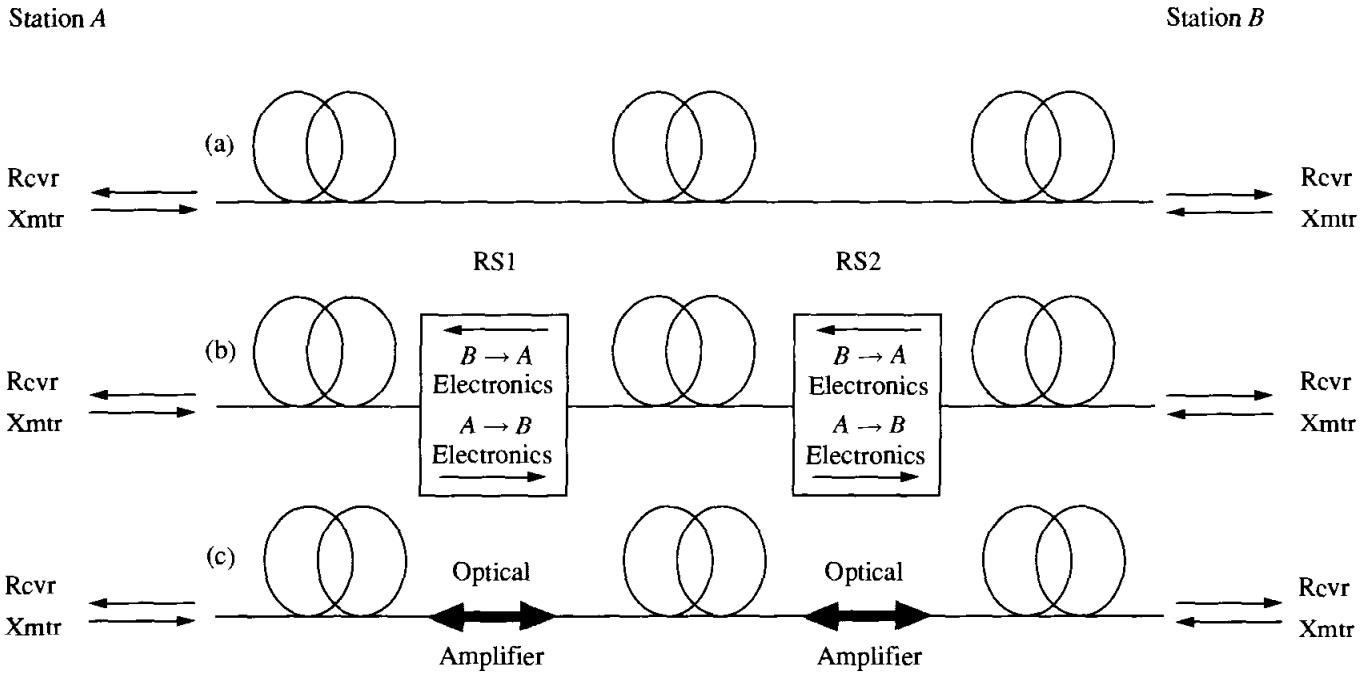


FIGURE 10.10. A long haul data link between stations A and B: (a) direct link; (b) detection and regeneration at the repeater stations; (c) optical amplifiers.

To regenerate the signal and send it on toward RS2 at the 10 mW level, we must amplify the 125 μA back toward the level required of the RS1 laser transmitter (about tens of mA) and the digitally coded information be re-established. The process is then duplicated at repeater R2. Note that B would also be sending information to A. Hence, the above scenario must be duplicated for the $B \rightarrow A$ link.

The use of an optical amplifier at RS1 and RS2 greatly simplifies and enhances the performance of the system. Not only is the electronics greatly simplified (only a DC power supply is needed for the laser pump) but an optical amplifier is naturally bilateral and thus will amplify the $B \rightarrow A$ signals just as well as $A \rightarrow B$. Thus only one pump, one power supply, and no fast regenerative electronics for the two directions are needed. All of these features can be obtained at 1.55 μm with an erbium-doped-fiber-amplifier (EDFA).

The energy level diagram for the rare earth erbium in YAG and in a silica host, which is codoped with aluminum and phosphorus, is shown in Fig. 10.11.* Erbium in these hosts acts as if it were triply ionized (similar to neodymium) with the $2J + 1$ levels in each manifold split and shifted by the local electric field owing to the Stark effect. For the YAG host, the site-to-site variation for the dopant is small, and thus the various split levels are distinct and have been identified rather precisely by careful spectroscopy.

The levels in the glass host are not yet identified to the precision of those in YAG because there is a site-to-site variation in the local field that shows up in a variation of the levels and various broadening processes, homogeneous and inhomogeneous in nature, conspire to wash out the line structure leaving a more or less continuous band of wavelengths

*The purpose of aluminum and phosphorus is to increase the solubility of erbium before clustering occurs. See the reference by Miniscalco [53].

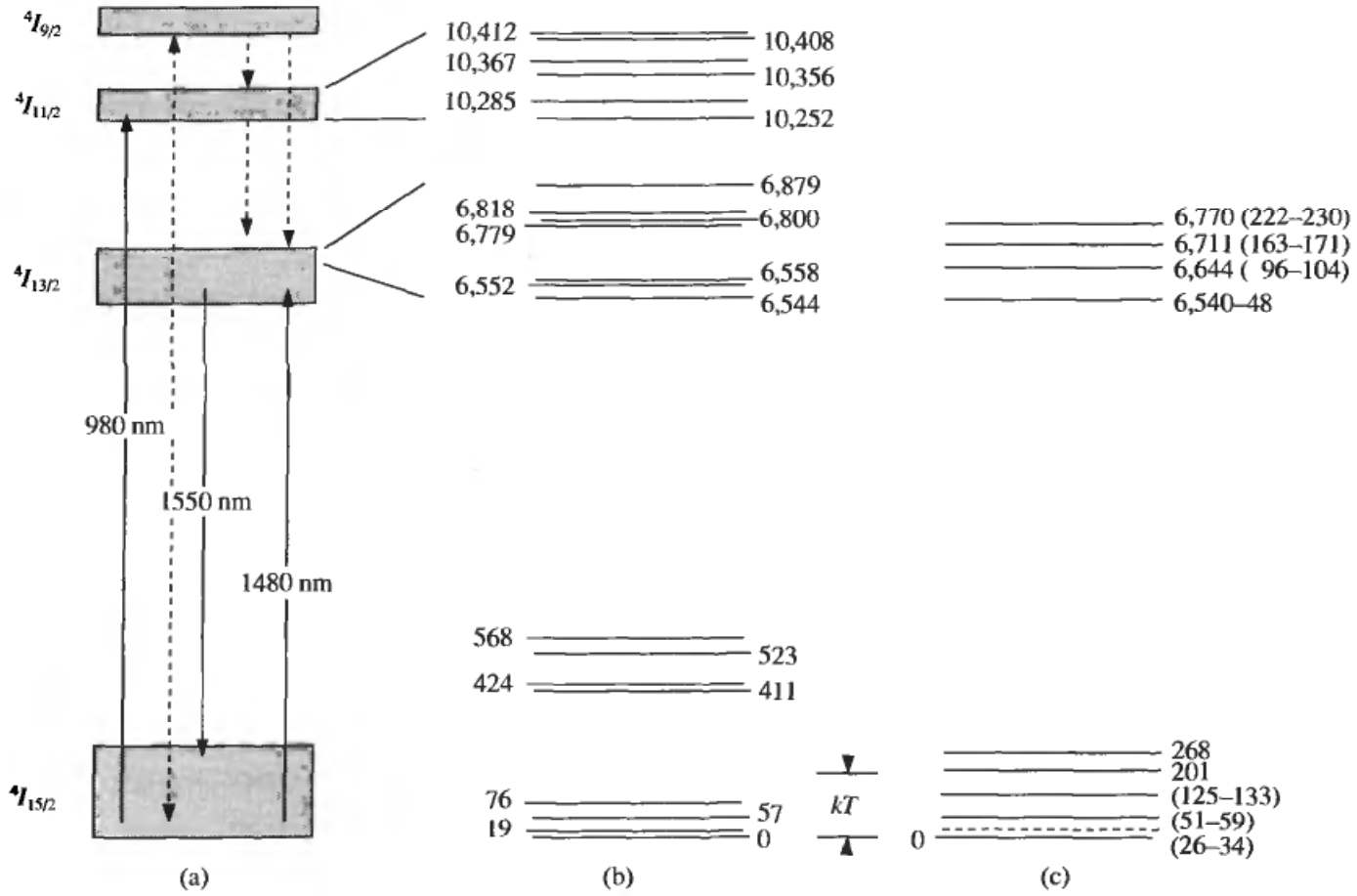


FIGURE 10.11. (a) Simplified energy level diagram of erbium in a solid host; (b) the Stark levels of Er^{3+} in YAG (data from Table 4.9 of Kaminski [39]); (c) Stark levels in aluminosilicate glass (data from Desurvire and J. Simpson [40]).

connecting the two manifolds. Those levels shown in Fig. 10.11(c) were established by careful spectroscopy (Simpson and Desurvire [40]) by using the spectral dependence of both emission and absorption as a function of temperature.

We might expect $(2J + 1)/2$ to be eight levels* in the $^4I_{15/2}$ manifold and seven in the $^4I_{13/2}$ manifold leading to 56 possible transitions as is shown for Er : YAG. Miniscalco [53] suggests that inhomogeneous processes contribute a 40-50 cm^{-1} (120-150 GHz or $\sim$ 10 nm) to the width and homogeneous ones contribute about the same. Thus we obtain a significant spectral bandwidth from the overlapping of many different transitions. Thus, the EDFA can be considered a 2 or 3 manifold or a 20 level system (with the last possibility inducing considerable dismay).

Let us focus on Fig. 10.11(a) for a moment. Owing to the availability of efficient InGaAs semiconductor lasers at 980 nm, we can use that energy to excite the erbium atoms to the $^4I_{11/2}$ manifold since those atoms relax primarily back to the $^4I_{13/2}$ manifold.

*Remember the Stark shift is proportionate to the magnitude of the electric field $|E|$. Hence, for a low symmetry site for an odd number of electrons the $2J + 1$ levels will split into $(2J + 1)/2$ states with each being doubly degenerate (see Kaminski [39] p. 120).

Amplification takes place between the low lying levels of the $^4I_{13/2}$ manifold and the high lying ones in $^4I_{15/2}$, and residual absorption remains between the opposite extremes of the two manifolds. Hence, we can also pump the system at $\lambda \approx 1480$ nm and amplify at 1550 nm.

The combination of an overlap of the broadened 20 to 56 possible transitions between the two manifolds leads to a practical advantage and a key simplification. The practical advantage is that many different wavelengths can be amplified simultaneously, and hence wavelength division multiplexing (WDM) is possible. The simplification is that we can use the experimentally determined absorption and emission cross sections,* which automatically accounts for the multiple levels and large number of possible transitions. Such data are shown in Fig. 10.12.

These cross sections can be applied directly to the gain coefficient.

$$\gamma(\nu) = N_2\sigma_{em}(\nu) - N_1\sigma_{ab}(\nu) \tag{10.4.24}$$

where N_1 is the density of Er^{3+} ions in the $^4I_{15/2}$ manifold and N_2 is that in the $^4I_{13/2}$ one. The only serious assumption used in deriving (10.4.24) is that the time scale for establishing a Boltzmann population within each manifold is fast compared to the time scale for relaxation between them.

If we pump at $\sim$ 1480 nm, then $N_1 + N_2 = N$ is the doping density of erbium, and we can easily compute the threshold pump power to establish optical transparency.

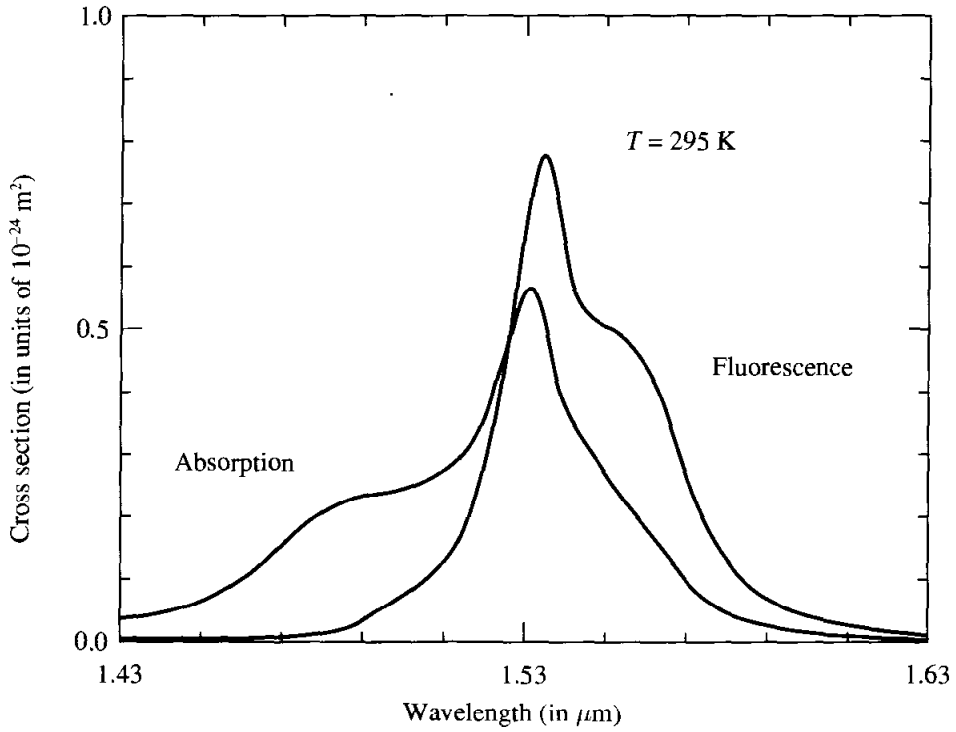


FIGURE 10.12. The experimental emission (fluorescence) and absorption cross section of the aluminosilicate fiber glass (data from Desurvire and J. Simpson [40]).

*The origin of the fact that $\sigma_{em}(\nu) \neq \sigma_{ab}(\nu)$ will be discussed in Sec. 10.5.

For optical transparency at ν, $\gamma(\nu) = 0 = N_{2t}\sigma_{em}(\nu) - N_{1t}\sigma_{ab}(\nu)$

$$\therefore \frac{N_{2t}}{N_{1t}} = \frac{\sigma_{ab}(\nu)}{\sigma_{em}(\nu)} \qquad N_{1t} + N_{2t} = [N] \text{ or } N_{1t} = [N] - N_{2t} \tag{10.4.25}$$

$$\therefore N_{2t} = N\frac{\sigma_{ab}(\nu)}{\sigma_{em}(\nu) + \sigma_{ab}(\nu)} \tag{10.4.26}$$

Manifold 2 ($^4I_{13/2}$) decays primarily by radiation back to the ground state ($^4I_{15/2}$) at a rate of $1/\tau_2$, and thus the pump must supply this fluorescence power.

$$\frac{P}{\text{vol}}(\text{opt. transp.}) = h\nu\frac{N}{\tau_2}\frac{\sigma_{ab}(\nu)}{\sigma_{em}(\nu) + \sigma_{ab}(\nu)} \tag{10.4.27}$$

Note that this condition is frequency (wavelength) dependent, but not drastically so, for the long wavelength side of Fig. 10.12.

If we pump at 980 nm, which appears to be the favorite practical route, then we should also include the population in the $^4I_{11/2}$ manifold in the conservation of atoms equation. However, its lifetime is very short (3.0 μs) compared to that of the $^4I_{13/2}$ manifold ($\sim$ 10 ms) [48]. Hence we can ignore the population in 3 and use (10.4.26) anyway. We must multiply (10.4.27) by the quantum efficiency $\lambda_{amp}/\lambda_{pump}$ and the branching efficiency to predict the threshold for optical transparency. Data for the Al/P silica fiber is given in Table 10.5

The stimulated emission cross section is small, as is the doping density, and thus tens of meters of doped fibers are needed to establish the 30 to 40 dB gain required to compensate for the fiber loss. Such amplifiers are, by necessity, pumped from one end or the other (or both), and thus the 980 nm (or 1480 nm) must be injected into the fiber through a coupler. Both the signal and the pump interact with the populations in the large signal domain, and hence saturation phenomena is a bit more complicated than encountered to date. The fact that the $^4I_{13/2}$ lifetime is so long permits the use of WDM without cross talk. (The analysis of both issues is reserved for problems.)

TABLE 10.5 Al/P Silica Fiber Data

	Value	Units
$^4I_{13/2}$ lifetime	10.2	ms
Peak emission λ_{em}	1530.8	nm
peak emission cross section	5.7×10^{-21}	cm^2
Peak absorption λ_{ab}	1530.1	nm
peak absorption cross section	6.6×10^{-21}	cm^2
Gain bandwidth (FWHM)	43.3	nm
(20 dB peak, fully inverted)	7.9	nm
980 absorption peak	978.6	nm
peak absorption cross section	3.12×10^{-21}	cm^2
FWHM of absorption	16.6	nm

From Miniscalco [53].

10.5 BROAD-BAND OPTICAL GAIN

10.5.1 Band-to-Band Emission and Absorption

The material on EDFA introduced us to the fact that some systems can amplify over a broad band of optical frequencies, a highly desirable situation from a practical standpoint in that it allows us to use one gain medium to amplify many wavelengths or to construct a tunable oscillator with the proper feedback. However, the data presented in Fig. 10.12 and the analysis of Sec. 10.4.6 introduced (and glossed over) a conceptual problem that is apparently at odds with the elementary atomic theory used to date.

For instance, the stimulated emission cross section for the EDFA of Fig. 10.12 is *not* equal to absorption cross section (except at one wavelength where the curves cross). We cannot explain this fact by a degeneracy ratio in the manner of Eq. 7.7.3, since $\sigma_{em}(\nu)/\sigma_{ab}(\nu)$ is also a function of frequency. Have we reached the limit of applicability of the Einstein A and B coefficient approach (which led to the relationship $\sigma_{em}/\sigma_{ab} = g_1/g_2$)?

No, the apparent problem arises because we are considering transitions between bands or manifolds of states and by using experimental absorption measurements to define $\sigma_{ab}(\nu)$ and fluorescence variations to define $\sigma_{em}(\nu)$. The simple sketch of Fig. 10.13 illustrates why the emission from the $^4I_{13/2}$ manifold to the ground $^4I_{15/2}$ states of erbium will not match the absorption variation. Shown in Fig. 10.3(a) is the case of a system—say a EDFA—with minimal (or no) pumping and thus $[^4I_{13/2}] \ll [^4I_{15/2}]$. Within both manifolds the population is distributed according to the Boltzmann relationship, but the upper state manifold is essentially empty. Hence the absorption coefficient as a function of frequency will reflect the relative population of the lower manifold. Because of the Boltzmann factor there will

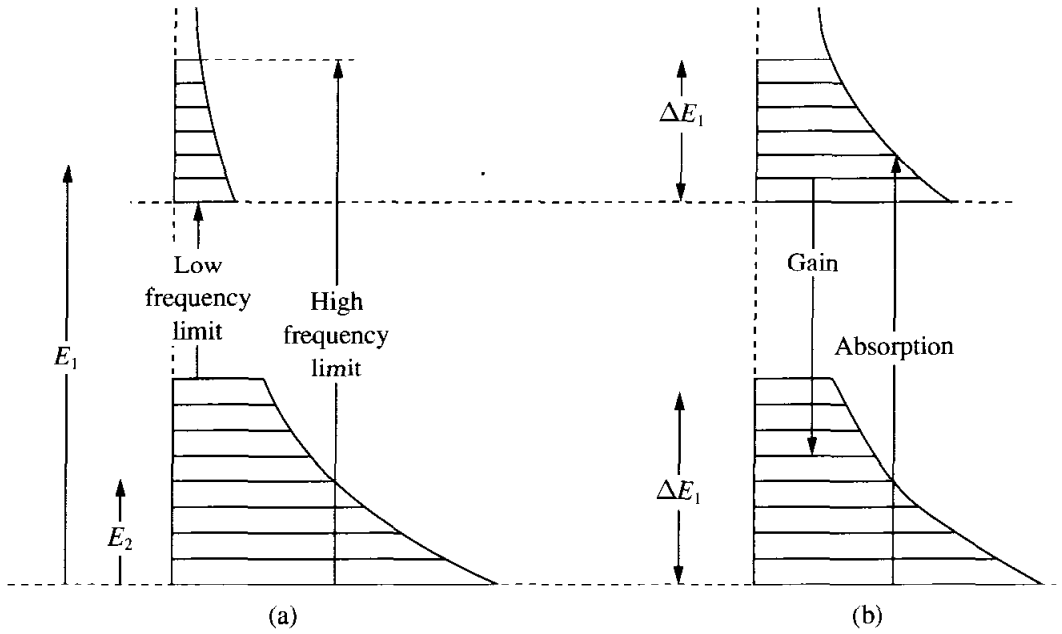


FIGURE 10.13. A sketch showing transitions between two bands whose populations are given by the Boltzmann factor for states within a band. The system is in (a) equilibrium and (b) pumped.

be small absorption at the low frequency limit since the population in the highest state of the lower manifold is down by $\exp[-(\Delta E_1/kT)]$ compared to the bottom.

If the system is pumped so that atoms populate the upper manifold in a Boltzmann fashion, then there is the possibility of absorption on the high frequency (short wavelength) limit with even gain at the other extreme as shown in Fig. 10.13(b). Provided the selection rules are obeyed, there can be considerable fluorescence at the low frequency (long wavelength limit) with smaller amounts at the high frequency (short wavelength) limit where the absorption can still be significant. Thus the spectral dependence of the emission and absorption can be quite different.

If we knew the energy levels, the transition probabilities between all combinations of states, and the broadening for each transition, then we could predict $\sigma_{em}(\nu)$ and $\sigma_{ab}(\nu)$ by summing over all combinations of E_2 and E_1 that yield the same photon energy $h\nu = E_2 - E_1$. Whether the transition is from $E_2 \rightarrow E_1$ by spontaneous and/or stimulated emission yielding a photon $h\nu$, or from $E_1 \rightarrow E_2$ by the absorption of the photon would not be important. However, this theoretical route is seldom open, but the path of measurement of absorption and emission is always available, and thus it is appropriate to examine this situation in greater detail.

Our goal is to establish a general relationship between the emission and absorption cross sections. Although we will choose the simple band model involving a continuous Boltzmann distribution of states within the two manifolds, the primary result is applicable to a very general set of circumstances involving phonon assisted transitions (see the discussion on the alexandrite and Ti:sapphire lasers) and to systems involving changes in atomic coordinates (see discussion of the dye laser and Ti:sapphire). It could be adapted to semiconductor lasers or to a molecular system in a gas, but they will be considered separately to accommodate their selection rules.

Along the way we will have the various opportunities to check our calculations by invoking some cases with known limits.

1. All populations must be related by the Boltzmann factor if the system were in a local thermodynamic equilibrium (LTE) environment.
2. In that case, the rate of spontaneous and stimulated emission at ν must *exactly* balance the amount absorbed from the blackbody radiation.
3. If the widths of the bands approach zero, then we must recover the atomic system results.

10.5.2 Theory of Band-to-Band Emission and Absorption

An analysis of a real system with a large number of transitions at different wavelengths, each having different widths and different transition probabilities, and overlapping lines is exceedingly tedious. For instance, we could expect 56 different transitions for the ${}^4I_{13/2} \rightarrow {}^4I_{15/2}$ in the Er:YAG crystal of Fig. 10.11(b) and 20 for alumino-silicate case shown in Fig. 10.11(c). The combination of a large number of levels, site-to-site variation in the

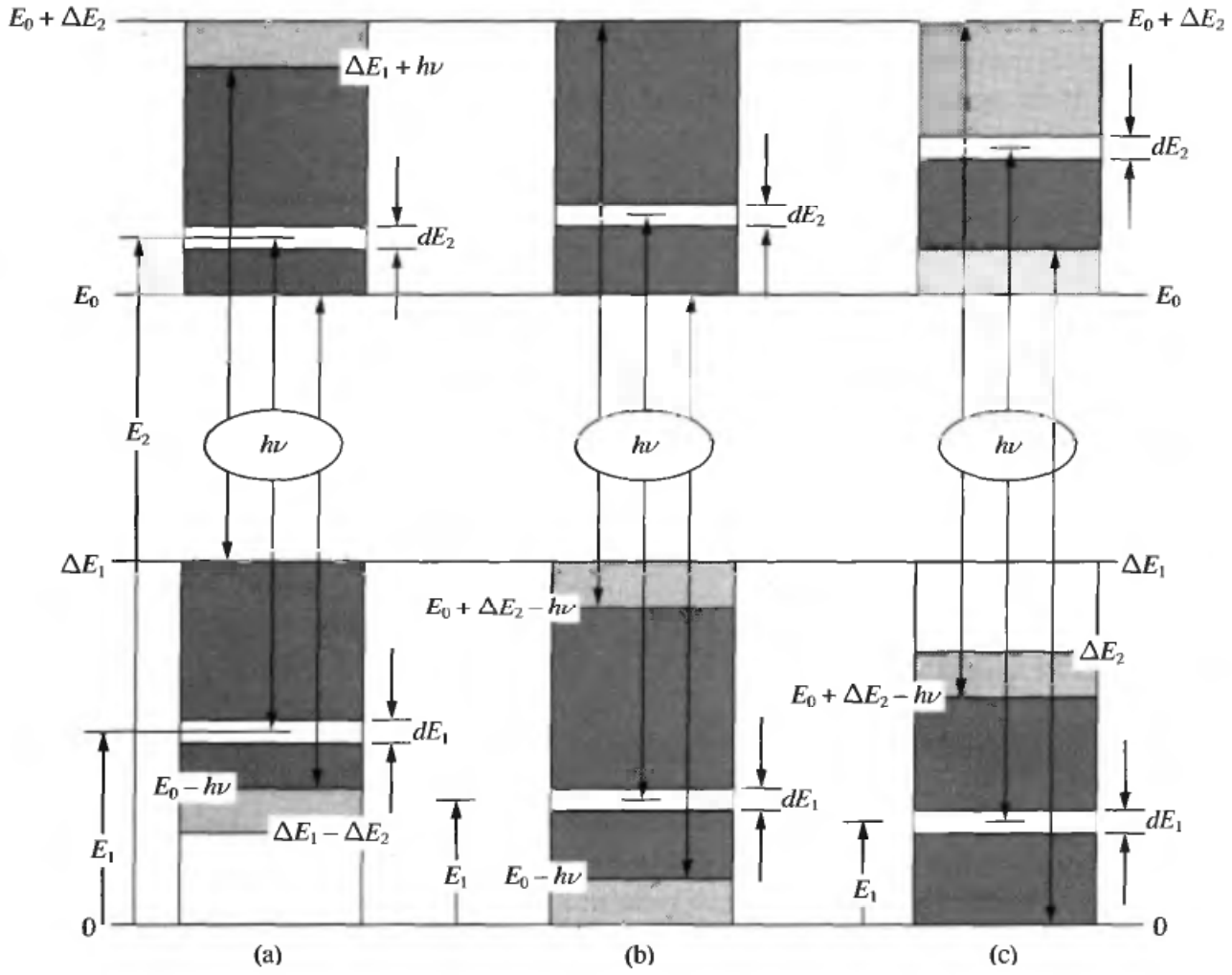


FIGURE 10.14. The energy levels involved in the emission or absorption of a photon of energy $h\nu$. (a) presumes a photon energy defined by $E_0 - \Delta E_1 < h\nu < E_0 - \Delta E_1 + \Delta E_2$. (b) is for $E_0 - \Delta E_1 + \Delta E_2 < h\nu < E_0$. (c) $E_0 < h\nu < E_0 + \Delta E_2$ with the limits shown by the light shading. The extremes on the darker parts are the limits of integration for each case.

energy of each state, and a broadening comparable to the spacing between transitions leads one to consider a simpler system depicted in Fig. 10.14 showing two manifolds whose lowest states are separated by the difference E_0.

To make the problem simple and tractable, we assume that the (g_1, g_2) states of each manifold are smeared into a continua of widths ΔE_1 and ΔE_2 with a Boltzmann distribution within each maintained by some very fast thermally driven process. Let $N^0_{(2,1)}$ denote the density per *quantum level* at the respective band edges so that the total density of atoms in manifold 2 is

$$N_2 = N_2^0 \int_{E_2=E_0}^{E_2=E_0+\Delta E_2} \frac{g_2}{\Delta E_2} \exp -\left[\frac{E_2 - E_0}{kT}\right] d(E_2 - E_0)$$

$$= N_2^0 g_2 \frac{kT}{\Delta E_2}(1 - e^{-\Delta E_2/kT}) \qquad \text{or} \qquad N_2^0 = \frac{N_2}{g_2}\frac{1}{Z_2(T)}\frac{\Delta E_2}{kT} \tag{10.5.1a}$$

where N_2 is the total density in manifold 2 and $Z_2(T) = (1 - e^{-\Delta E_2/kT})$. (10.5.1b)

Note that if $\Delta E_2 \ll kT$, then the density in 2 is simply $N_2^0 \cdot g_2$ as we would expect. The analogous quantities for state 1 are

$$N_1^0 = \frac{N_1}{g_1} \frac{1}{Z_1(T)} \frac{\Delta E_1}{kT} \tag{10.5.2a}$$

with

$$Z_1(T) = (1 - e^{-\Delta E_1/kT}) \tag{10.5.2b}$$

Conservation of atoms requires that $N_2 + N_1 = N =$ the total density of active atoms.

If the system were in an LTE environment, even the populations of the manifolds would be described by a Boltzmann factor, and thus the total number of atoms is given by

$$N = N_{eq}^0 \left\{ \int_0^{\Delta E_1} \frac{g_1}{\Delta E_1} e^{-E/kT}\, dE + \int_{E_0}^{E_0+\Delta E_2} \frac{g_2}{\Delta E_2} e^{-E/kT}\, dE \right\}$$

$$= N_{eq}^0 \left\{ g_1 \frac{kT}{\Delta E_1} Z_1(T) + e^{-E_0/kT} g_2 \frac{kT}{\Delta E_2} Z_2(T) \right\} = N_{1eq} + N_{2eq}$$

where

$$N_{1eq} = N_{eq}^0 g_1 \frac{kT}{\Delta E_1} Z_1(T) \qquad \text{and} \qquad N_{2eq} = N_{eq}^0 g_2 \frac{kT}{\Delta E_2} Z_2(T) e^{-E_0/kT} \tag{10.5.3}$$

Note that

$$\frac{N_{2eq}}{N_{1eq}} = \frac{g_2}{g_1} e^{-E_0/kT} \left[\frac{\Delta E_1}{\Delta E_2} \frac{Z_2}{Z_1} \right] \tag{10.5.4}$$

where the subscript eq denotes an LTE relation. An easy check on the development is to consider the limit of $\Delta E_{2,1} < kT$, which is the normal situation encountered in atomic systems. For such a case, the bracket approaches 1 and we recover our usual result.

In such a system, transitions are possible at photon energies between $(E_0 - \Delta E_1) < h\nu < (E_0 + \Delta E_2)$, which is conveniently broken into three intervals according to the situations shown in Fig. 10.14.

$$E_0 - \Delta E_1 < h\nu < E_0 - \Delta E_1 + \Delta E_2 \tag{10.5.5a}$$

$$E_0 - \Delta E_1 + \Delta E_2 < h\nu < E_0 \tag{10.5.5b}$$

$$E_0 < h\nu < E_0 + \Delta E_2 \tag{10.5.5c}$$

A complicating issue is the fact that the same photon energy $h\nu$ can be produced by the emission from many different states in manifold (2) and can be absorbed by many different states in (1) as shown by the multiple uses of the same populations in Fig. 10.14. Furthermore, the transition probability (i.e., the A and B coefficients) most likely depends upon E_2 in manifold 2 and E_1 in 1 because of some selection rule (but $h\nu = E_2 - E_1$ in any case). Consider the interval specified by (10.5.5a): the spontaneous emission from atoms in the interval $E_2 + dE_2$ making a transition to $E_1 + dE_1$ into 4π directions producing photons in the frequency interval between ν and $\nu + \delta\nu$ is given by

$$Sp(\nu)\delta\nu = \int_{E_2=E_0}^{E_2=h\nu+\Delta E_1} [h\nu\delta\nu] A_{21}(E_2, E_1) \left[N_2^0 \frac{g_2}{\Delta E_2}\, dE_2 \right] e^{-(E_2-E_0)/kT}$$

$$Sp(\nu)\delta\nu = \int_{E_2=E_0}^{E_2=h\nu+\Delta E_1} [h\nu\delta\nu]A_{21}(E_2, E_1)\frac{N_2}{Z_2(T)} \times e^{-(E_2-E_0)/kT}\, d[(E_2 - E_0)/kT] \tag{10.5.6}$$

where (10.5.1) has been used to express the result in terms of the total density in manifold 2. The stimulated emission rate by photon energy in this same interval ν to $\nu + \delta\nu$ is given by

$$\begin{aligned} St(\nu)\delta\nu &= \int_{\text{low}(2)}^{\text{up}(2)} [h\nu\delta\nu]B_{21}(E_2, E_1)\rho(\nu)\left[N_2^0\frac{g_2}{\Delta E_2}\, dE_2 e^{-(E_2-E_0)/kT}\right] \\ &= \int_{\text{low}(2)}^{up(2)} [h\nu\delta\nu]B_{21}(E_2, E_1)\rho(\nu)\frac{N_2}{Z_2(T)} e^{-(E_2-E_0)/kT}\, d[(E_2 - E_0)/kT] \end{aligned} \tag{10.5.7}$$

where $\rho(\nu) = I(\nu)/(c/n_g)$ is energy density (joules L^{-3}) per unit of frequency. The absorptive rate is proportional to the density in dE_1 at E_1 in the lower manifold

$$\begin{aligned} Ab(\nu)\delta\nu &= \int_{\Delta E_1=\text{low}(1)}^{E_1-h\nu=\text{up}(1)} [h\nu\delta\nu]B_{12}(E_2, E_1)\rho(\nu)\left[N_1^0\frac{g_1}{\Delta E_1}\, dE_1 e^{-E_1/kT}\right] \\ &= \int_{\text{low}(1)}^{\text{up}(1)} [h\nu\delta\nu]B_{12}(E_2, E_1)\rho(\nu)\frac{N_1}{Z_1(T)} e^{-E_1/kT}\, d[E_1/kT] \end{aligned} \tag{10.5.8}$$

Special features and various notations are used in (10.5.6) to (10.5.8): the upper and lower limits for the (2,1) manifolds are to be chosen from Fig. 10.14(a) to Fig. 10.14(c), where the dark shaded region indicates the range of states involved in a transition for a chosen $h\nu$, and the notation $A_{21}(E_2, E_1)$ signifies that the transition strength may depend upon where E_2 lies in the manifold 2 as well as where E_1 is in 1.

Equations (10.5.6) to (10.5.8) hold for all circumstances with the only stipulation being that the relative populations within the manifolds are given by the Boltzmann relation. In particular, they can be used for a system in LTE in which case the following conditions apply:

1. $N_2 \to N_{2eq}$ and $N_1 \to N_{1eq}$ and the two densities are related by (10.5.4).
2. The spectral distribution of radiation becomes the Planck blackbody value.

$$\rho(\nu) \to \rho_{eq}(\nu) = \frac{8\pi n^2 n_g}{c^3}\frac{h\nu^3}{e^{h\nu/kT} - 1} \qquad \text{Recall (7.2.10)}$$

3. The net change of population owing to the three radiative processes is identically zero.

$$[Sp_{eq}(\nu) + St_{eq}(\nu) - Ab_{eq}(\nu)]\delta\nu \equiv 0 \tag{10.5.9}$$

4. Not only does the algebraic sum of the integrals have to be zero, the sum of the integrands must also be zero for each dE around (E_2, E_1) with $E_2 - E_1 = h\nu$ in order to maintain a Boltzmann distribution among all of the states. (See Appendix II on detailed balancing.)

If we substitute the population ratio indicated by (10.5.4), shift the range of integration on (10.5.6) and (10.5.7) by using the fact that $E_2 = E_1 + h\nu$ and $dE_2 = dE_1$, and extract the frequency dependent part of the integrands, we require that

$$A_{21}\left[\frac{g_2/\Delta E_2}{g_1/\Delta E_1}\right]e^{-h\nu/kT} + B_{21}\left[\frac{g_2/\Delta E_2}{g_1/\Delta E_1}\right]e^{-h\nu/kT}\rho_{eq}(\nu) - B_{12}\rho_{eq}(\nu) \equiv 0 \quad (10.5.10)$$

in order for detailed balancing to be satisfied. The solution for $\rho_{eq}(\nu)$ from (10.5.10) and a comparison with the Planck expression (7.2.10) leads to the following familiar relationships:

$$B_{12}(E_2, E_1) = \left[\frac{g_2/\Delta E_2}{g_1/\Delta E_1}\right]B_{21}(E_2, E_1) \quad (10.5.11a)$$

$$\frac{A_{21}(E_2, E_1)}{B_{21}(E_2, E_1)} = \frac{8\pi n^2 n_g}{c^3}h\nu^3 \quad (10.5.11b)$$

Equations (10.5.11a) and (10.5.11b) are generalizations of (7.3.10a) and (7.3.10b) and are specific to the case addressed here with smearing of the degenerate levels into a uniform continua of widths $\Delta E_{2,1}$. In a real case, all of the integrals become a summation and (10.5.11a) retreats to (7.3.10a) for each transition.

While these facts have been established with the help of an LTE environment, the relations expressed by (10.5.11) are associated with the atom and not the environment. Hence, they hold under all circumstances; that is, indeed they must hold when the manifolds are far from LTE and the system is irradiated by a spectrally pure signal.

$$\rho(\nu) \rightarrow g(\nu)\rho_\nu \qquad \text{and} \qquad \rho_\nu = \frac{I_\nu}{c/n_g} \quad (10.5.12)$$

where $g(\nu)$ is the line shape centered at $h\nu = E_2 - E_1$ and I_ν is the intensity of that monochromatic wave. The net production (which may be negative, implying that we have lost photons) of new photons is given by the difference between (10.5.7) and (10.5.8) with the above facts inserted.

$$\begin{aligned} St(\nu) - Ab(\nu) = {} & \int_{\text{low}(2)}^{\text{up}(2)} h\nu \left\{B_{21} = \frac{c^3}{8\pi n^2 n_g}\frac{A_{21}(E_2, E_1)}{h\nu^3}\right\} \\ & \times \left\{\rho_\nu = \frac{I_\nu}{c/n_g}\right\} g(\nu) \cdot \frac{N_2}{Z_2(T)} \cdot e^{-(E_2 - E_0)/kT} d[E_2/kT] \\ & - \int_{\text{low}(1)}^{\text{up}(1)} h\nu \left\{B_{12} = \frac{c^3}{8\pi n^2 n_g}\frac{A_{21}(E_2, E_1)}{h\nu^3}\left[\frac{g_2/\Delta E_2}{g_1/\Delta E_1}\right]\right\} \\ & \times \left\{\rho_\nu = \frac{I_\nu}{c/n_g}\right\} g(\nu) \cdot \frac{N_1}{Z_1(T)} e^{-E_1/kT} d[E_1/kT] \end{aligned} \quad (10.5.13)$$

We shift the range of integration on the first integral by making the substitution: $E_2 = E_1 + h\nu$, thus $dE_2 = dE_1$, change the limits of integration to compensate for the

shift, and evaluate the line shape at $h\nu = E_2 - E_1$ so that $g(\nu) = 2/(\pi \Delta\nu_h)$ to obtain

$$St(\nu) - Ab(\nu) = \frac{\lambda^2}{8\pi n^2} \left\{ \frac{N_2}{Z_2(T)} e^{-(h\nu - E_0)/kT} - \left[\frac{g_2/\Delta E_2}{g_1/\Delta E_1} \right] \frac{N_1}{Z_1(T)} \right\} \times \frac{2}{\pi \Delta\nu_h} I_\nu \int_{\text{low}(1)}^{\text{up}(1)} A_{21}(E_2, E_1) e^{-E_1/kT} \, d[E_1/kT] \quad (10.5.14)$$

The gain coefficient is defined by

$$\boxed{\gamma(\nu) \triangleq \frac{St(\nu) - Ab(\nu)}{I_\nu} \triangleq N_2 \sigma_{em}(\nu) - N_1 \sigma_{ab}(\nu)} \quad (10.5.15)$$

where $\sigma_{em}(\nu)$ and $\sigma_{ab}(\nu)$ are the emission and absorption cross sections defined by

$$\sigma_{em}(\nu) \triangleq \frac{e^{-(h\nu - E_0)/kT}}{Z_2(T)} \frac{\lambda^2}{8\pi n^2} \frac{2}{\pi \Delta\nu_h} \int_{\text{low}(1)}^{\text{up}(1)} A_{21}(E_2, E_1) e^{-E_1/kT} \, d[E_1/kT] \quad (10.5.16)$$

$$\sigma_{ab}(\nu) \triangleq \left[\frac{g_2/\Delta E_2}{g_1/\Delta E_1} \right] \frac{1}{Z_1(T)} \frac{\lambda^2}{8\pi n^2} \frac{2}{\pi \Delta\nu_h} \int_{\text{low}(1)}^{\text{up}(1)} A_{21}(E_2, E_1) e^{-E_1/kT} \, d[E_1/kT] \quad (10.5.17)$$

The ratio of (10.5.17) to (10.5.16) yields a very important relation.

$$\frac{\sigma_{ab}(\nu)}{\sigma_{em}(\nu)} = \left[\frac{g_2/\Delta E_2}{g_1/\Delta E_1} \frac{Z_2(T)}{Z_1(T)} \right] e^{(h\nu - E_0)/kT} \quad (10.5.18a)$$

which can be rewritten so as to identify an important relation

$$\frac{\sigma_{ab}(\nu)}{\sigma_{em}(\nu)} = e^{h\nu/kT} \left[\frac{g_2/\Delta E_2}{g_1/\Delta E_1} \frac{Z_2(T)}{Z_1(T)} e^{-E_0/kT} \right] = e^{h\nu/kT} \frac{N_{2eq}}{N_{1eq}} \quad (10.5.18b)$$

An "excitation" potential, ϵ, can be defined such that $\exp[-\epsilon/kT] = N_{2eq}/N_{1eq}$ and the above written in a compact form.

$$\boxed{\sigma_{ab}(\nu) = e^{(h\nu - \epsilon)/kT} \sigma_{em}(\nu)} \quad (10.5.19)$$

where ϵ is a temperature dependent excitation potential. Equation (10.5.19) was first derived by McCumber [55] (for a very general system) in connection with phonon terminated lasers, but it is applicable to a broad class of systems. For the system considered here, ϵ is defined by

$$\boxed{e^{-\epsilon/kT} \triangleq \frac{N_{2eq}}{N_{1eq}}} = \left[\frac{g_2/\Delta E_2}{g_1/\Delta E_1} \frac{Z_2(T)}{Z_1(T)} e^{-E_0/kT} \right] \quad (10.5.20)$$

where the boxed part of (10.5.20) is perfectly general and independent of second equality which is specific to the model considered here. McCumber [55] shows that the excitation

potential ϵ is the energy required to move one atom from the manifold 1 to 2 keeping the system at a temperature T.

For the simple academic system analyzed here, we obtain an explicit relationship between σ_{em} and σ_{ab}, and there is no need to invoke the definition of ϵ expressed by (10.5.20). Sufficient information is seldom known about real systems to permit the evaluation of the integrals, and hence the boxed expressions become very useful in relating different measurements such as the spectral distribution of the fluorescence (from an optically thin sample*) and that of absorption. The first yields the shape of the emission cross section and the latter yields the value for the absorption. The two are related by (10.5.19) and thus ϵ can be estimated. McCumber suggests that a 20% error in $\exp[-\epsilon/kT]$ is tolerable.

The gain coefficient can be expressed in a very familiar format

$$\gamma(\nu) = [N_2\sigma_{em}(\nu) - N_1\sigma_{ab}(\nu)] \tag{10.5.20}$$

For our simplified model of this section, we obtain

$$\gamma(\nu) = \left\{ N_2 - \left[\frac{g_2/\Delta E_2}{g_1/\Delta E_1} \frac{Z_2(T)}{Z_1(T)} e^{(h\nu - E_0)/kT} \right] N_1 \right\} \sigma_{em}(\nu)$$

For instance, if $(\Delta E_2, \Delta E_1) \ll kT$, then $h\nu - E_0 \ll kT$ and the exponent becomes 1, $Z_2/Z_1 \rightarrow \Delta E_2/\Delta E_1$, the interior bracket becomes g_2/g_1, and we recover our familiar expression for the gain coefficient for an atomic system.

McCumber also identified a most useful relationship between the stimulated cross section and the spontaneous fluorescence that follows from (10.5.6) through (10.5.17). Rather than use the integral forms, let us retreat to the atomic case and express our usual relationship in terms of measurable physical parameters.

$$\sigma_{em}(\nu) \triangleq A_{21} g(\nu) \frac{\lambda_0^2}{8\pi n^2}$$

or

$$N_2\sigma_{em}(\nu) = \left\{ \frac{[A_{21} N_2 g(\nu)]}{2 \cdot 4\pi} \right\} \cdot \frac{\lambda_0^2}{n^2}$$

Since $A_{21} N_2 g(\nu)$ is the fluorescent photons emitted per unit of volume per unit of frequency into both polarizations into all angles of a sphere, 4π is number of steradians (solid angles) in a sphere, 2 is different polarizations, and $g(\nu)$ = line shape function, which is probability of emission into a frequency ν to $\nu + d\nu$, thus

$$N_2\sigma_{em}(\nu) = F_p(\nu) \frac{c^2}{\nu^2 n^2} \tag{10.5.21a}$$

where $F_p(\nu)$ is fluorescent photons emitted per unit of volume per unit of frequency per unit of solid angle for one polarization or in terms of the energy interval $h\nu = E$, we have

$$N_2\sigma_{em}(\nu) = F_p(E) \frac{h^2 c^2}{E^2 n^2} \tag{10.5.21b}$$

*The fluorescence from the upper states can be reabsorbed by the ground state manifold. If this occurs, the measured fluorescence is not representative of the $\sigma_{em}(\nu)$. One can use a completely inverted stem with all of the atoms in manifold 2 or use a small enough sample to avoid significant reabsorption.

The utility of either form of (10.5.21) stems from the fact that it is relatively easy to measure the relative fluorescence as a function of frequency/energy.

We cannot go further without some sort of a statement about the dependence of the transition probability on (E_2, E_1), which is seldom known from an experimental viewpoint and extremely difficult to predict. To illustrate a few important issues, this hypothetical example is carried further and the integrals of (10.5.16) and (10.5.17) were evaluated assuming that the transition probabilities could be expressed as

$$A_{21}(E_2, E_1) \rightarrow A_0 \cdot \exp[-\{[h\nu - E_0 - \Delta E]/\delta E\}^2]$$

which amounts to assuming that the transition probability is maximum when $h\nu = E_0 + \Delta E$, an arbitrary but tractable assumption that has some basis in experiment. The parameters chosen for the calculations were $E_0 = 6544\ \text{cm}^{-1}$, $\Delta E_1 = 318\ \text{cm}^{-1}$, and $\Delta E_2 = 306\ \text{cm}^{-1}$, which were guided by the Er:silica energy level diagram presented in the last section although the values for $\Delta E = 90\,\text{cm}^{-1}$ and $\delta E = 125\,\text{cm}^{-1}$ were chosen arbitrarily. A graphical presentation of the cross sections are given in Fig. 10.15 for two temperatures, 188 K and 300 K.

There are two important lessons to be learned from this hypothetical example.

1. The stimulated emission cross section can be bigger than the absorption for parts of the spectral region of interest.
2. Both are temperature dependent. This is due to the redistribution of atoms in the two manifolds according to the Boltzmann factor rather than due to a broadening mechanism.

If we presume that a pump has promoted a fraction of the atoms to the upper manifold, we can predict the spectral dependence of the gain according to (10.5.20). This is shown in Fig. 10.16.

There are three additional lessons to be learned:

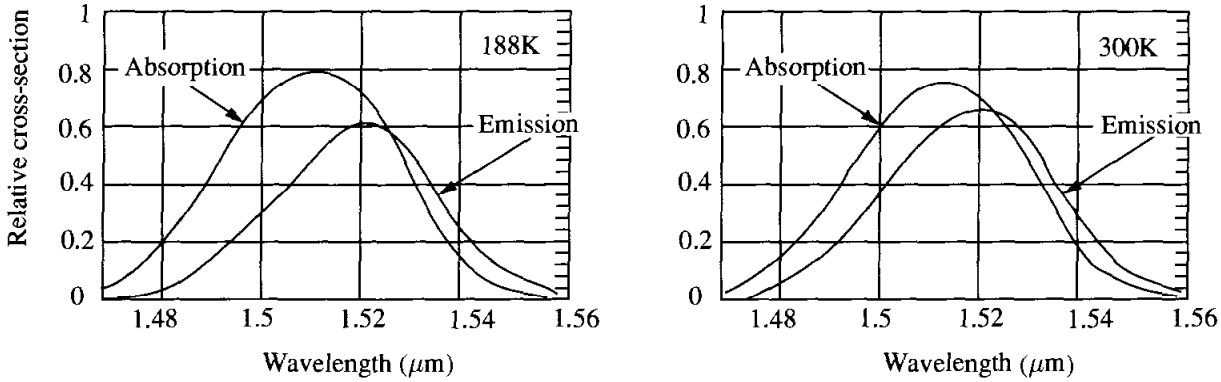


FIGURE 10.15. The relative cross sections as a function of wavelength at 77 K and 300 K for the example chosen.

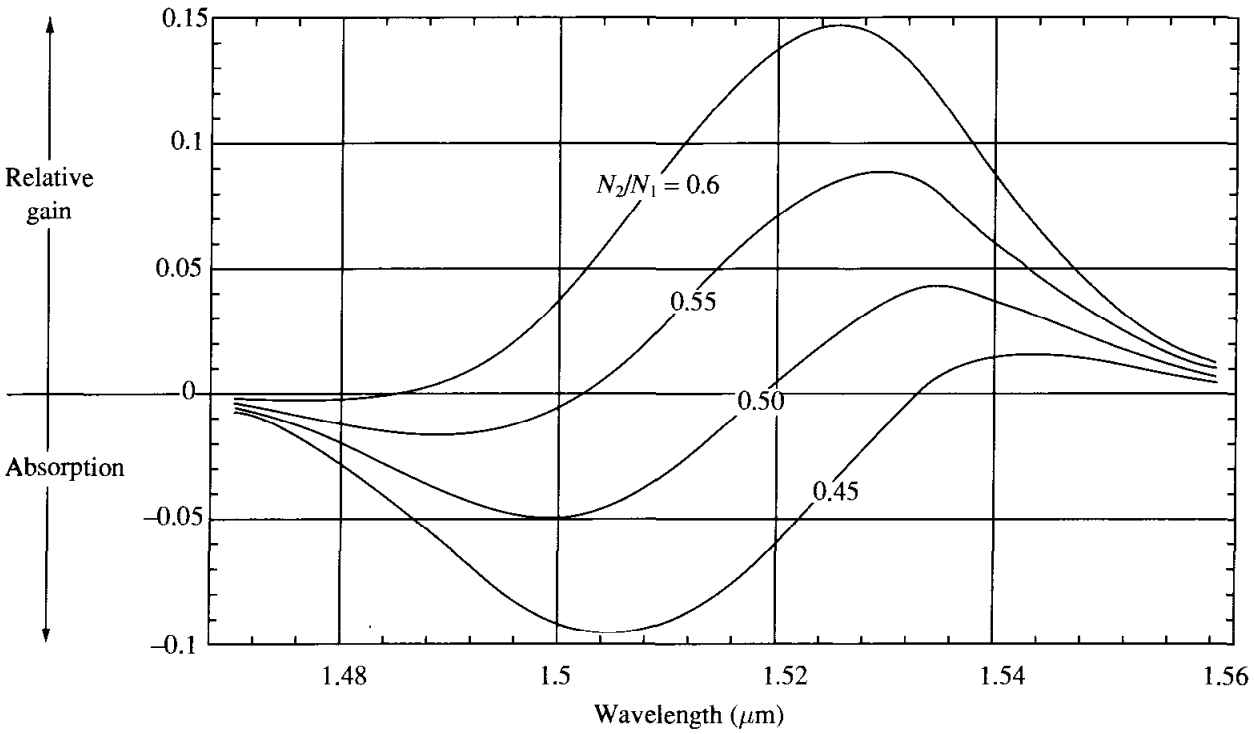


FIGURE 10.16. The relative gain as a function of wavelength for the various ratios of the densities in the two manifolds.

1. One can have, simultaneously, gain on the long wavelength side of the band and absorption of the high frequency side. This is essential for fiber amplifiers of 1.55 μm signals excited with the 1.48 μm laser pumps.
2. We do not need N_2 to be bigger than N_1 to amplify at some wavelength.
3. However, the bandwidth for positive gain does increase with the inversion ratio.

While the calculations and theory have been performed for the academic model and are of limited utility, the theory can be adapted to more realistic problems at the expense of considerably more tedious arithmetic. However, the results expressed by (10.5.20) and (10.5.21) are independent of our model and can be used for these cases.

10.6 TUNABLE LASERS

10.6.1 General Considerations

For most of the lasers discussed to this point of the book, it is easy to identify the active atom or pairs of excited atoms responsible for the gain. The energy levels are distinct and sharp and hence the tuning range is limited to the linewidth of the transition (tens of GHz) or to "jumps" between adjacent transitions in the case of a molecular laser such as CO_2. Based upon our experience at low frequencies (RF and microwave), we would prefer a system that is tunable over a significant fraction—say 5 to 10%—of its mean wavelength. To achieve

such a goal, we must have a system involving as many processes as possible: electronic, vibrational, rotational, and phonons (i.e., lattice vibrations) in an environment that promotes the coupling of these characteristic motions so as to have overlapping transitions. This last requirement indicates super-high pressure gases, or dense media, such as a liquid or a solid.

There have been two practical systems that have achieved this goal of significant tunability: a dye in a suitable solvent and various metallic atoms doped into a crystalline environment. Since the physics of these two systems are considerably different, they will be discussed separately.

10.6.2 Dye Lasers

The laser medium is usually a dye (mostly poisonous but sometimes digestible) dissolved in a solvent such as water or alcohol or ethylene glycol (antifreeze) at $\sim 10^{-4}$M concentration, which is optically pumped by a flashlamp or a beam from another laser. The dye is usually composed of hydrocarbons and other atoms tied in a complicated chain involving 50 or more atoms. A small sampling of the molecular structure of the dyes, the tuning range, and some of the solvents is shown in Fig. 10.17. Figure 10.18 shows the tuning range of various dyes when pumped by various common sources. One can see that there is a complicated interplay between the concentration of the dye, the solvent, and the pump wavelength with all affecting the tuning range and the efficiency. The chemical lifetime of the dye is also involved.

Since there are so many possibilities of coupling between the complicated configurations of atoms, one searches for a simple approximate model that matches the experimental observations. Probably the most successful one is the simple molecular model shown in Fig. 10.19. One assumes that there is some interaction potential dependent upon a molecular coordinate q (a distance) that describes the classical motion of the vibrations corresponding to the energy levels in both the ground and excited states. The most critical issue in this sketch is the fact that the minimum of the potential for the first excited state is shifted along this coordinate axis with respect to the minimum of ground state.

The Boltzmann factor, $\exp[-(E - E_{\min})/kT]$, describes the occupation probability of the "molecules" in each manifold. This is similar to the case analyzed in Sec. 10.5. Thus the optical absorption is from the heavily populated region near q_0 in S_0 along a vertical path to a high vibrational state of S_1.* Once the molecule is in S_1, the excess energy ΔE_2, is rapidly apportioned to the myriad of vibrational modes of the complicated molecule and thus the occupation probability of the states in S_1 is also described by the Boltzmann factor referenced to the lowest state in S_1. Hence most of the atoms in S_1 reside at E_1 with a coordinate q_1, which is shifted with respect to q_0.

The $S_1 \rightarrow S_0$ emission is also vertical and hence the lower state of the emitted photons is a highly excited one in ground manifold. Thus the lasing or fluorescence can be indicated by the following kinetic sequence:

*Since the absorbed photon carries so little momentum, the molecular coordinates cannot change in a transition—hence the path must be vertical. This is the basis of the Frank-Condon principle.

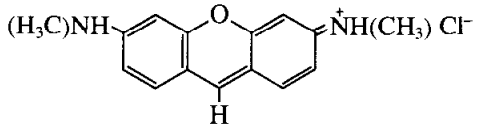
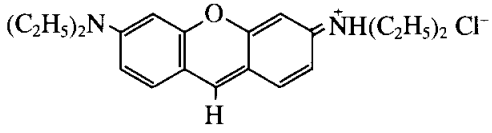
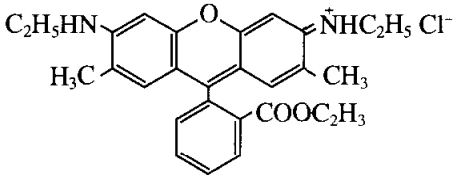
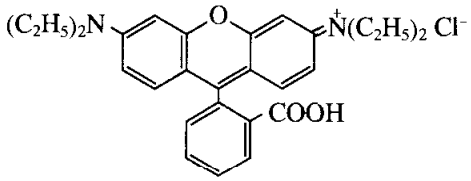
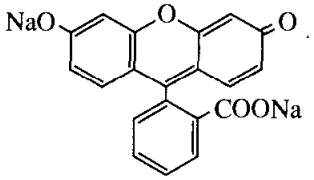
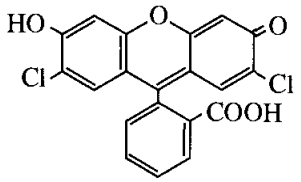
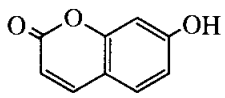
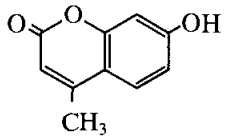
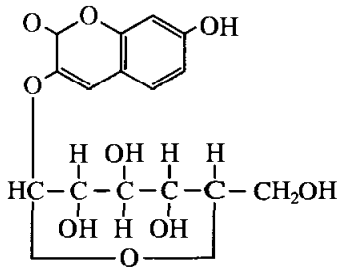
Dye	Structure	Solvent	Wavelength
Acridine red		EtOH	Red 600–630 nm
Puronin B		MeOH H_2O	Yellow
Rhodamine 6G		EtOH MeOH H_2O DMSO Polymethyl-methacrylate	Yellow 570–610 nm
Rhodamine B		EtOH MeOH Polymethyl-methacrylate	Red 605–635 nm
Na-fluorescein		EtOH H_2O	Green 530–560 nm
2,7-Dichloro-fluorescein		EtOH	Green 530–560 nm
7-Hydroxy-coumarin		H_2O (pH ~ 9)	Blue 450–470 nm
4-Methylem-belliferone		H_2O (pH ~ 9)	Blue 450–470 nm
Esculin		H_2O (pH ~ 9)	Blue 450–470 nm

FIGURE 10.17. Molecular structure, laser wavelength, and solvents for some laser dyes. (Data from Snavely [8].)

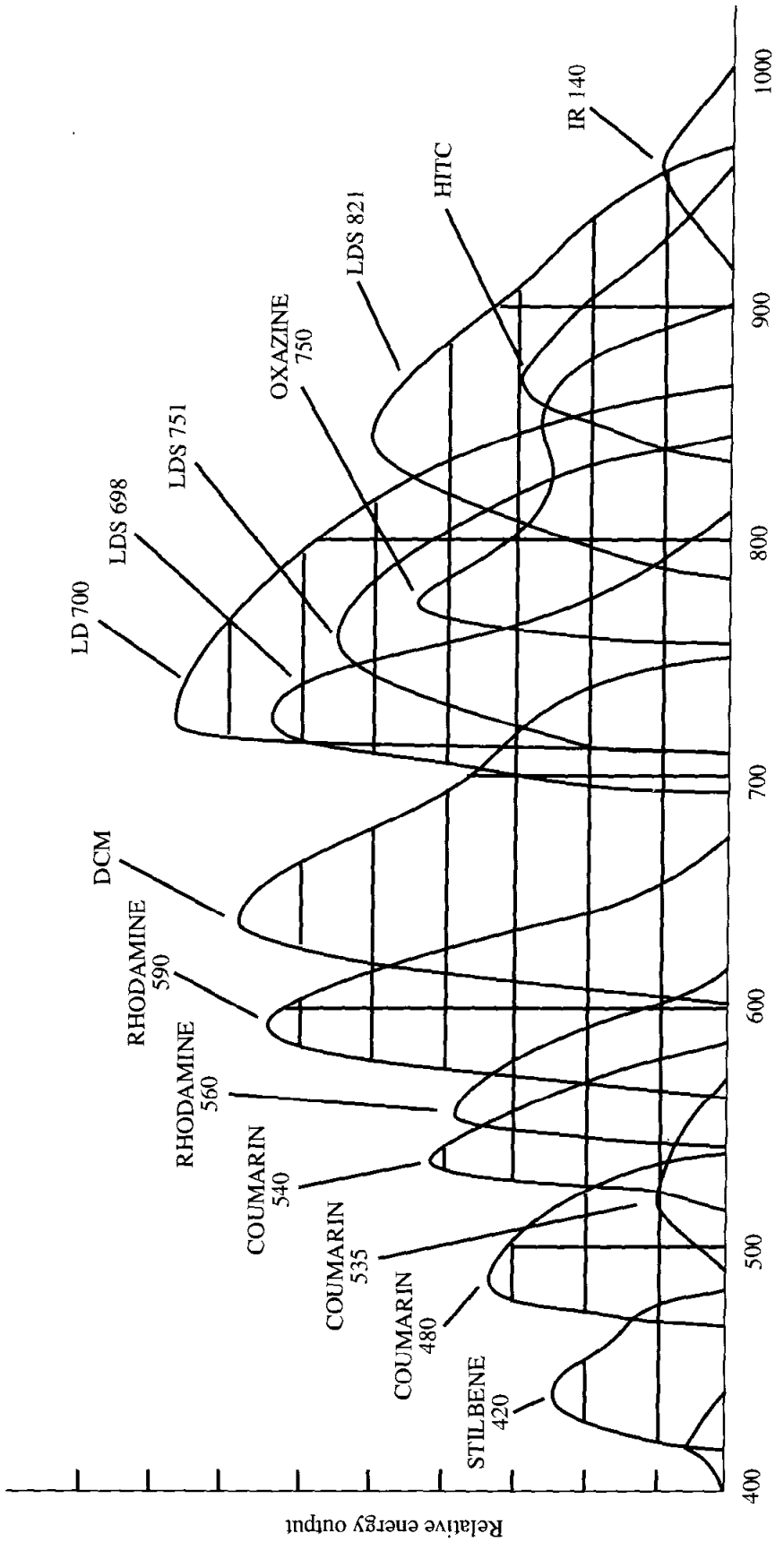


FIGURE 10.18. Performance of various dyes when pumped with an argon-ion or Krypton-ion laser (Data from Spectra-Physics and advertised in Exuton, Inc. catalog, p.4 [61].)

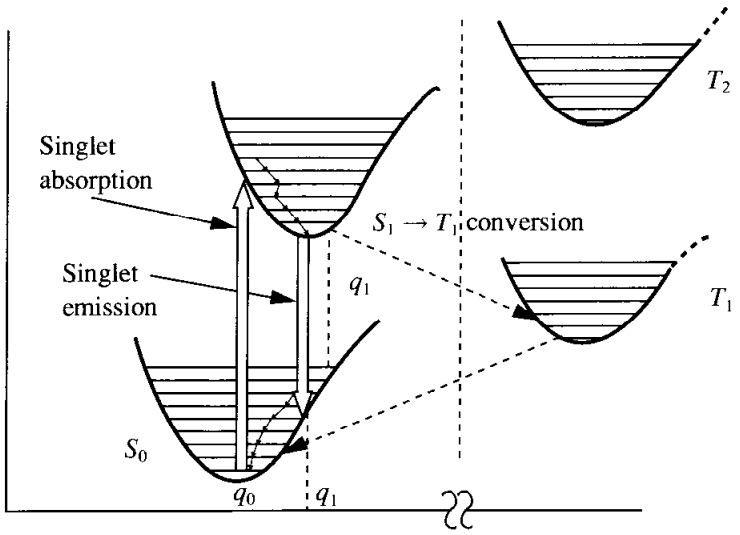


FIGURE 10.19. Energy-level diagram typical of a dye. (Data from Bass et al. [9]).

Absorption of a photon from the pump:

$$h\nu_{\text{pump}} + (\text{dye at } q_0 \text{ in } S_0) \rightarrow (\text{dye at } q_0 \text{ in } S_1) \tag{10.6.1a}$$

Relaxation in the upper S_1 manifold:

$$(\text{dye in } S_1 \text{ at } q_0) + (\text{solvent}) \overset{\text{very fast}}{\rightarrow\rightarrow\rightarrow} (\text{dye in } S_1 \text{ at } q_1) \tag{10.6.1b}$$

Spontaneous or stimulated emission back to S_0

$$(\text{dye at } q_1 \text{ in } S_1) + (h\nu_{\text{laser}}) \overset{\text{stimulated emission}}{\rightarrow\rightarrow\rightarrow} (\text{dye at } q_1 \text{ in } S_0) \tag{10.6.1c}$$

Relaxation in the lower S_0 manifold:

$$(\text{dye at } q_1 \text{ in } S_0) + (\text{solvent}) \overset{\text{very fast}}{\rightarrow\rightarrow\rightarrow} (\text{dye at } q_0 \text{ in } S_0) \tag{10.6.1d}$$

Such a four-level model conforms to the experimental observation that the absorption peaks at short wavelengths, whereas the emission peaks at longer ones in the manner shown in Fig. 10.20. The fact that the fluorescence and absorption are mirror images is a convincing argument that the minimum of S_1 state is displaced with respect to the minimum of S_0.

Thus the pumping sequence is to excite the ground-state molecules into the S_1 state fast enough to overcome the thermal population of the high vibration states of S_0. Whereas the radiative lifetime of the upper laser level of ruby or neodymium was hundreds of microseconds, here the lifetime is submicroseconds, typically 10 ns. Consequently, the pumping must be intense and *fast*.

This last requirement is necessary because of the presence of the triplet system in the dye, a system that is normally empty. Unfortunately, some of the molecules in the S_1 state are converted to a triplet configuration—by interaction with the solvent or by a hundred other processes—and accumulate in the state T_1. Now absorptive transitions can take place

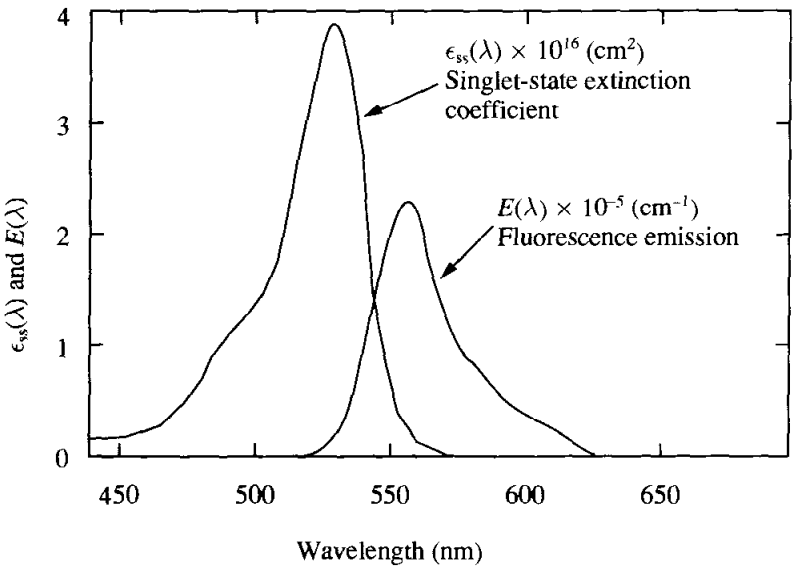


FIGURE 10.20. Singlet-state absorption and fluorescence spectra of rhodamine 6G obtained from measurements with a 10^{-4}M ethanol solution of the dye. (Data from Snavely [8].)

between T_1 and T_2. By conforming to Murphy's law,* the T_1 and T_2 absorptions coincide with part of the emission profile of S_1 to S_0.

Thus what started out to be a four-level system has many of the undesirable features of the three-level one. We can enhance the reconversion of that triplet back to the S_0 state by adjusting the dye chemistry. For instance, the lifetime of T_1 can be as long as 10^{-3} sec or as short as 10^{-7} sec in a solution in which oxygen is dissolved. But even this short lifetime is long compared to the radiative lifetime of S_1. As a consequence, a dye laser tends to be self-terminating, owing to the buildup of the triplet T_1 population. Thus most flashlamp pumped dye lasers are pulsed with pulse widths on the order of 100 to 1000 ns.

But, of course, the striking beauty of a dye laser is its tunability over a majority of its fluorescence spectrum. Indeed, the dye laser has provided the answer to the dream of having a tunable oscillator in the visible portion of the spectrum. This device is every bit as useful as the corresponding tunable oscillator is for the low-frequency domain (maybe even more).

The pulsed dye laser tends to be of the high-power, short-pulse variety. The former characteristic is caused by the very intense fast pumping and the latter by self-quenching by the triplet buildup. However, there is a way to avoid this last difficulty—flow the dye out of the cavity—to prevent the triplets from building up too high of a level.

This simple line of reasoning has led to a CW dye laser that has the form shown in Fig. 10.21. Here the dye solution flows across the cavity at a position where the spot size of the $TEM_{0,0}$ dye laser mode is a minimum so as to enhance the stimulated emission rate. Furthermore, the continuous stream of dye solution is oriented at Brewster's angle to enhance the Q of the dye-laser cavity.

On the other hand, the angle of the incoming pump—usually another laser, such as an argon-ion or krypton-ion laser—is adjusted to give maximum gain along the region where the pump and dye laser beams overlap.

*"If anything can go wrong, it will."

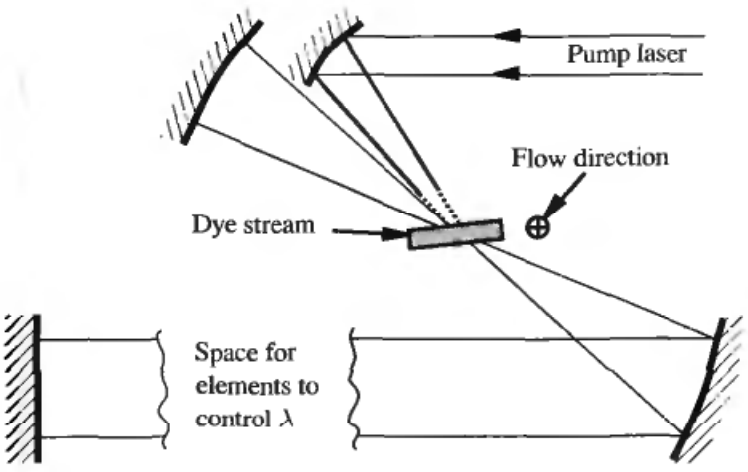


FIGURE 10.21. Typical configuration for a CW dye laser.

The used dye is collected and recirculated. Now, there are literally minutes available for the triplet T_1 to be reconverted back to S_0. The more pertinent time interval is the one corresponding to the passage of the stream past the focus of the pump beam. If this is a small spot size, a nominal flow replenishes the dye before the triplets can quench the laser action.

10.6.3 Tunable Solid State Lasers

One of the major drawbacks of the dye lasers is the relative short chemical life of the dye in the solvent, a problem solved by solid-state lasers capable of significant tuning ranges. Two of the most common are the Ti:sapphire and Alexandrite systems, which have many similarities but are sufficiently different to warrant a separate discussion. Both are "phonon terminated" lasers, a term that will be explained shortly.

Ti:Sapphire Lasers. In the Ti:sapphire case, TiO_3 is doped (say 1% by atomic weight) into a crystal of Al_2O_3 (sapphire) with Ti substituting for some of the aluminum atoms in the lattice structure. There are two electrons filling the 4s orbitals in Ti even though the 3d system is only partially filled with two electrons in an orbital capable of holding 10. When Ti substitutes for an aluminum at an octahedral site as shown in Fig. 10.22(a), the two 4s and one 3d electrons are occupied with the bonding to the oxygen atoms leaving a single 3d electron active at the substitutional site and Ti acts as if it were triply ionized.

The energy degeneracy and simple symmetry of the 3d active electron is broken by the local crystalline fields and spin–orbit interactions to yield the simplified energy levels diagram of Fig. 10.22(b). Furthermore, these states are influenced by (or coupled to) the local phonons (or lattice vibrations) leading to the energy level diagram shown in Fig. 10.22(c), which is very similar to the molecular model of the dyes given in Fig. 10.19 where q is some configuration coordinate.

Notice that the minimum of the upper manifold is shifted with respect to the minimum of the ground state. The vibronic states of $^2T_{2g}$ are normally occupied according to the Boltzmann distribution and thus the high lying levels of the manifold are mostly empty. The pump promotes the system to the ^2E_g manifold where it quickly ($\sim 10^{-12}$ sec) adjusts its

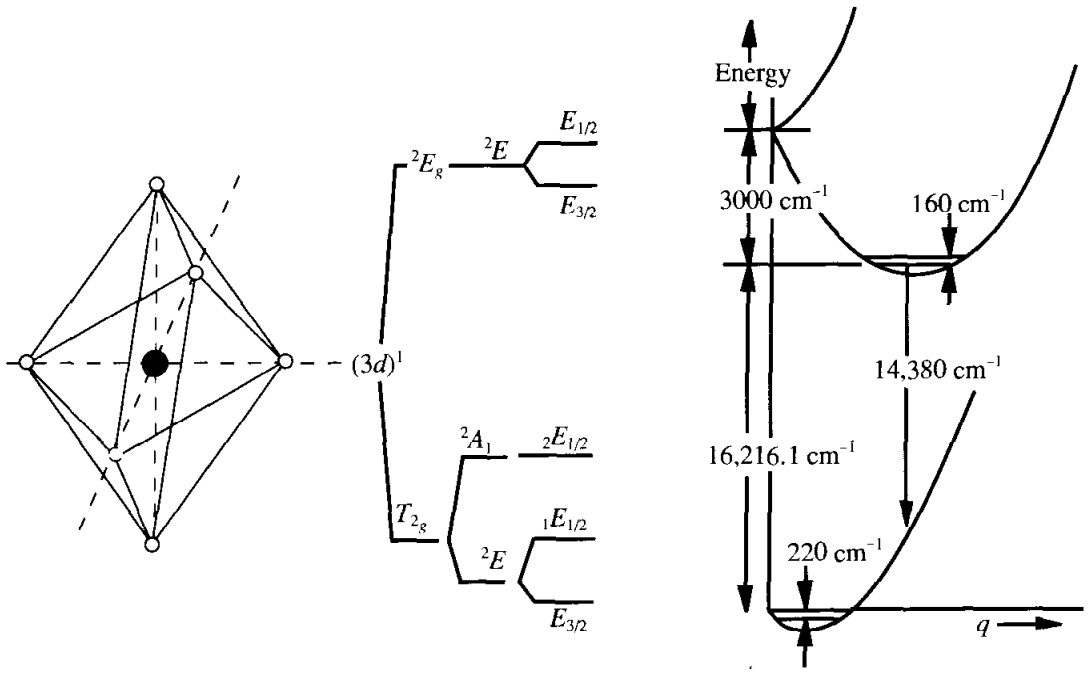


FIGURE 10.22. (a) The octahedral site for the titanium ion (solid) surrounded by the six oxygen atoms in sapphire. (b) Term splitting by the crystalline fields. (c) Simplified schematic of the potential energy curves for the $^2T_{2g}$ and the ^2E_g states of Ti^{3+} in sapphire. (Same as Fig. 1 of Byvik and Buoncristiani [79]. Numerical values for (c) from Fig. 3 of Gächter and Koningstein [85].)

configuration to those minimum energy states by the emission of one or more phonons. The lifetime of the ^2E_g manifold is relatively long $\sim$ 3.9 μs and the probability of nonradiative transitions is small. The stimulated emission by another photon along with a simultaneous emission of a one or more phonons returns the system back to the $^2T_{2g}$ vibronic states. Because of the importance of the phonons, it is described as a "phonon terminated" laser.

The energy difference between the minimum of the ground state ($^2T_{2g}$) and the ^2E is about 19,400 cm^{-1}, which is a very good match to the strong argon ion laser line at 5145 Å or the doubled YAG wavelength at 532 nm. The absorption cross section for this transition is shown in Fig. 10.23 (on page 393) and illustrates that even the argon ion line at 4800 Å can be used.

The information of greatest significance to laser applications is shown in Fig. 10.24 (on page 393) where the stimulated emission cross section is given as a function of wavelength. That information, along with the upper state lifetime (τ = 3.9 μs) and is sufficient to enable one to do a reasonable job of modeling of this laser. Notice that the Ti:sapphire laser is capable of amplification over wavelength range of 700 to 1020 nm. Such a large gain-bandwidth product implies that mode locking can yield extremely short pulses.

Alexandrite lasers. The active atom in this laser is chromium (same as in the ruby laser), which is doped into a crystal, chrysoberyl ($BeAl_2O_4$). It can function as a three-level laser in the manner of ruby or as a vibronic or phonon terminated four-level laser that can be tuned. A simplified energy level diagram due to Walling is shown in

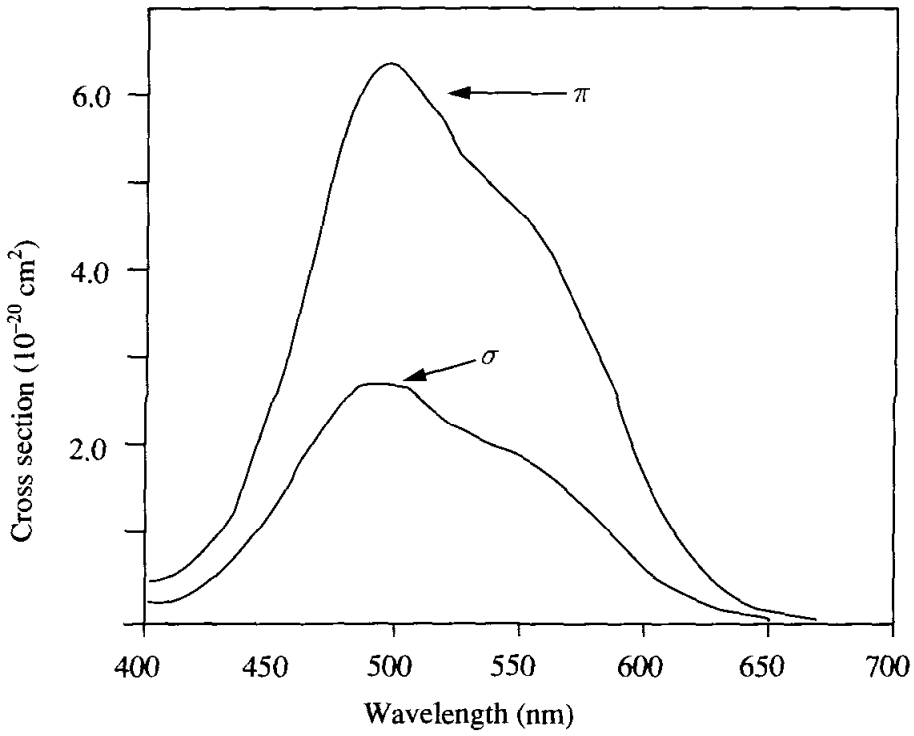


FIGURE 10.23. The polarized absorption cross section for the $^2T_{2g} \rightarrow {}^2E$ transition in Ti:Al_2O_3 (Data from Fig. 1 of Moulton [73].)

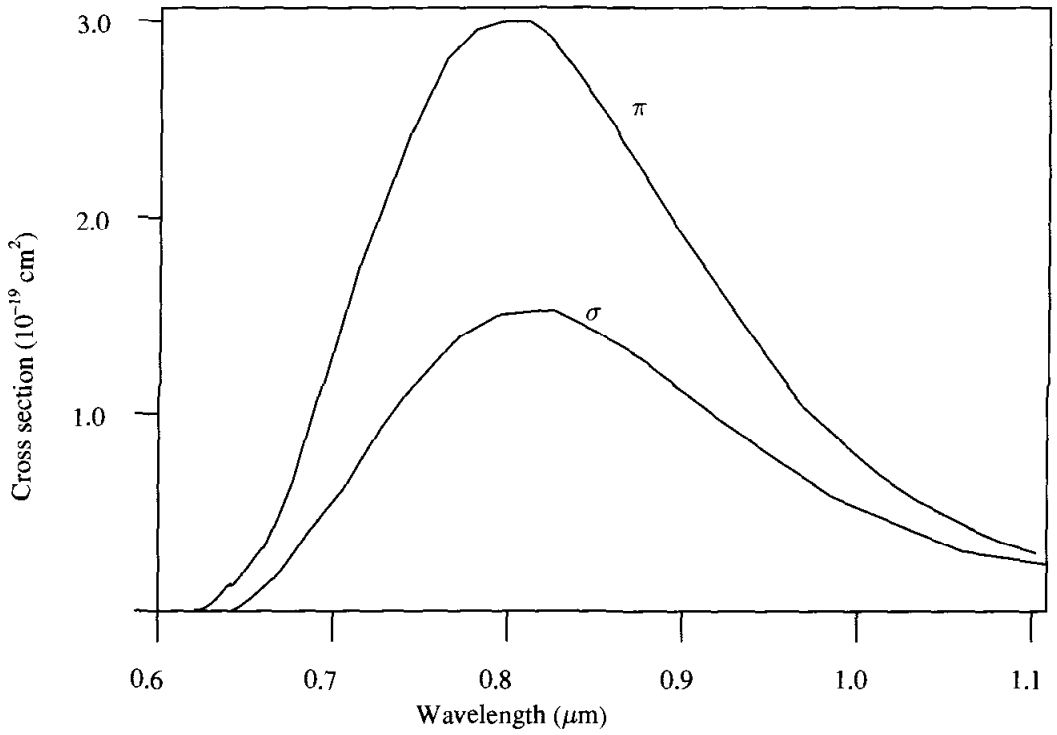


FIGURE 10.24. The polarized fluorescence spectra and relative gain cross section for Ti:Al_2O_3 (Data from Fig. 2 of Eggleston et al., [66].)

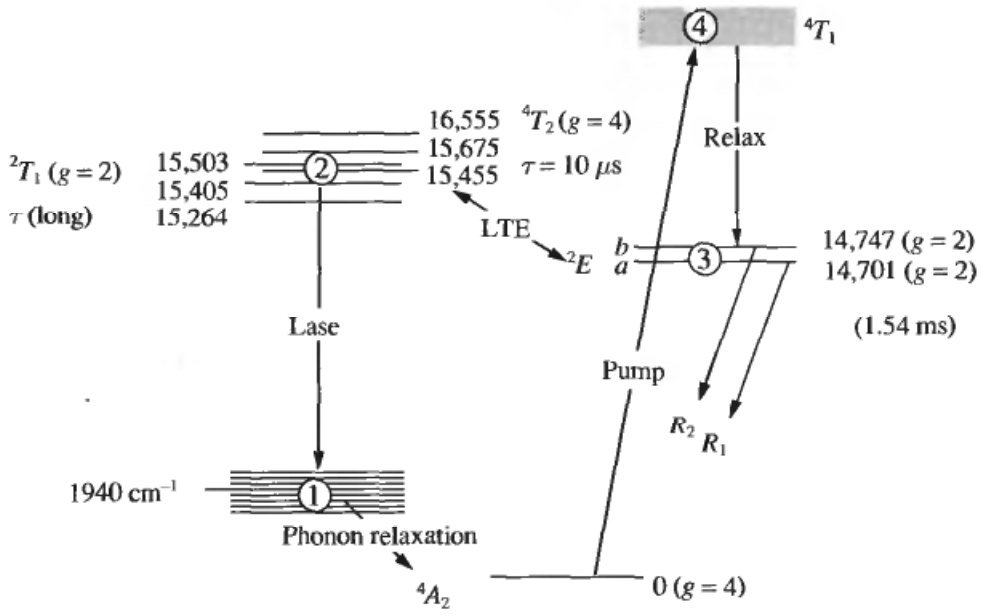


FIGURE 10.25. The simplified energy level diagram of alexandrite using the data presented in Walling et al [87, 89]. The lifetime shown for the $^4T_2$ of 10 μs was for a weakly doped crystal. A simplified model of Walling assigns a lifetime of state 2 to be 6.6 μs separated from 3 by 800 cm^{-1}. A typical doping density of $[Cr^{3+}] = 2.2 \times 10^{19}$ cm^{-3}. The crystal is biaxial with $n_a = 1.7381$, $n_b = 1.7436$, and $n_c = 1.7361$ at 700 nm.

Fig. 10.25, where the R lines are the usual ones associated with the transitions between the ^2E and $^4A_2$ states of Cr^{3+} (see Fig. 10.2, the energy level diagram of ruby).

This crystal was proposed as an alternative to sapphire (Al_2O_3) as a host for the Cr^{3+} atoms, but it rapidly came into its own when it was realized that it could function as a phonon terminated laser and thus be tuned over a significant wavelength interval. The system is pumped through the broad absorption bands $0 \rightarrow 4$, which are similar to those in ruby and shown in Fig. 10.3 Those atoms then relax back to the ^2E "storage levels," which would decay with a lifetime of 1.54 ms if they were not connected to higher levels by fast thermal processes that maintain the populations in $^2T_1$, $^4T_2$, and ^2E manifolds in LTE. The system can lase between 3 and 0 as a three-level laser like ruby or between 2 and 1 with the latter requiring the simultaneous emission of one or more phonons.

Figure 10.25 indicates a new situation that has not been encountered previously. Most of the excited atoms will accumulate in the "storage level" with its lifetime of 1.54 ms with only a small fraction $\Sigma_g \exp[-(\Delta E/kT)] \sim 2.5 \times 10^{-2}$ (at 300 K) would be in the upper laser level with its lifetime of 10 μs. Thus, the excited state population will decay with a time constant that depends upon the temperature, decreasing with increasing temperature because of the higher fraction of the excited state population residing in the shorter lifetime states. Thus we have one of these unique circumstances where the laser performance actually improves with temperature. There is a trade off however. The lower laser level, 1, is also in thermodynamic equilibrium with the lowest state and hence its population increases with temperature. Thus, only the wavelengths longer than 730 nm benefit from the increased temperature.

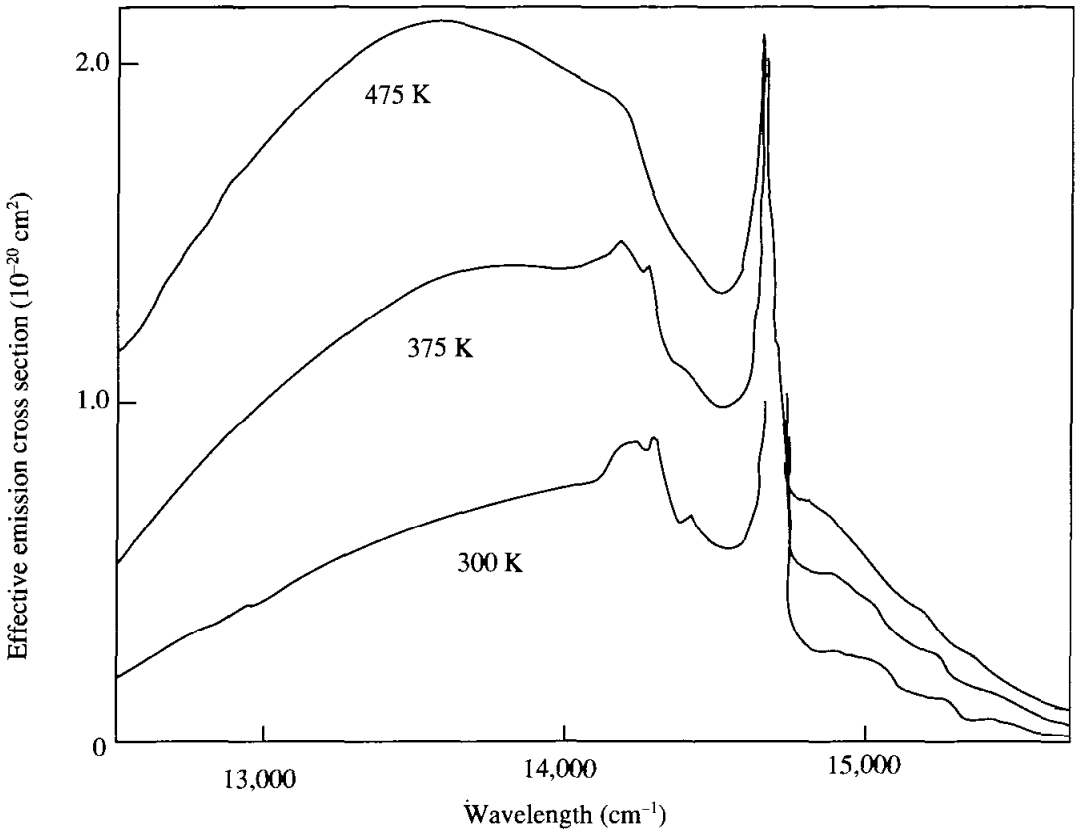


FIGURE 10.26. Variation of the stimulated emission cross section with wavelength for various temperatures (after Fig. 8 of Walling, [87]).

The variation of the stimulated emission cross section with wavelength and temperature are shown in Fig. 10.26. A simplified model of the laser due to Walling assigns a lifetime of the storage level to be 1.54 ms, the lifetime of the upper level to be 6.6 μs, with an energy gap between the two of 800 cm^{-1}, numbers not significantly different than the data presented in Fig. 10.25. The lower state relaxes in subnanoseconds.

10.6.4 Cavities for Tunable Lasers

With gain persisting over such a wide bandwidth of the above lasers, it is appropriate to consider the means by which any single wavelength in the tuning interval can be chosen. If we were to use just a simple two-mirror cavity, then the laser will oscillate at the wavelength where the gain-to-loss ratio is the largest, and with broadband gain and broadband mirrors, there would not be much discrimination nor wavelength control.

Prisms, gratings, and Fabry-Perot etalons, separately and in many wondrous combinations, are used to select the wavelength and to restrict the oscillation bandwidth of these tunable lasers. Two samplings are shown in Fig. 10.27.

In Fig. 10.27(a), one depends upon the chromatic dispersion of the glass in the prisms, and cascades many such prisms for additional selectivity. By changing the angle of the mirror M_2 or the orientation of the prisms, one can select the wavelength. In 10.27(b), we

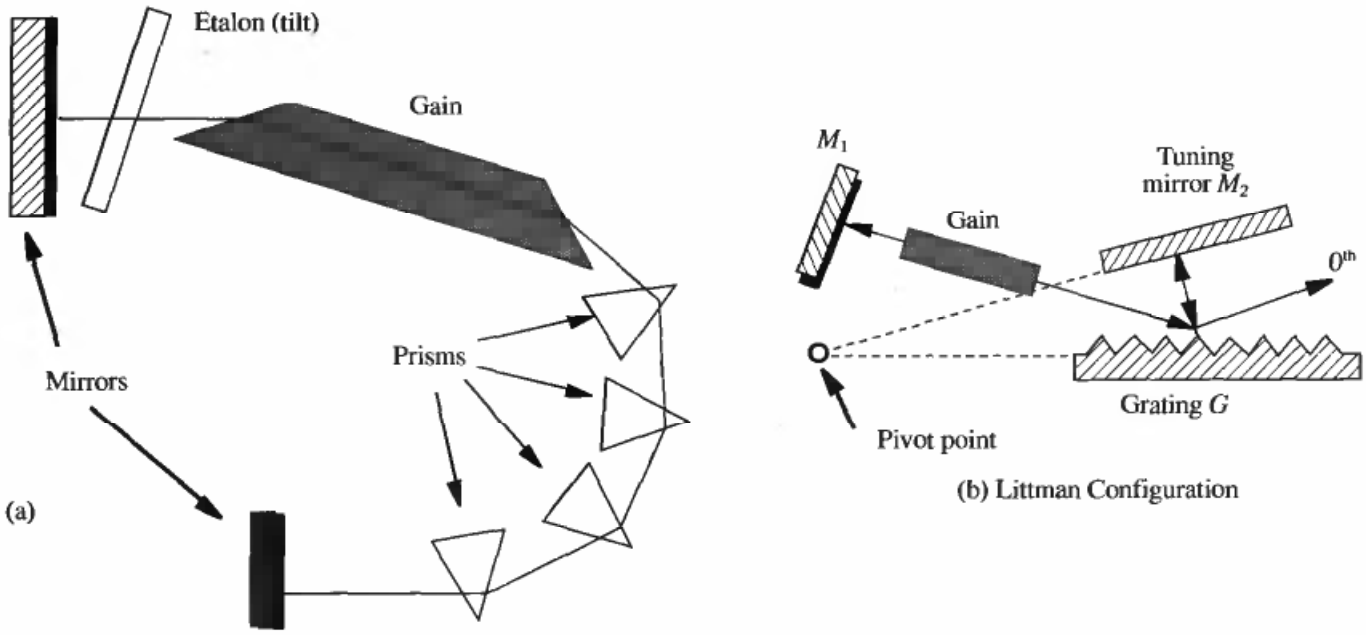


FIGURE 10.27. Various methods of tuning a laser.

use a diffraction grating at grazing incidence to increase the number of lines of the grating being illuminated, which, in turn, increases its resolving power. Feedback is provided by the tuning mirror M_2, which is adjusted to be perpendicular to the first order diffraction pattern. M_2 sends the diffracted beam back to the grating and back to M_1 (since the grating is a reciprocal device). The "output" is through the 0^{th} order of the grating (i.e., the grating appears as a simple turning mirror). A rather unique feature of this configuration is that a single cavity mode can be tuned over a significant wavelength interval, without mode hops, if both M_2 and G are rotated about a common pivot point found by extrapolating both surfaces back to the intersection.

10.7 GASEOUS-DISCHARGE LASERS

10.7.1 Overview

Shortly after the demonstration of the ruby laser, Javan, Bennett, and Herriott (Ref. 10) demonstrated oscillation in an RF excited gas mixture of helium and neon at a wavelength of 1.1523 μm. Since that time, literally hundreds of gases have been excited and made to lase at literally thousands of different wavelengths. In fact, Ref. 11 lists 159 different transitions for neon alone, with wavelengths ranging from the UV at 267.9 nm to 132.8 μm at the ultramicrowave portion of the spectrum. (One now asks the question of where [in λ] and with what efficiency a gas will lase, rather than if it will lase.) Obviously, we have to restrict our attention to a few of the more common ones, namely, the helium-neon laser, the argon ion laser, and the CO_2 laser, all being excited by a simple discharge.

Gas lasers have many advantages over other types of lasers, primarily because the active medium—a gas—is very homogeneous even when mistreated in a violent fashion. As we will see later, they are relatively efficient. Because of the homogeneity, most gas lasers produce a nearly perfect Gaussian beam mode (with proper optics), and some can be scaled to large volumes and extreme powers.

However, these very desirable characteristics have disadvantages. Gas lasers are big devices, huge compared to a semiconductor laser, for instance. They require voltages and currents for excitation that are not normally compatible with most modern solid-state electronic devices. Nevertheless, the gas laser, in particular the CO_2 laser, has a commanding lead in CW power and efficiency. In many respects, the gas laser is the simplest to understand and analyze.

10.7.2 Helium-Neon Laser

The energy-level diagram for the helium-neon system is shown in Fig. 10.28 together with some common laser transitions shown as solid lines. For instance, the common "red" laser line at 6328 Å is a transition from the $3s_2$ to $2p_4$ state.* The numbers in parentheses are the A coefficients. Also shown in this figure, as dashed lines, are the spontaneous decay routes from the lower laser level $2p_4$ down to the $1s$ states.

These latter transitions, $2p_4 \rightarrow 1s_{2,4,5}$, and many others originating at the rest of the $2p$ states, give neon signs their characteristic red color. Note that the A coefficients (the numbers in parentheses in Fig. 10.28) for these spontaneous transitions are quite large compared with that of the laser transition. This illustrates an important point referred to in Chapter 8, where a generalized pumping scheme of a laser was discussed. We do *not* look for strong spontaneous lines for a laser; rather, it is the weak lines, which terminate on the upper states of the strong ones, that are prime candidates for a laser.

To illustrate this point further, consider the data given in Table 10.6 for the two states of interest in the 6328 Å transition. Here we list the many transitions out of the two states and their relative intensities as listed in the *Handbook of Chemistry and Physics* (CRC Press). *All* the $3s \rightarrow 2p_{1-8}$ transitions have lased, and few of the $2p_4 \rightarrow 1s_{2-5}$ have oscillated (in CW operation). This is in spite of the fact that the A coefficient is in the numerator of the laser gain equation:

$$\gamma_0(\nu) = A_{21} \frac{\lambda^2}{8\pi} g(\nu) \left(N_2 - \frac{g_2}{g_1} N_1 \right) \tag{10.7.1}$$

However, as we will see, it is difficult to obtain an inversion on the $2p \rightarrow 1s$ transitions; hence, most do not lase. But the rapid spontaneous decay of the $2p$ states empties the lower level of the $3s \rightarrow 2p$ laser transitions.

A typical helium-neon laser is a glass tube 10 to 100 cm long, 2 to 8 mm in diameter, and filled with helium and neon gas in a 5 : 1 to 20 : 1 ratio to a pressure prescription of

*The spectroscopic notation for neon is particularly confusing but widely used. Paschen notation, as used in Fig. 10.28 was an attempt to fit the neon spectra to a hydrogen-like theory. It did not work! Nevertheless, the notation is still with us. As far as we are concerned, the numbers and letters are names of the states, nothing more.

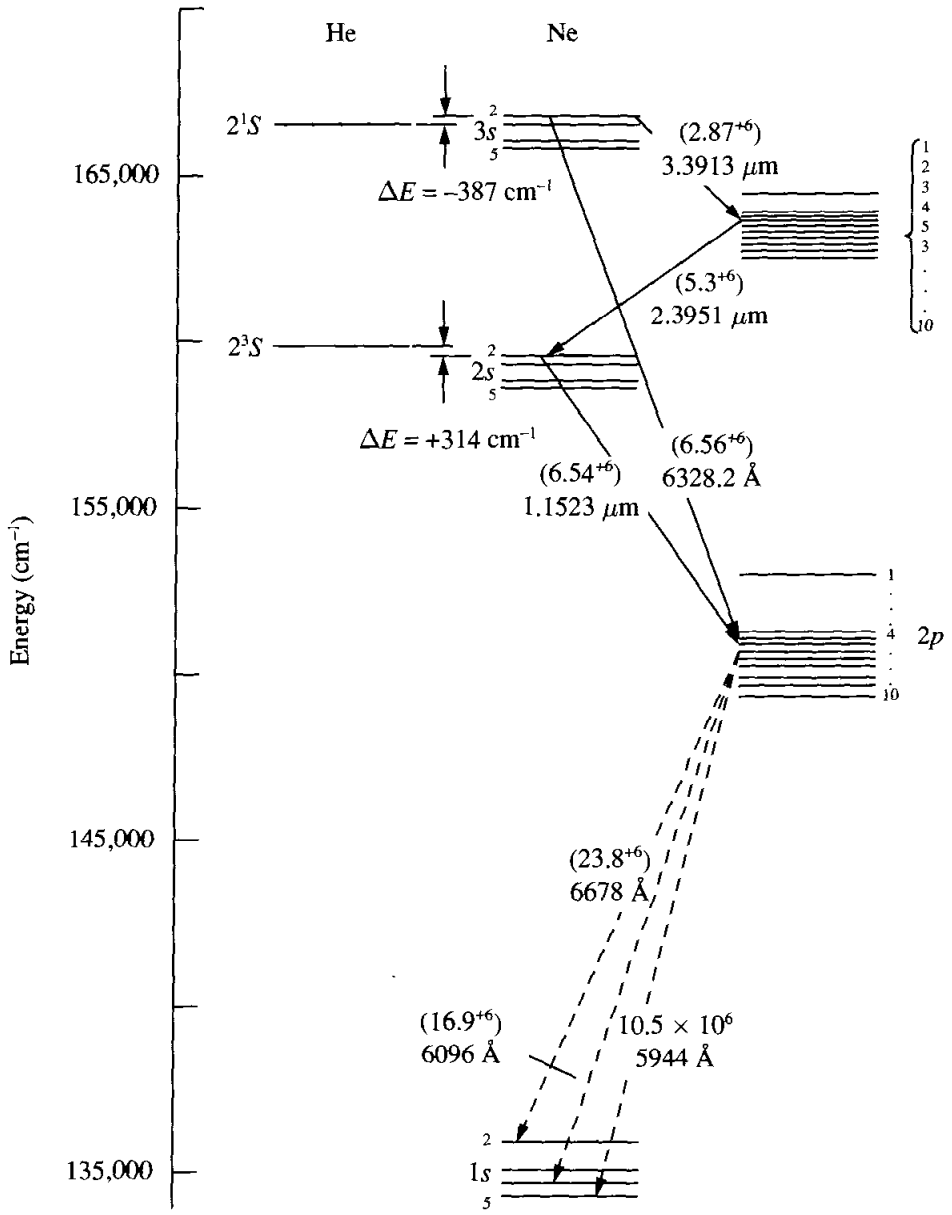


FIGURE 10.28. Energy-level diagram for the helium-neon laser. The solid line represents the common laser line; the dashed lines are spontaneous. The numbers in parentheses are the A coefficients.

reference [12].

$$p \cdot d \sim 0.36 \text{ torr-cm} \qquad d = \text{diameter of bore (cm)}$$

Typical currents through the discharge tube range from 5 to 100 mA for CW operation.

Such a prescription satisfies the scaling laws of a positive column discharge, a subject to be discussed in greater detail in Chapter 17. For now, we need only the experimental facts of life about low-pressure discharges.

TABLE 10.6 Data Associated with the Various States of Neon

Transition	J_{upper}	J_{lower}	λ (Å)	$A(10^6\ sec^{-1})$	Relative Intensity	
$3s_2 \rightarrow 2p_1$	1	0	7304.9	0.48	30	
$3s_2 \rightarrow 2p_2$	1	1	6401.1	0.6 (est.)	100	
$3s_2 \rightarrow 2p_3$	1	0	6351.9	0.7	100	(common "red" laser)
$3s_2 \rightarrow 2p_4$	1	2	6328.2	6.56	300	(common "red" laser)
$3s_2 \rightarrow 2p_5$	1	1	6293.8	1.35	100	(common "red" laser)
$3s_2 \rightarrow 2p_6$	1	2	6118.0	1.28	100	
$3s_2 \rightarrow 2p_7$	1	1	6046.1	0.68	50	
$3s_2 \rightarrow 2p_8$	1	2	5939.3	0.56	50	
$3s_2 \rightarrow 2p_9$	1	3	5882.5	Forbid $\Delta J = 2$	Not observed	
$3s_2 \rightarrow 2p_{10}$	1	1	5433.6	0.59	250	
$3s_2 \rightarrow \Sigma 2p$	1	—	Red-orange	12.8	—	
$3s_2 \rightarrow 3p_4$	1	2	33913	2.87	—	
$3s_2 \rightarrow \Sigma 3p$	1	—	IR	5.24	—	
$2p_4 \rightarrow 1s_2$	2	1	6678.3	23.8	500	
$2p_4 \rightarrow 1s_3$	2	0	6234.5	Forbid $\Delta J = 2$	Not observed	
$2p_4 \rightarrow 1s_4$	2	1	6096.2	16.9	300	
$2p_4 \rightarrow 1s_5$	2	2	5944.8	10.5	500	
$2p_4 \rightarrow \Sigma 1s$			Red-orange	51.2		

Other transitions	ΣA		λ	$A(\times 10^6)$	
$2p_1 \rightarrow \Sigma 1s$	87.9	$2s_2 - 2p_1$	1.5231 μm	0.802	
$2p_2 \rightarrow \Sigma 1s$	116.6	$2p_2$	1.1767 μm	4.089	
$2p_3 \rightarrow \Sigma 1s$	61.7	$2p_3$	1.1602 μm	0.801	(first gas laser)
$2p_4 \rightarrow \Sigma 1s$	51.7	$2p_4$	1.1523 μm	6.537	(first gas laser)
$2p_5 \rightarrow \Sigma 1s$	53.3	$2p_5$	1.1409 μm	2.301	(first gas laser)
$2p_6 \rightarrow \Sigma 1s$	53.6	$2p_6$	1.0844 μm	7.543	
$2p_7 \rightarrow \Sigma 1s$	49.3	$2p_7$	1.0621 μm	0.816	
$2p_8 \rightarrow \Sigma 1s$	41.2	$2p_8$	1.0295 μm	0.726	
$2p_9 \rightarrow \Sigma 1s$	43.3	$2p_9$	Forbidden	—	
$2p_{10} \rightarrow \Sigma 1s$	33.6	$2p_{10}$	0.8895 μm	1.708	

1. The electron temperature* is directly related to the ratio of the electric field to number density of gas atoms (i.e., $T_e \propto E/N$ or E/p).
2. To a reasonable approximation, E/N is a function only of the pressure times the diameter product.
3. The electron temperature (or E/N) is either constant or decreases slightly with increasing current.

Thus keeping the pd product constant implies a discharge with more or less the same voltage drop per unit length and same electron temperature. Typical values for T_e are $\sim$ 80, 000 K and $E/P = 28$ V/cm/torr at $pd = 0.36$ torr-cm.

*The word *temperature* implies a Maxwellian distribution of electron velocities. In fact, it is not! But it is a reasonable approximation to use for the initial understanding.

The pumping sequence of the red laser is as follows: Helium, being the majority gas present, is excited by the energetic electrons in the high-energy tail of the Maxwellian distribution. This is represented by the following chemical equation:

$$e(\text{K.E.}) + \text{He}(1^1S) \longrightarrow \text{He}(2^1S) + e(\text{K.E.} - 20.6\ \text{eV}) \qquad (10.7.2)$$

where the potential energy of the $\text{He}(2^1S)$ state relative to the ground state is 20.6 eV.

Now this particular state is metastable; that is, it cannot decay back to the ground state by emitting a photon, as such a transition would require $J = 0 \rightarrow J = 0$ step. Consequently, such a state will live for a long time—many seconds in fact—unless something else collides with the excited helium atom. That something else may be another electron, which could excite it further up the helium ladder, or the electron could reverse the arrow of (10.7.2). The excited helium state could also diffuse to the wall to be deactivated there (and warm up the universe). None of these possibilities help the laser a bit, and, in fact, they are deleterious to its operation.

But a neutral neon atom can be that something else that collides with that excited helium atom. The potential energy of the He state is transferred to the neon according to

$$\text{He}(2^1S) + \text{Ne(ground)} + 387\ \text{cm}^{-1} \longrightarrow \text{He}(1^1S) + \text{Ne}(3s_2) \qquad (10.7.3)$$

where the excess energy, 387 cm^{-1}, is provided by the kinetic energy of the colliding atoms.

Thus we have a *selective* mechanism for pumping the upper laser level. Furthermore, since the lower $2p_4$ state has a much shorter radiative lifetime than the $3s_2$ level, it contributes significantly to the creation of the population inversion.

Unfortunately, the $2p_4$ level can also be excited by the discharge independently of the $3s_2 \rightarrow 2p_4$ route, and this fact ultimately dooms the He:Ne laser system to low power. For instance, the helium 2^3S state is excited at an even greater rate than is the 2^1S state. This lower helium metastable state transfers its energy to the $2s$ manifold in neon, which, in turn, radiates to $2p$ level. Indeed, this is the excitation route for the laser transition at 1.1523 $\mu m(2s_2 - 2p_4)$, the first gas laser. If the 6328-Å transition is desired, the $2^3S \rightarrow 2s_2 \rightarrow 2p_4$ route is an unavoidable pumping of the lower laser level. Alternatively, the $2^1S \rightarrow 3s_2 - 2p_4$ is bad for the 1.1523 μm laser. This is an example where one laser transition tends to fill the lower state of another.

A classic example of the competition for the upper state is provided by the 6328-Å and 3.39-μm transitions. The latter, being of longer wavelength and thus having a smaller Doppler width, has a much higher stimulated emission cross section than does the visible transition. Thus, unless special precautions are taken to reduce the feedback at 3.39 μm, the $3s_2 \rightarrow 3p_4$ transition will lase and deplete the population inversion of the $3s_2 \rightarrow 2p_4$.

There is another route for pumping the $2p$ levels: electron impact excitation from the neon ground state or from the neon $1s$ manifold. After all, the neon sign depends on the $2p \rightarrow 1s$ transitions for its characteristic color, and it does indeed work. Therefore the $2p$ states will be excited by a discharge. As we will see shortly, this process is the ultimate culprit for the termination of the laser action at high current.

To model this common 6328-Å laser is not a simple task, because there are at least four $3s$, ten $3p$, four $2s$, ten $2p$, and four $1s$ neon levels and the two helium metastable states to be related to the voltage and current through the discharge. This would require the

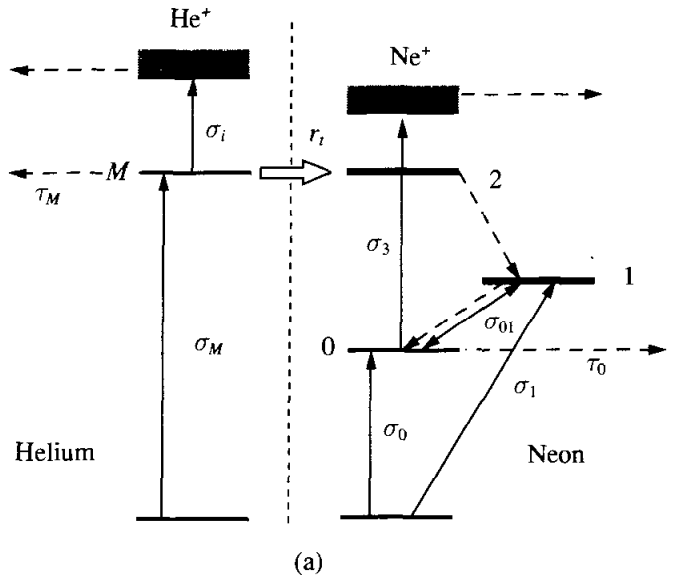
simultaneous solution of 34 rate equations for the atomic populations and one rate equation for the electron and ion density. Obviously, that is not feasible here.

But, with some judicious approximations, we can gain an understanding of the gross behavior of the laser as the current through the discharge is varied. This is done in the following example.

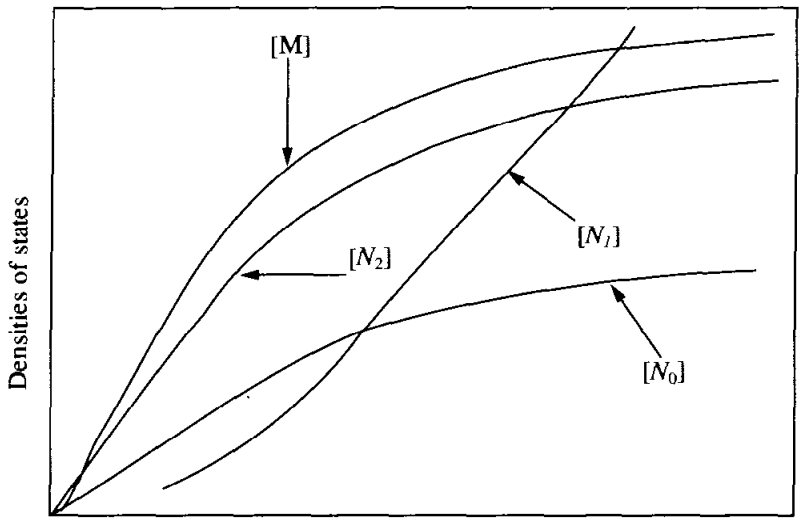
Example: Pumping of the helium-neon 6328-Å laser

It is obvious that we must reduce the number of variables and cross couplings to even remotely handle the situation by analytic techniques. The model chosen is shown in Fig. 10.29(a). Our approximations are

1. Ignore the rate equation for the electrons and ions and assume that the electron density is directly proportional to the current, in agreement with experiment [12].



(a)



(b)

FIGURE 10.29. (a) Simplified model of the helium-neon laser. The solid lines represent electron excitation routes, the dashed lines represent diffusion to the walls of the tube, and the dashed lines represent spontaneous decay; and (b) the variation of the states with current.

2. The excitation of the helium metastable proceeds by electron impact with helium in the ground state. Once formed, M can transfer its energy to neon at a rate r_t (per helium metastable atom per neon atom), diffuse to the wall and be deactivated, or be ionized by a second collision with another electron. We ignore the fact that the neon state N_2 can transfer its potential energy back to helium and also ignore the deactivation of M by superelastic collisions with electrons.
3. The only pumping mechanism allowed for the upper laser level is the transfer indicated: N_2 decays by spontaneous emission to the [1] manifold and subsequently to the [0] state (the $1s$ levels).
4. There are two primary pumping routes for the lower laser level; that is, electron impact excitation from the ground state and from the neon $1s$ manifold N_0. We ignore the pumping of N_1 by spontaneous decay from N_2 as being small compared to the electron pumping rate. The lower state decays by spontaneous decay back to N_0.
5. The neon metastable states are excited by electron impact with ground-state neon atoms and by spontaneous decay from N_2. We can balance the spontaneous decay into N_0 against the electron deexcitation routes to N_1 from N_0 and the direct electron excitation of N_1. The net deexcitation routes are diffusion to the walls and ionization by a second electron collision.

With these assumptions, the 35 rate equations simplify to four:

$$\frac{d[M]}{dt} = n_e \langle \sigma_M v_e \rangle [\text{He}] - \frac{[M]}{\tau_M} - r_t[M][\text{Ne}] - n_e \langle \sigma_i v_e \rangle [M] \tag{10.7.4}$$

$$\frac{d[N_2]}{dt} = r_t[M][\text{Ne}] - A_2[N_2] \tag{10.7.5}$$

$$\frac{d[N_1]}{dt} = n_e \left\{ \langle \sigma_1 v_e \rangle [\text{Ne}] + \langle \sigma_{01} v_e \rangle [N_0] \right\} - A_1[N_1] \tag{10.7.6}$$

$$\frac{d[N_0]}{dt} = n_e \langle \sigma_0 v_e \rangle [\text{Ne}] + A_1[N_1] - n_e \langle \sigma_{01} v_e \rangle [N_0] + A_2[N_2] - \frac{[N_0]}{\tau_0} - n_e \langle \sigma_3 v_e \rangle [N_0] \tag{10.7.7}$$

where all quantities in square brackets are densities and the symbol $\langle \sigma v_e \rangle$ indicates that the cross section for the process must be multiplied by the electron velocity and averaged over the electron distribution function (see Chapter 17).

We set all time derivatives equal to zero and set about the unpleasant task of solving these coupled equations. The first two are simple:

$$[M] = \frac{n_e \langle \sigma_m v_e \rangle [\text{He}]}{\tau_M^{-1} + r_t[\text{Ne}] + n_e \langle \sigma_i v_e \rangle} \tag{10.7.8}$$

$$[N_2] = \frac{r_t[\text{Ne}][M]}{A_2} = \frac{1}{A_2} \frac{n_e \langle \sigma_m v_e \rangle r_t[\text{He}][\text{Ne}]}{\tau_M^{-1} + r_t[\text{Ne}] + n_e \langle \sigma_i v_i \rangle} \tag{10.7.9}$$

By combining (10.7.6) with (10.7.7) we first find the density of the neon metastable state $[N_0]$:

$$[N_0] = \frac{n_e \langle \sigma_0 v_e \rangle + \langle \sigma_1 v_e \rangle [\text{Ne}]}{\tau_0^{-1} + n_e \langle \sigma_3 v_e \rangle} + \frac{A_2[N_2]}{\tau_0^{-1} + n_e \langle \sigma_3 v_e \rangle} \tag{10.7.10}$$

where $[N_2]$ is given by (10.7.9). Then we return to (10.7.6) to relate $[N_1]$ to all the prior terms.

$$[N_1] = n_e \langle \sigma_1 v_e \rangle [\text{Ne}] + n_e \langle \sigma_{01} v_e \rangle [N_0] \tag{10.7.11}$$

To evaluate these equations is to experience an arithmetic nightmare. The trend should be clear if we mentally substitute current I for the electron density in the above and sketch the variation of the states with this externally controlled variable.

This is done in Fig. 10.29(b). In examining this sketch, one should follow the order of the previous four equations. For instance, (10.7.8) indicates that $[M]$ should increase linearly with n_e until ionization collisions become more frequent than the deactivation rates of diffusion to the walls or transfer to neon. Equation (10.7.9) indicates that $[N_2]$ has more or less the same functional dependence on current as does $[M]$.

Equation (10.7.10) is complicated unbearably by the last term, but since $[N_2]$ saturates (i.e., becomes independent of current) in the same manner as does the first term, the variation of $[N_0]$ is similar to $[M]$.

The two terms of (10.7.11) play a crucial role in establishing a limitation to the helium-neon laser and relegate it to the low-power domain. Both terms state that the lower laser level varies linearly with current. Thus, since $[N_2]$ saturates at high currents, $[N_1]$ is bound to catch up and exceed* $[N_2]$. The inversion is destroyed, and the laser ceases operation!

Although the helium-neon laser is doomed to small power, it has practically cornered the market for alignment and other low-power applications. Even its paltry 1 to 10 mW is very intense compared to incoherent light sources. In numbers produced and sold, it is far in front of whichever gas laser is in second place.

10.7.3 Ion Lasers

The atomic levels associated with a laser can be the excited states of an ion, and many elements have been used in this manner.† Two of the most common ion lasers use the heavy rare gases krypton and argon for the donor atoms.

The energy-level diagram for the Ar^+ laser is shown in Fig. 10.30. As is shown in this figure, we can obtain nearly 50 different laser transitions in the excited neutral argon atom at wavelengths ranging from 0.7 to 30 μm, and "visible" transitions in argon ion with the common ones shown there.

Such a diagram allows us to review the selection rules for atomic transitions (see Chapter 15). The small numbers to the right of each energy level are J values of each state, and the superscript preceding the capital letter is the multiplicity $(2S + 1)$ of the group of states. Note that all transitions obey $\Delta J = \pm 1, 0$ and $\Delta L = \pm 1, 0$. As indicated in Chapter 15, the $\Delta S = 0$ rule is usually the first to be broken, and the strong 5145-Å line is an example.

It is also important to note that the energy scale of Fig. 10.30 has *two* breaks: one between the ground state of neutral argon and the first ionization at 15.75 V and the other between this value and the lower laser level. To reach the upper laser level of the 4880-Å

*A more complicated model would show that eventually the ratio of N_2/N_1 would be $g_2/g_1 \exp(-\Delta E/kT_e)$, a normal population equilibrated at the electron temperature.

†We have laser lines in the excited ions of Cd, Se, Hg, Ge, Ag, Au, Pb, and so on. Unfortunately, many of these elements require extreme temperatures to obtain a reasonable vapor pressure. (For instance, the vapor pressure of gold at 900°C is only 10^{-11} torr.)

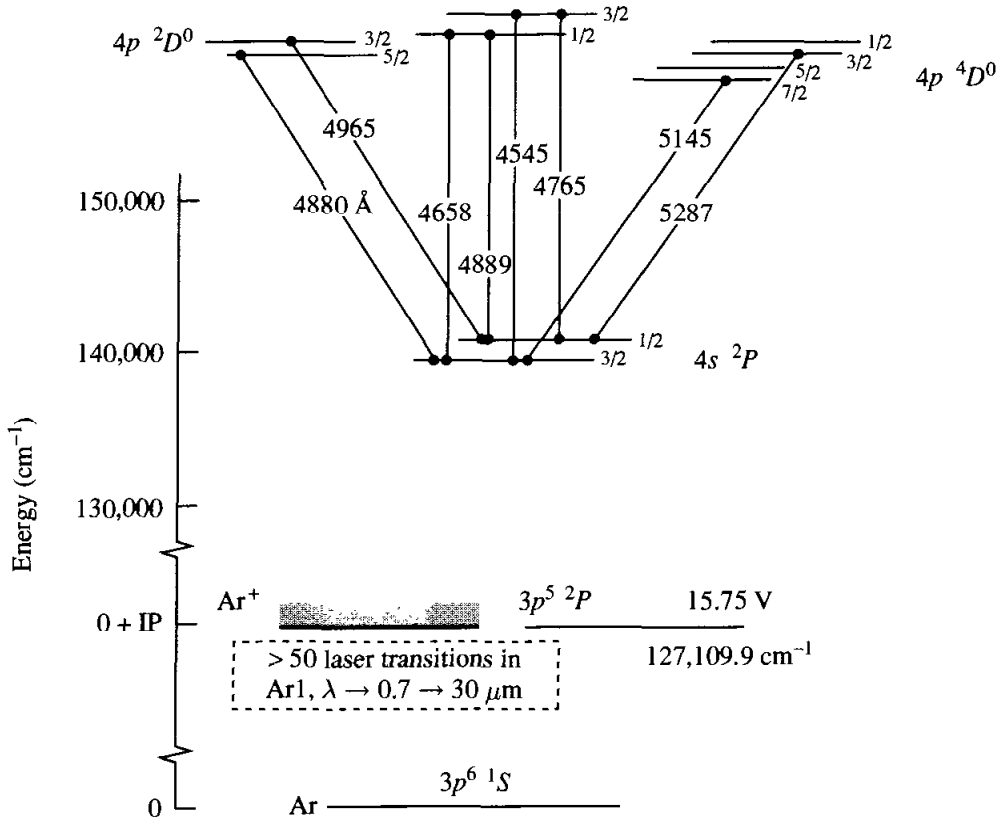


FIGURE 10.30. Energy-level diagram for the argon-ion laser.

line at 158,731.20 cm^{-1}, we must provide an additional 127,109.9 cm^{-1} to first create an ion, or a total of 35.43 V.

There is a reason for the emphasis on numbers here. First, the maximum efficiency of this laser—the quantum efficiency—is low, 7.2%. Furthermore, the upper state of the laser is higher than the highest first ionization potential of any gas. Thus we cannot hope for a selective excitation transfer mechanism as found in the helium-neon laser. Rather, we must pump the upper level from the ion ground state or by direct excitation from the neutral. In view of the fact that the high-energy tail of the electron distribution drops as $\exp(-\epsilon/kT)$, there are many more electrons capable of first ionizing the atom (15.75 V) and then another electron exciting the ion (19.68 V) than have the energy to excite the $4^2pD^0_{5/2}$ state directly from the neutral argon ground state.

One consequence of this two-step excitation process is that an argon-ion laser discharge is very intense, with amperes being carried by a small-bore discharge, which, in turn, requires water cooling to survive such treatment. Indeed, the survival of the wall material is the major limitation to the generation of intense visible CW radiation.

Typical efficiencies of the laser are far worse than the quantum efficiency noted above, with 5-W output (in all lines) being typical from a laser run from a power supply requiring $\sim$ 12 kW from a three-phase 460-V, 60-Hz plug. This is an overall efficiency of 0.03%. Thus there is considerable laser engineering in the not-so-trivial problem of building the power supply.

In spite of the poor efficiency, the argon (and krypton) ion lasers are very common, a major use being the optical pump for the tunable dye lasers. In that application, we can expect a conversion efficiency of 10% to 30% between the pump and the dye laser output.

10.7.4 CO_2 Lasers

The CO_2 laser is one of the most spectacular, most efficient, and because of these characteristics, most useful and most studied lasers discovered to date. Whereas most gas laser efficiencies are measured in tenths of a percent, the CO_2 laser is measured in tens. A "small" laser can produce tens of watts and, by using a master oscillator power amplifier (MOPA) chain, we can achieve terawatts of pulsed power,* or we can obtain hundreds of kilowatts in a CW mode. The latter two devices are not small. Applications of this laser have been numerous and diverse. For instance, CO_2 is used in trimming operations in industrial manufacture, pattern cutting, communications, welding and cutting of steel (many inches thick), weaponry, and laser fusion, to list only a few applications.

Oscillation can be obtained on more than 200 vibrational-rotational (VR) transitions in the 8 to 18 μm wavelength range. Many of the desirable characteristics of the CO_2 laser can be attributed to the kindness of nature in providing us with a linear symmetric molecule that has a characteristic vibrational energy in a near-perfect match with the characteristic vibrational energy of nitrogen (N_2).

The pertinent energy levels of the CO_2 and N_2 molecules are shown in Fig. 10.31, where some classic artifacts are added to aid the interpretation.

The states are usually described by the number of quanta in the characteristic vibration of the classic linear molecule shown at the top. Thus the $10^{\circ}0$ state has one quanta in the characteristic ν_1 or symmetric stretching mode, the $02^{\circ}0$ state has two quanta in the bending mode, and the $00^{\circ}1$ state has a single quanta in the asymmetric stretch.† As shown in the diagram, lasing takes place between the *odd* rotational levels of the 001 state and the *even* rotational states of either the 100 or 020 states [13, 14].

Note that the first vibrational level of N_2 coincides very closely with the upper laser level (001). Indeed, most of the lower vibrational levels of N_2 from $\upsilon = 1$ to 8 are spaced to have an excellent match with the (000) to (001) separation of CO_2. Recall that N_2 is a homonuclear molecule, and thus radiative transitions between vibrational levels are strictly forbidden (see Chapter 15). In other words, the vibrational levels of N_2 are metastable. Thus, if one can vibrationally excite these N_2 molecules, they can transfer this trapped energy selectively to the upper laser level.

Finally, note that the excited-state energy levels of helium could not be placed on this diagram without adding 68 additional pages pasted side by side in a foldout fashion. The first excited state of helium occurs at 159,850 cm^{-1}, whereas the upper laser (001) level is only at 2349 cm^{-1}. This contrast explains in part why the CO_2 laser is so much more

*At that power level, the laser has to be pulsed. After all, the total installed power-generating capacity in the United States in 1975 was only 525 million kilowatts, or 0.525 terawatt.

†Unfortunately, this simple one-to-one identification with the classic vibration is only an approximation. However, it suffices for the initial introduction to this laser. In the same spirit, we ignore the meaning of the superscript on υ_2.

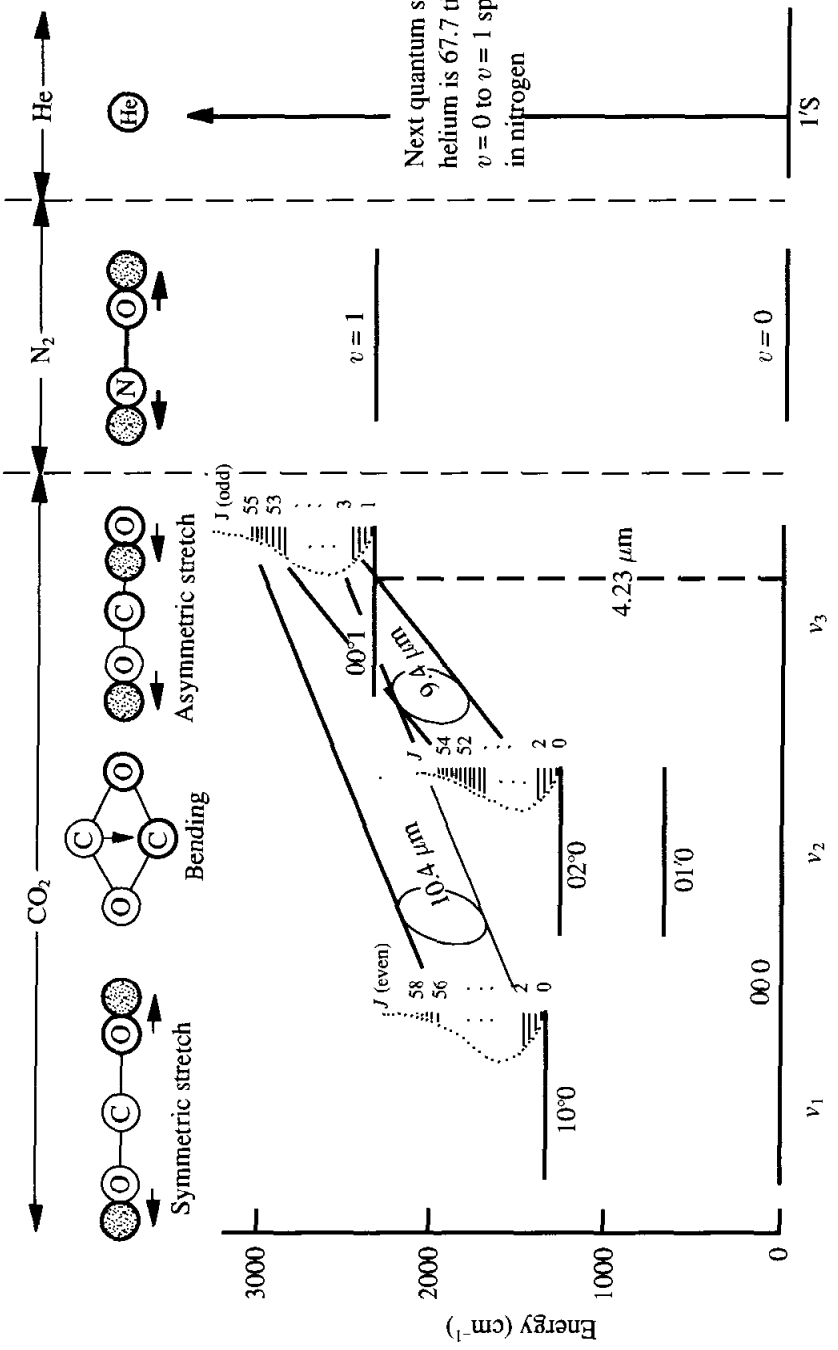


FIGURE 10.31. Energy-level diagram of the CO_2-N_2-He laser.

efficient than the helium-neon laser. One needs to supply $\sim \frac{1}{2}$ eV to excite the (001) state of CO_2, whereas ≥ 19.8 eV is required to excite the upper state of the helium-neon laser.

An elementary pumping sequence for an electrical discharge in a CO_2-N_2-He mixture (for a typical 1 : 2 : 3 pressure ratio) is as follows:

1. The electrical power is transferred to the electrons (as is the case in *all* discharges) by the electric field.
2. The electrons transfer this power by collisions to the neutral gas atoms. This power is apportioned to the gas in three different categories:
   a. *Gas heating:* This is caused by the "elastic" collisions of the very light electrons with the more massive neutral atoms. Although these collisions are mostly elastic, there are many such collisions, and some energy is expended in raising the kinetic temperature of the gas.
   b. *Vibration excitation:* This is an inelastic collision process represented by the following chemical equations:
      1. For the upper state:

$$e + N_2(v = 0) \longrightarrow N_2(v = n \leq 8) + (e - \text{KE}) \tag{10.7.12}$$

followed by

$$\begin{aligned} N_2(v = n) + CO_2(000) \\ \longrightarrow N_2(v = n - 1) + CO_2(001) \end{aligned} \tag{10.7.13}$$

or

$$e + CO_2(000) \longrightarrow CO_2(001) + (e - \text{KE}) \tag{10.7.14}$$

      2. For the lower state:

$$\begin{aligned} e + CO_2(000) \longrightarrow CO_2(010) \\ \text{or } CO_2(020) \\ \text{or } CO_2(100) + (e - \text{KE}) \end{aligned} \tag{10.7.15}$$

   c. *Electronic excitation and ionization:* Although ionization is essential to maintain an active discharge, the fraction of the electrical power used to do so is usually insignificant in discharges in molecular gases.
3. Theory and experiment show that 60% of the electrical power can be funneled into pumping the upper laser level (see Chapter 17).

Now the quantum efficiency of the 9.4-μm laser is 45%, and with 60% entering the upper state, we could expect a "wall plug" efficiency of $\sim$ 27%. Such numbers have been approached.

To understand why this is achieved in a CO_2-N_2-He laser (but not in most other molecular lasers) requires us to examine the details of the foregoing reactions. This will

TABLE 10.7 Data Associated with the CO_2 Laser

Fundamental frequencies

$01^{1}0 \rightarrow 00^{0}0 = 667.3\ \text{cm}^{-1}$	$00^{0}1 \rightarrow 00^{0}0 = 2349.3\ \text{cm}^{-1}$
$02^{0}0 \rightarrow 00^{0}0 = 1285.5\ \text{cm}^{-1}$	$00^{0}1 \rightarrow 10^{0}0 = 960.8\ \text{cm}^{-1}$
$10^{0}0 \rightarrow 00^{0}0 = 1388.3\ \text{cm}^{-1}$	$00^{0}1 \rightarrow 02^{0}0 = 1063.6\ \text{cm}^{-1}$

Rotational data

$$B(00^{0}1) = 387140.44 \times 10^{-6}\ \text{cm}^{-1}$$
$$D(00^{0}1) = 13.252 \times 10^{-8}\ \text{cm}^{-1}$$
$$B(10^{0}0) - B(00^{0}1) = 3047.389 \times 10^{-6}\ \text{cm}^{-1}$$
$$D(10^{0}0) - D(00^{0}1) = -1.8366 \times 10^{-8}\ \text{cm}^{-1}$$
$$B(02^{0}0) - B(00^{0}1) = 3340.757 \times 10^{-6}\ \text{cm}^{-1}$$
$$D(02^{0}0) - D(00^{0}1) = 2.3816 \times 10^{-8}\ \text{cm}^{-1}$$

Transition probabilities (Einstein A coefficients)

	$00^{0}1 \rightarrow 00^{0}0$* (4.2-$\mu$m band)	$00^{0}1 \rightarrow 10^{0}0$ (10.4-μm band)	$10^{0}0 \rightarrow 02^{0}0$ (9.4-μm band)
P branch	$2.1 \times 10^{2}\ (\text{sec}^{-1})$	0.34	0.20
R branch	2.0×10^{2}	0.33	0.19
N_2 data:	$\omega_e = 2359.1\ \text{cm}^{-1}$	$\omega_e x_e = 14.456\ \text{cm}^{-1}$	$\omega_e y_e = 0.00751\ \text{cm}^{-1}$
	$B_e = 2.010\ \text{cm}^{-1}$	$\alpha_e = 0.0187\ \text{cm}^{-1}$	$r_e = 1.094$ Å

*Resonance trapping of this radiation has a profound effect on the effective radiative lifetime. For instance, the P branch has an effective transition probability of 10.1 sec^{-1} rather than 210 sec^{-1}, as listed.

Data from Cheo [13].

be done shortly. Before doing so, let us return to the energy-level diagram (Fig. 10.31) to discuss some of the details of the laser transitions.

Table 10.7 lists some of the pertinent data for CO_2. The fundamental "frequencies" indicate the energy levels of the various states involved, ignoring rotation. Thus the $J = 0 \rightarrow J = 0$ of the $00^{0}1 - 10^{0}0$ transition would appear at 960.8 cm^{-1} (were it not for the fact that such a transition is forbidden). The position of the rotationless $J = 0$ state of $00^{0}1$ is 2349.3 cm^{-1}, which is the sum of the corresponding frequencies of $10^{0}0 \rightarrow 00^{0}0$ and $00^{0}1 \rightarrow 10^{0}0$, $1388.3 + 960.8 = 2349.1$, or the sum of $(02^{0}0 \rightarrow 00^{0}0) + (00^{0}1 \rightarrow 02^{0}0)$, $1285.5 + 1063.6 = 2349.1$. The fact that these sums do not exactly match the specified energy of the $00^{0}1$ state indicates a small error in one or more of the measurements.

Note the value of the Einstein A coefficients, or the inverse, the radiative lifetime for the laser transitions. This lifetime is 3 to 5 seconds—or 3 to 5 heartbeats. Even the resonance transition $00^{0}1$ to $00^{0}0$ has a paltry 10/sec rate. These numbers illustrate a very important point about the CO_2 laser (and others operating on VR transitions):

> *Spontaneous* radiation is not a significant loss process for the laser states and is usually ignored in a kinetic rate equation.

The energy exchange between colliding gas molecules is overwhelmingly more significant than is spontaneous emission.

There are two consequences that might seem paradoxical at first glance, given the wonders and efficiency of the laser.

1. In a pulsed CO_2 laser, it takes a significant amount of time to generate enough photons in the proper cavity mode to start the laser oscillating. Indeed, the gain can build up faster than the photons and leads to a *gain-switched* pulse (see Sec. 9.3).
2. The stimulated emission cross section is small, typically less than 10^{-18} cm^2.

Consequently, there must be an appreciable fraction of the molecules excited to obtain a reasonable gain.

The last item explains the high-power, high-energy characteristics of the CO_2 laser. If we excite a large percentage of the fill gas to the proper state, each molecule can contribute one photon by the stimulated emission process and we obtain many, many photons. Let us illustrate this point by an example.

Example

Consider a gas fill of 1 : 2 : 3 mixture of CO_2-N_2-He at 10 atm and assume that the excitation enters the 001 level by the N_2 route (10.7.12) and (10.7.13). Let us assume a pulsed discharge that is to create a vibrational "temperature" in N_2 of 0.5 eV. If every vibrational state in N_2 from $\upsilon = 1$ to $\upsilon = 8$ is capable of creating a 10.6-μm CO_2 photon, how much energy is involved?

Solution: Partial pressure of $N_2 = 2/(1 + 2 + 3) \times 10 = 3.33$ atm. Thus the concentration of N_2 is $3.33 \times 2.69 \times 10^{19} = 8.96 \times 10^{19}$ cm^{-3}. The number of the nitrogen molecules in vibrational states $\upsilon = 1$ to 8 is given by the probabilities that the energies, $hc\omega_e(\upsilon + \frac{1}{2})$, can occur, divided by the sum of the probabilities for *all* vibrational states, $\upsilon = 0$ to $\upsilon = \infty$, times the concentration of nitrogen molecules.

$$N_2(\upsilon = 1 \rightarrow 8) = [N_2] \frac{\sum_{1}^{8} \exp[-hc\omega_e(\upsilon + \frac{1}{2})/kT]}{\sum_{0}^{\infty} \exp[-hc\omega_e(\upsilon + \frac{1}{2})/kT]} \tag{10.7.16}$$

where the anharmonic terms for the vibrational energy levels have been ignored. For $\omega_e = 2359.1$ cm^{-1} (see Table 10.7) and $kT/e = 0.5$ eV, the ratio of the two series is 0.553, or there are 4.48×10^{19} cm^{-3} molecules capable of producing a 10.6-μm photon (0.1169 eV/photon). Thus, in principle, we should be able to extract 0.838 J/cm^3 or 838 J/liter. We do not do quite this well, but the calculation is close. Such numbers indicate the energy-storage capability of the CO_2-N_2 laser system.

As mentioned earlier, the CO_2 laser can oscillate on any of the P or R branches shown in Fig. 10.31. In a simple laser cavity without any wavelength discrimination, the P branch of the 10.4-μm band will dominate. This is because the factors in the laser-gain equation

$$\gamma(\nu) = A_{21} \frac{\lambda^2}{8\pi} g(\nu) \left(N_2 - \frac{g_2}{g_1} N_1 \right)$$

all favor this collection of lines: the A coefficient is bigger for the $00^\circ 1 \rightarrow 10^\circ 0$ group of transitions (see Table 10.7), the wavelength is larger, and the P branch is always favored

over the R branch (see Chapter 15 for an analogous situation in CO). Thus the P-branch oscillation on the 10.4-μm band competes for and wins the photons associated with the 001 $\rightarrow$ 100 (or 020) inversion at the expense of the gain on the R branch of the 10.4-μm band and both branches of the 9.4-μm band.

With sufficient frequency discrimination in the cavity, such as that provided by a grating, the P (4), P (6), . . . , P (56) and R (4), . . . , R (54) of the 00°1 $\rightarrow$ 10°0 transition lase as well as the P (4) to P (60) and R (4) to R (56) of the 00°1 $\rightarrow$ 02°0 band.* To compute the wavelengths of these transitions is a straightforward but dull arithmetic task (which is best left as a problem). Incidentally, some of these wavelengths have been measured by beating the laser against a high harmonic of a microwave-crystal-controlled source, and this accounts for the extreme precision of the data presented in Table 10.7.

Let us now return to the excitation process. The figures presented for the energy storage are so much chicanery if we cannot establish a vibrational temperature with electron collisions. We can. As mentioned earlier, nature kindly provides us with a large cross section for electron excitation of the vibrational levels of N_2. Figure 10.32(a) shows the total excitation cross section of $v = 1$ to 8 states, with Fig. 10.32(b) showing only the excitation of the $v = 1$ state as measured by G. Schulz.

There is considerable physics buried beneath such simple-looking curves. First note that the cross section is virtually zero below about 1.5 eV. It only takes an electron of energy $G(1) - G(0) \simeq 0.3$ eV to excite the $v = 1$ state. Why is the apparent threshold so high? The reason is that the excitation proceeds via a complex route. The electron is first attached to the N_2 molecule, makes a few swings around it, and then flies off, leaving the N_2 in the

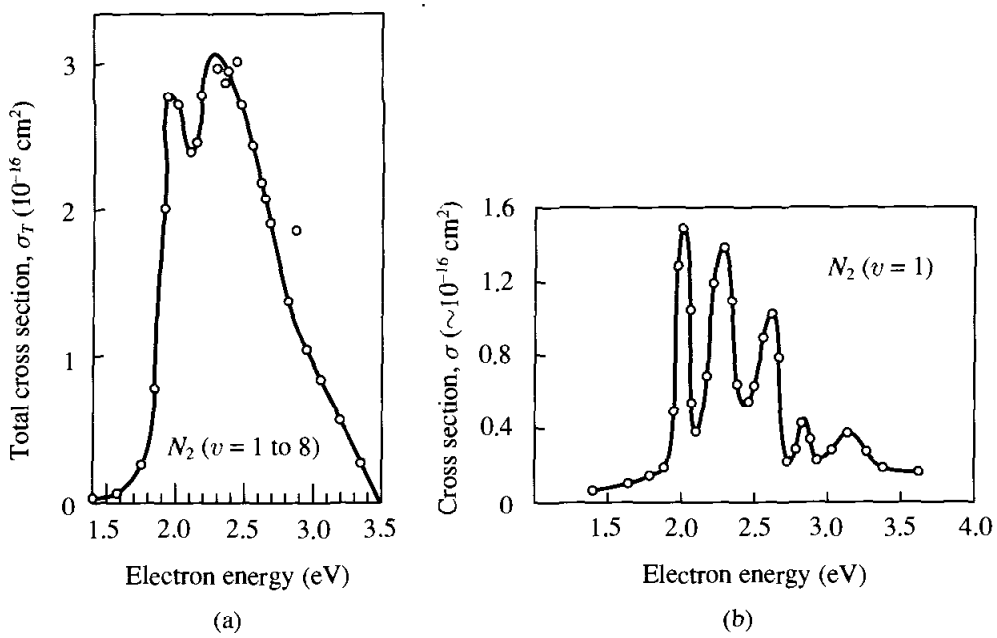


FIGURE 10.32. Vibrational excitation cross section. (Data from Schultz [15].)

*Remember that only even J values appear in the rotational structure of the 100 and 020 states, and only odd J values appear in the 001 state.

vibrational excited state. The chemical equations describing the process are

$$e(E) + N_2 \longrightarrow N_2^- | \longrightarrow N_2(v) + e(E - G(v))$$

Thus the upward shift in the energy scale is caused by the shift of N_2^- with respect to N_2.

We studiously avoided any all discussion of the role of helium except to note that none of its excited states are even on the map of the energy-level diagram of Fig. 10.31. Yet, more often than not, it is the majority gas. Let us now indicate its role.

Helium has many purposed in the CO_2 laser, some rather mundane and others quite subtle. A rather mundane role is to help transport the useless and deleterious heat generated by the discharge to the walls of the container for cooling. A more subtle role is to relax the lower laser level. However, its primary purpose is to control the E/N (or E/p) and thus the electron "temperature" of the discharge, which, in turn, controls the relative division of the electrical power to the processes noted previously. We postpone discussion of this most important role until Chapter 17.

The CO_2 laser is probably the most extensively studied and analyzed system in existence. It certainly has a commanding lead in performance.

10.8 EXCIMER LASERS: GENERAL CONSIDERATIONS

The rare gases (helium, neon, argon, krypton, and xenon) have always been considered models for chemical *inertness*. For the most part, these gases do not form compounds with anything else, and hence they normally exist as simple monatomic gases. The crude and simple explanation usually given is that these gases have "closed" or "filled" shells for the orbiting electrons and thus do not participate in a chemical reaction.

However, it was recognized that if the rare gas was excited, this simple line of reasoning would no longer apply. For then, one of the electrons would be in the next "shell" and thus an excited rare gas (Rg*) would appear as the next element over in the periodic chart. Thus excited neon (Ne*) might act as Na in a chemical reaction, Ar* as potassium, Kr* as rubidium, and Xe* as cesium. Since an alkali-halogen reaction such as

$$K + F_2 \longrightarrow (K^+F^-) + F \tag{10.8.1}$$

proceeds with vigor, one might also expect a reaction such as

$$Ar^* + F_2 \longrightarrow (Ar^+F^-)^* + F \tag{10.8.2}$$

to do likewise. Note that the "salts", such as NaCl, can be considered as bound by ionic attraction, and thus we expect the state formed by (10.8.2) to be similar. Here, however, we would expect the rare gas-halide "salt" to be in an excited state whose energy is related to the initial excited state of the rare gas. If the rare gas were not excited, no reaction with the halogen donor should take place because of the filled shells. They would merely repel each other at small nuclear spacing. This simple line of reasoning led to the invention of the *r*are *g*as-*h*alide (RgH) lasers, which are part of a class of lasers called "excimers," a system that only exists for a measurable time in an excited state.

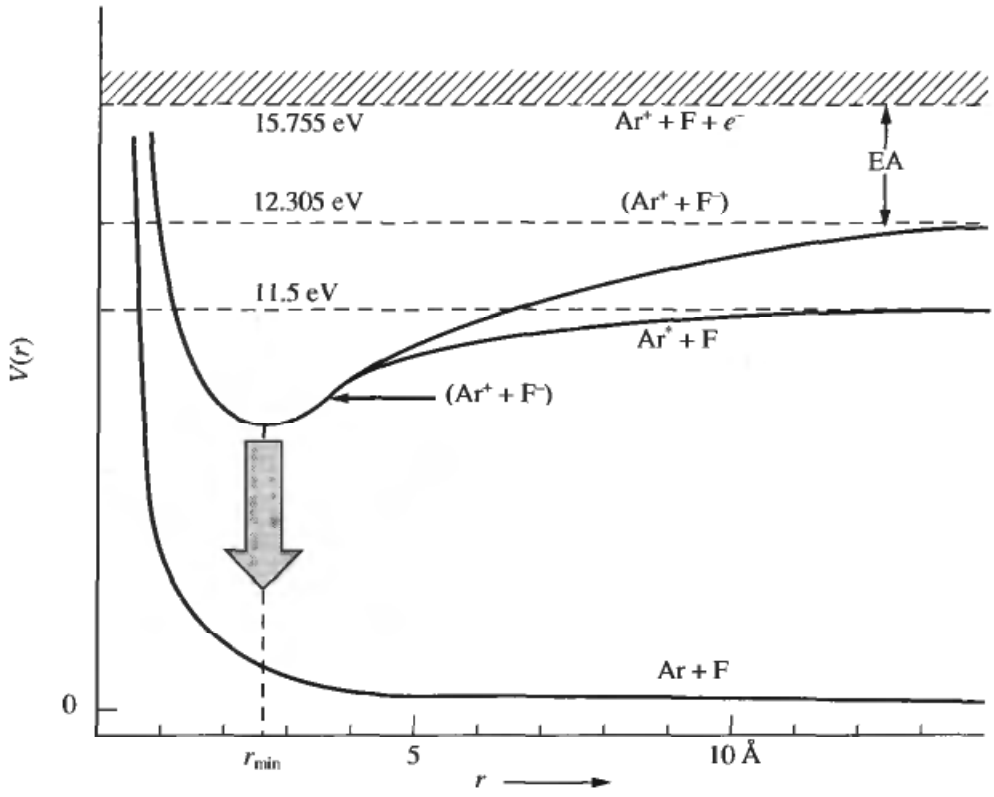


FIGURE 10.33. Energy-level diagram associated with the formation of the (Ar^+F^-) excimer.

10.8.1 Formation of the Excimer State

As a specific example, consider Fig. 10.33, which shows the energy level diagram associated with interaction of an argon ion, an electron, and a fluorine atom. At very large inner nuclear spacing, one could have an argon ion plus a free electron and a neutral fluorine atom with a total potential energy of the interaction equal to the ionization value of argon (15.755 eV). Now let us get rid of the electron by attaching it to fluorine making a negative ion. The electron affinity of F is about 3.45 eV; hence the asymptotic limit of potential energy between Ar^+ and F^- at large nuclear spacing is $15.755 - 3.45 = 12.305$ eV. These two ions attract each other according to Coulomb's law, and thus the potential energy curve decreases as the internuclear spacing decreases. At very small distances, $1 \rightarrow 2$ Å, other repulsive short-range forces come into play so as to prevent two atoms being at the same place at the same time, and thus there is a potential minimum at about 2.3 Å (from experiment).

Let us estimate the potential energy depression between $r = \infty$ and $r = 2.3$ Å due to this Coulomb attraction between Ar^+ and F^-. The electrostatic energy is simply

$$\Delta V = \frac{e}{4\pi\epsilon_0 r_{\min}} = 6.54 \text{ eV} \tag{10.8.3}$$

Thus the potential depression due to the Coulomb attraction is 6.54 eV, and the total energy is now $12.305 - 6.54 = 5.77$ eV. Now we have to add some of the repulsive interaction to

prevent $r \rightarrow 0$, about which we know little except that it exists, but 5.77 eV is a reasonable estimate for the potential energy of the lowest ionically bound state of (Ar^+F^-).* (The actual value is $\sim$ 6.7 eV.)

If neither of the participants, Ar and F, is excited, the ground state of the complex consists merely of a repulsive potential preventing the two atoms from getting too close together and the actual energy at $r_{min} = 2.2$ Å amounts to $\sim$ 0.29 eV. Any argon and fluorine atoms appearing at this spacing will immediately fly apart in a time estimated by the following simple calculation.

> They only have to move 1 Å before the repulsive forces are negligible and in that distance accelerate to a kinetic velocity corresponding to 0.29 eV. Thus setting $\frac{1}{2} M v^2 = e\,(0.29)$, and using M as the sum of the two masses (Ar = 40 amu and F = 19 amu), we obtain a velocity of 9.7×10^4 cm/sec. If they start at zero velocity and end with 9.7×10^4 cm/sec, pick an average velocity of 4.85×10^4 cm/sec for their speed. Thus the lifetime of the lower state of ArF at $r_{min} = 2.2$ Å is only $(10^{-8}$ cm) $\div$ 4.8×10^4 cm/sec $= 2 \times 10^{-13}$ sec, which is quite an adequate approximation to zero. It is so short that there is no reason to make a better calculation.

The radiative lifetime of the ionically bound excimer state at $r = 2.2$ Å is short by normal electronic standards but huge compared to the lower state lifetime. Thus we have one of the best approximations to an *ideal laser* system known at the present time. The main question is whether we can *produce* this ionically bound upper state in a practical sense. Obviously, the previous kinetic sequence requires copious quantities of Ar^+ and F^- (hence it is called the ionic channel) and something to stabilize the complex at $r_{min} = 2.2$ Å. Let us postpone the details of the production process for a moment so as to describe another channel for the formation of the ionically bound excimer.

This second process is very similar to the "burning" of an alkali metal in a halogen gas as described by (10.8.1). Since (10.8.2) implies the "alkali-like" atom is an excited state of the rare gas, when it and a neutral halogen donor, F_2, NF_3, and so on, get close together at $r \sim 3$ to 4 Å, the excited electron of the rare gas finds that it is energetically favorable to jump to the halogen atom creating a negative ion and leaving behind a positive rare gas ion. These two attract and follow the same interaction curves discussed previously. This extraction of a halogen from its donor by the "alkali-like" excited state of a rare gas is aptly named "*the harpooning reaction.*" It occurs with nearly unit efficiency; that is, every collision of Rg* with these donors yields $(Ar^+F^-)^*$. The extra atoms of the fluorine donor and the $(Ar^+F^-)^*$ complex share the excess energy of the reaction and stabilize the excimer in the upper state. The excited rare gas precursor is usually a metastable state, and thus this last process is called the "metastable channel."

The choice of the ArF system was motivated by the fact that the photon obtained by the radiation from the excimer state to ground at $(6.7 - 0.29 = 6.41$ eV) is one of the shortest wavelengths available ($\lambda = 193$ nm). However, other rare gases and other halogens can be used to generate a variety of wavelengths. Table 10.8 illustrates the matrix formed by considering combinations between the various rare gases and various halogens. Note that the wavelengths become shorter as lighter rare gases are used, which have higher ionization potential and metastable levels, and heavier halogens, which have lower electron affinities.

TABLE 10.8 Rare Gas–Halide Wavelengths (nm)

Halogen	EA		Rare Gas			
			Neon	Argon	Krypton	Xenon
		IP (eV)	21.56	15.755	13.996	12.127
		M (eV)	16.6	11.55	9.92	8.31
			nm	nm	nm	nm
Fluorine	(3.45)		108	**193**	**249**	**351**
Chlorine	(3.61)		—	**175**	**222**	**308**
Bromine	(3.36)			161	**206**	**282**
Iodine	(3.06)				185	253

IP, ionization potential; M, metastable level; EA, electron affinity. Wavelengths in boldface type refer to the peak of the laser; those in lightface type refer to the fluorescence assignable to an excimer (see text) and have not yet lased.
Data from Brau [31].

This follows from the scenario described for ArF (and allows the student to follow it and predict the wavelengths).

All the combinations in Table 10.8 follow the same general scenario as described for ArF. There is a slightly amusing wrinkle for XeF in that the ground state is actually bound by about 1200 cm^{-1}, and thus there are a few stable vibrational states in it. (This proves once again that the molecules cannot read our textbooks!) However, since the binding is so small (only $\sim$ 6 kT), it is easily dissociated and does not prevent lasing.

Table 10.9 presents some of the pertinent data for the common rare gas–halogen lasers and was taken from the work by Brau (Ref. 31).

TABLE 10.9 Data on Rare Gas–Halide Laser Systems

Excimer	r (Å)	ω_e (cm^{-1})	σ (10^{-16} cm^2)	τ (ns)	λ (nm)
XeBr	3.1	120	2.2	12–17.5	282
XeCl	2.9	194	4.5	11	308
XeF	2.4	309	5.3	12–18.8	351
KrCl	2.8	210	—	—	222
KrF	2.3	310	2.5	6.7–9	249
ArCl	2.7	(280)	—	—	175
ArF	2.2	(430)	2.9	4.2	193

r_e, minimum of the lowest ionically bound excimer state; ω_e, vibration constant representative; σ, stimulated emission cross section; τ, radiative lifetime; λ, dominant laser wavelength. These lasers hold a commanding lead as far as efficiency in the production of UV and near-UV power. Values in parentheses are estimates. Data from Brau [31].

10.8.2 Excitation of the Rare Gas–Halogen Excimer Lasers

It should be clear that any excitation scheme requires the production of copious quantities of rare gas ions or excited states, or both. Before describing various techniques for producing them, let us estimate a lower bound for the pumping power required to establish a typical gain coefficient of 10%/cm. Using a typical stimulated cross section from Table 10.9 of $\sim 4 \times 10^{-16}$ cm^2, we thus require the upper state density of 2.5×10^{14} cm^{-3}, presuming that the lower state is empty. Now a typical fluorescence lifetime is 10 ns, yielding a photon energy of $h\nu/e \sim 5$ eV; hence the spontaneous power is

$$P_{\text{spont.}} = \left(\frac{h\nu}{e}\right) \cdot \left(\frac{N_2 e}{\tau_{\text{sp}}}\right) = 20 \text{ kW/cm}^3$$

This is not an insignificant power loading, and we have not included any other losses associated with the formation of the excimer; one loses 6.54 eV via the ion channel (10.8.3). Thus it is safe to estimate that we need at least 100 kW/cm^3 to excite such a laser. The power supply electronics to provide this excitation is a tough electrical engineering problem, and removal of the waste heat is a good mechanical one. In any case, such lasers are pulse excited.

There are two primary methods: nearly relativistic electron beam excitation (hundreds of kVs, hundreds of amp/cm^2) and a discharge excitation with fast radar-type modulator circuits. Sometimes the two are combined whereby the E-beam current only sustains the discharge and the main power is supplied by the modulator circuit. It is not appropriate to discuss the electronic circuitry here, other than to say that it is a significant problem. However, the physical mechanisms for achieving the excited state are considerably different. Thus we discuss the physics of the excitation separately.

10.8.2.1 E-Beam (or nuclear) excitation. The geometry of this type of excitation is shown in Fig. 10.34 where it is presumed that an energetic, high-current electron beam is injected through the foil separating the high vacuum where the E-beam is formed and the high-pressure gas where its energy is deposited. The key parameter in such a deposition

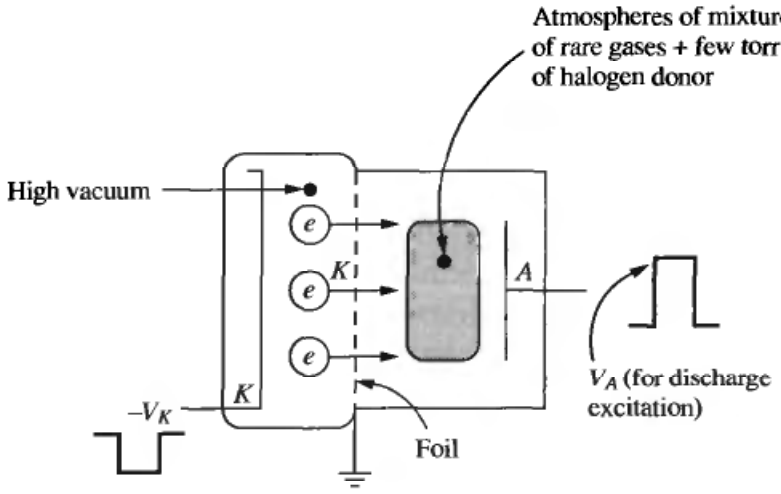


FIGURE 10.34. Excitation of the rare gas–halide laser.

is the so-called W value, which is the necessary energy deposited in a neutral gas to create 1 electron-ion pair, which in the case of argon is about 26 eV for electron energies greater than 1 keV. Thus, for instance, a 500 keV electron will generate some $5 \times 10^{+5}/26 \sim 19,000$ secondary electron-ion pairs. Inasmuch as the ionization potential of argon is 12.755 eV, not 26 eV, the energetic beam electron will also produce some excited states that decay to the metastable level. For argon, which is representative of all rare gases, only one third of an excited state is produced for each electron-ion pair. Consequently, most of the excitation from an E-beam proceeds through the ion channel.

The rare gas ions, of course, are exactly the items necessary for the formation of the excimer. The electrons are rapidly attached to a halogen by a dissociative attachment process typified by

$$e + NF_3 \longrightarrow F^- + \text{Products } (NF_2?) \tag{10.8.4}$$

Such processes occur very fast with rate coefficients k_a between 10^{-9} and 10^{-7} cm^3/ sec. Thus, if one had a background NF_3 pressure of 3 torr, then its density is ($[NF_3] = 3$ torr $\times$ 3.54×10^{16} cm^{-3}/torr $= 1.06 \times 10^{17}$ cm^{-3}), and the electrons have a lifetime for conversion into a negative ion of $(k_a[NF_3])^{-1} = 9.4$ ns, assuming $k_a = 10^{-9}$ cm^3/ sec. It will be much shorter if the higher rate coefficient is used. The negative and positive ions get together in the presence of a third body, usually another rare gas atom, to form the excimer.

$$Ar^+ + F^- + Ar \longrightarrow (Ar^+F^-)^* + Ar \tag{10.8.5}$$

The third body is necessary so that it as well as the complex can share the excess energy of the reaction, which is $\sim$ 6.54 eV as computed by (10.8.3). If the pumping is intense enough, enough excimer states will be formed to yield enough gain for the spontaneous emission to build up to a coherent amplitude so as to stimulate the complex down to the ground state (after which it dissociates in about 0.1 ps).

Now there are many competing channels for this energy, so many that we have to consult the current literature and monographs for a more complete model. However, there is one basic point to keep in mind: The energy flow to the excimer state with E-beam excitation is from above, starting at the ion level and terminating in the excimer state.

10.8.2.2 Discharge pumping. In any gas discharge, the power enters the system through the electron gas with the electric field moving the typical electron *up* the energy ladder. (This is discussed in detail in Chapter 17 for the CO_2 laser, but the physics is the same for any discharge.) When the electron energy is low, it can only make "elastic" collisions with the majority rare gas. The small amount of energy lost in such a collision, a "ping-pong ball bouncing off a battleship," merely warms the gas. The high energy tail of the electron distribution can make *inelastic collisions*, creating an excited state such as

$$e(E) + Ar \longrightarrow Ar^* + e(E - E_x) \tag{10.8.6}$$

This is the primary means by which the electron loses energy. (In a fluorescent lamp, nearly 75% of V · I goes into such a process.) The remainder of the power appears as gas heating with a small amount, typically $\leq 2\%$, going into ionization. Thus the ion channel is not significant for the formation of the excimer state.

However, the "harpooning" reaction described by (10.8.2) using the excited state of (10.8.6) can lead to the excimer.

$$Ar^* + F_2 \longrightarrow (Ar^+F^-)^* + F \tag{10.8.2}$$

Such reactions occur very fast and with nearly unity efficiency. Thus, in principle, the discharge pumping could be very efficient.

Unfortunately, nature conspires against us and generates a number of obstacles. We obviously need electrons for (10.8.6) to take place. However, the halogen attaches the electrons (10.8.4), and now the negative ion does no good whatsoever, other than causing trouble in that a detachment reaction can occur.

$$e + F^- \longrightarrow F + 2e \tag{10.8.7}$$

which leads to a negative dynamic resistance and an unstable discharge. Worse yet, (10.8.6) is only efficient at high electron temperatures, which implies high electric fields (see Chapter 17), and this is also the regimen for (10.8.7).

Thus compromises have to be made, and this reduces the promise of high efficiency dramatically. A low-current, high-energy electron beam can be used to sustain the discharge and then use the maximum electric field compatible with discharge stability. (UV preionization also helps.) If too many electrons are present, then a *superelastic* process robs the energy from the excimer state.

$$e(E) + (Ar^+F^-)^* \longrightarrow Ar + F + e(E + E_{\text{excimer}}) \tag{10.8.8}$$

In spite of the problems, the discharge-pumped laser works quite well with wall-plug efficiencies on the order of 1% to 2%. Most small commercial tabletop excimer lasers are of this variety and yield pulsed UV energy on the order of 50 to 100 mJ/pulse in a width of $\sim$ 10 ns (power = 10 MW), with a repetition rate of $\sim$ 500 pps (average power $\sim$ 50 W). These lasers are surely the most efficient source of high-intensity UV radiation.

10.9 FREE ELECTRON LASER

The free electron laser (FEL) is a device capable of generating electromagnetic radiation over extremely wide ranges of wavelengths ranging from microwave frequencies to the ultraviolet. It is debatable whether it should be called a laser, since there are not distinct "energy levels" associated with the transition, but, since the rest of the world uses this idea, we will also. The idea is a take off of some of the older classical free electron oscillators and amplifiers, with the added complication of relativistic mechanics. To give a thorough exposition, a considerable background in relativity is required, a topic beyond the scope of this book. However, with only a minimal amount of "proof-by-intimidation," a fair idea of the basic principles can be obtained.

The FEL requires that the electron undergo periodic oscillations in the transverse plane while translating at nearly the velocity of light along the z direction. The requirement of periodic oscillation implies an acceleration, and that, in turn, implies that the electron

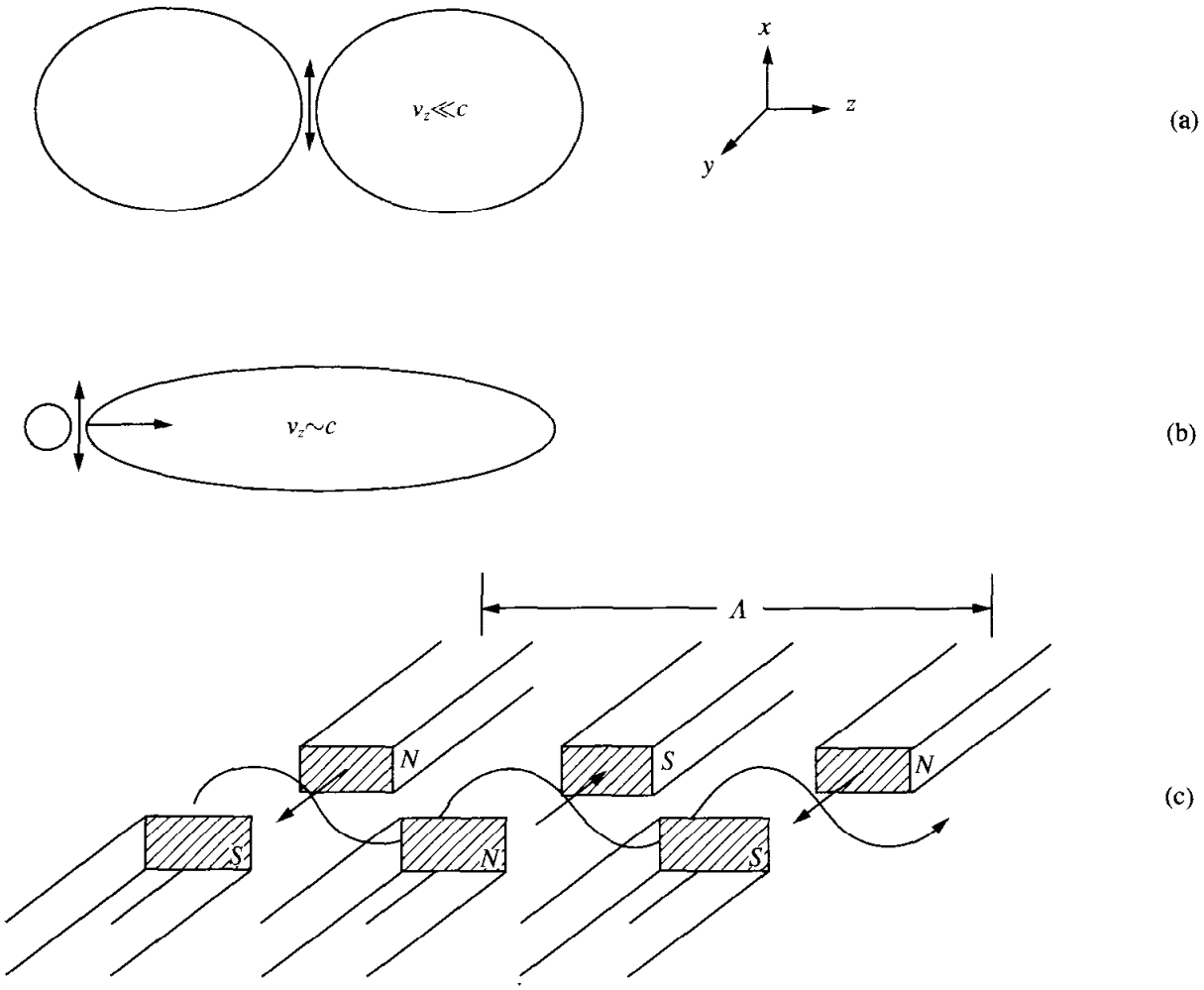


FIGURE 10.35. The radiation by an accelerated electron. (a) The pattern when the electron velocity $v_z \ll c$, (b) $v_z \sim c$, and (c) the electron trajectory in a wiggler.

radiates perpendicular to its undulatory motion (i.e., it is a dipole) as shown in Fig. 10.35(a). If the z component of the velocity is small compared to c, then the radiation pattern would be that of a simple dipole with equal amounts in the forward and reverse directions. If, however, $v_z \sim c$, then we can show that the radiation pattern is primarily in the $+z$ direction, a task that requires considerable work (which we will avoid). This is shown in Fig. 10.35(b).

The fact that the radiation pattern is becoming narrow should be attractive since this is one of the dominant characteristics of all other lasers. Actually, this characteristic of the spontaneous emission from the free electrons is not particularly useful and is most often ignored as is the spontaneous emission from atoms in conventional lasers. If many electrons were spaced along the z axis and either randomly spaced or phased, we would have to add the electric fields from each dipole and then square the total to obtain the power with the result being that the net power would be insignificant. We must ensure a phase coherence between the electrons before the situation becomes interesting. In the parlance of free electron devices, we say that the electron beam must be "bunched." Bunching of the

electron beam is a central idea to all free electron devices; fortunately, it is not as hard to accomplish as one might guess, but that comes later.

As with any laser, an increase in optical power comes at the expense of energy extracted from the pump, which is the kinetic energy, $(\gamma - 1)m_0c^2$ of the electrons in the case of the FEL.* Most are familiar with computing the energy (per unit of volume) transferred from the copropagating optical wave to the electron beam given by

$$w(\text{joules/volume}) = \int \mathbf{E}_{\text{opt}} \cdot [\mathbf{i}_b = n_{\text{beam}}(-e)\mathbf{v}_{\text{beam}}]\, dt \tag{10.9.1}$$

where $\mathbf{E}_{\text{opt}}$ is the electric field of the optical beam and $[\mathbf{i}_b = n_{\text{beam}}(-e)\mathbf{v}_{\text{beam}}]$ is the current density of the electron beam. Obviously, we wish for the energy to flow from the beam to the field, and thus the integral in (10.9.1) must be a negative number and remain so over a large number of optical cycles of copropagation.

Such a simple statement is not easily satisfied. For instance the optical field propagates in the z direction and hence it has a transverse character expressed by $\mathbf{E}_{\text{opt}} = E_0\mathbf{a}_t \cos \omega t$, whereas a first order description of the (DC) beam is $\mathbf{i}_b = i_0\mathbf{a}_z$. Hence, the integral in (10.9.1) is zero on two separate counts: $\mathbf{a}_t \cdot \mathbf{a}_z \equiv 0$ and $\int \mathbf{E} \cdot \mathbf{i}_b\, dt = 0$ for any multiple of the optical periods or multiple of the optical wavelength in the z direction. Thus some significant steps must be taken to cure this double problem.

1. The "wiggler" shown in Fig. 10.35(c) converts some of the relativistic axial velocity, $v_z = \beta_z c$, into a periodic perpendicular velocity $v_\perp$ and thus there is, at least, an outside chance for $\mathbf{E} \cdot \mathbf{v} \neq 0$.
2. The periodicity of the wiggler selects the frequency to be amplified by guaranteeing that the integral remains negative over many optical periods or a significant fraction of the copropagation path of the beam and the wave. Figure 10.36 illustrates this critical function.

In Fig. 10.36(a), the "wiggled" electron velocity is shown (with great exaggeration of the magnitude of the perpendicular velocity) but also emphasizing the fact that $v_\perp$ has a periodicity of the mechanical spacing of the magnets Λ_0. Let us follow the interaction of the electron (the solid dot) with the field (shown as 1 1/2 cycles or wavelengths) as both move along z. We presume that $\mathbf{E} \cdot \mathbf{v}_\perp$ is positive at the start of the interaction region corresponding to the electric field labelled A.

Both the field and the electron move along z, but the wave travels at a velocity of exactly c, whereas the electron is just "almost" c. Hence the electron is always going to lose the foot race and thus progressively slips behind wave train and there is not anything to be done about that fact. However, if we reverse the direction of $v_\perp$ when the slip amounts to $\lambda_0/2$ (at B), then again at the next $\lambda_0/2$ (at C), then the dot product is positive throughout the length. In this manner, the interaction of the beam and the field can accumulate over many spatial periods of the wiggler.

*An apology is offered for using the same symbol γ for the relativistic factor and for the laser gain coefficient. Only in the FEL is there a remote chance of confusing the two. The use of $\gamma = [1 - (v/c)^2]^{-1/2}$ is too ingrained in all of the physical sciences to even think of choosing a different symbol for it.

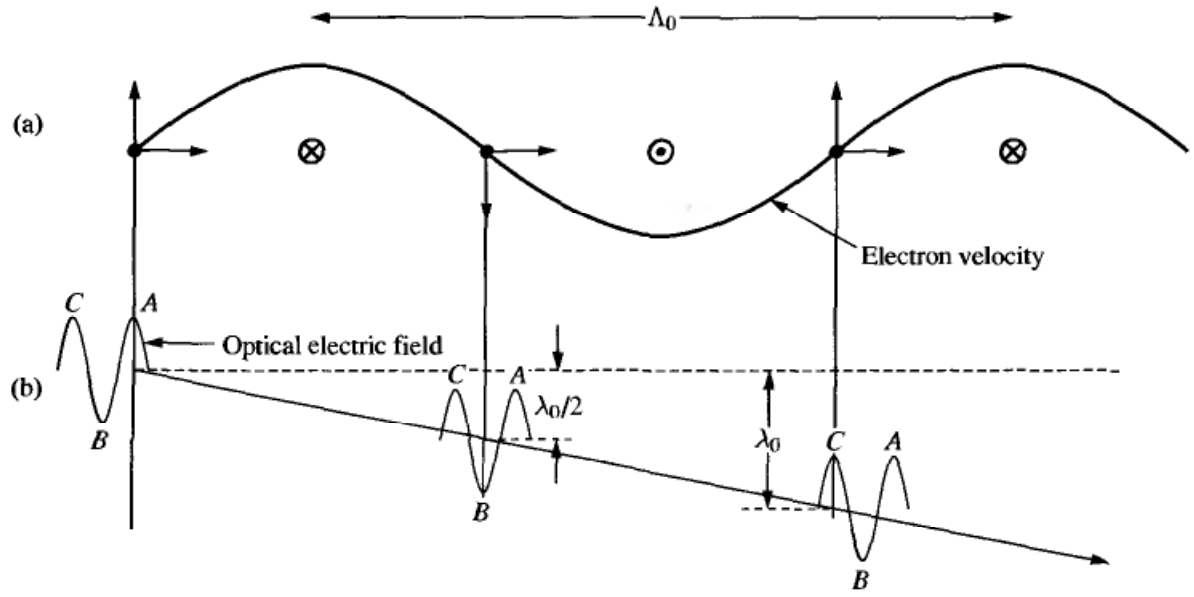


FIGURE 10.36. The accumulated interaction of the wiggled beam and the optical field. (Adaptation of Fig. 13.1 of Yario [17b].)

If we presume that the axial velocity, $v_z = \beta_z c$, is more or less a constant, then we need only require that the transverse velocity change its sign every $(\lambda_0/2) \div (c - v_z)$ seconds or every $(v_z\lambda_0/2) \div (c - v_z)$ units of length in order for $\mathbf{v} \cdot \mathbf{E}$ to have a constant value independent of z. This synchronism condition specifies the "tuning" relation between the mechanical periodicity of the magnet spacing and the free-space wavelength of the laser.

$$\Lambda_0 = \frac{\lambda_0 v_z}{c - v_z} = \frac{\lambda_0 \beta_z}{1 - \beta_z} \tag{10.9.2}$$

One can clearly see the exciting possibilities: While λ_0 might be very small, β_z is very close to 1 and hence, Λ_0, the magnet spacing could be measured with a wooden ruler! The requirement of $\beta_z \sim 1$ also points out the necessity of relativistic beams.

Consider the interaction with a 100 MeV beam: $\therefore (\gamma - 1)m_0c^2/e = 100$ MeV. Since $(m_0c^2)/e = 0.511$ MeV, we find that $\gamma = 196.7$. Now

$$\gamma^2 = \frac{1}{1 - \beta^2} \qquad \therefore \qquad \beta = 0.999987$$

If the transverse component of β were 10^{-3} of β_z, then

$$10^{-6}\beta_z^2 + \beta_z^2 = \beta^2 \qquad \text{or} \qquad \beta_z = 0.9999865$$

For $\lambda_0 = 1\mu$m in (10.9.2), the magnet spacing would be $\Lambda_0 = 7.45$ cm $= 2.93$ inches. This example points out the wide range of possibilities of the FEL. We can surely go to shorter wavelengths by either a smaller magnet spacing or with higher energy beams.

Equation (10.9.2) is correct as it stands, but the example illustrates that the value of β_z is critical. While we can safely set $\beta_z = 1$ in the numerator of (10.9.2), we need a more accurate expression for the denominator. We cannot just arbitrarily set $\beta_\perp = 10^{-3}\beta_z$ as was done for the example, but rather we must use the dynamics of the electron to solve for $v_\perp$

in terms of the specifications of the wiggler. We can make some progress by recalling the Pythagorean relationship between the components of β and the relativistic factor γ

$$v_z^2 + v_\perp^2 = v^2$$

or

$$\beta_z^2 + \beta_\perp^2 = \beta^2 = 1 - \frac{1}{\gamma^2}$$

or

$$(1 - \beta_z^2) = \beta_\perp^2 + \frac{1}{\gamma^2} = (1 - \beta_z)(1 + \beta_z) \approx 2(1 - \beta_z) = \beta_\perp^2 + \frac{1}{\gamma^2}$$

Hence, (10.9.2) becomes

$$\Lambda = \frac{2\lambda_0}{\beta_\perp^2 + 1/\gamma^2} \tag{10.9.3}$$

Solving for the optical wavelength λ_0 yields

$$\lambda_0 = \frac{\Lambda_0}{2}\left(\frac{1}{\gamma^2} + \beta_\perp^2\right) \tag{10.9.4}$$

Probably the easiest case to analyze in detail is that of a relativistic beam interacting with a circular polarized array of magnets (with the N-S flux twisted along z) rather than the linear form shown in Fig. 10.35(c). Such a static magnetic field can be expressed as

$$\mathbf{B} = B_0\left\{\cos\left(\frac{2\pi z}{\Lambda_0}\right)\hat{\mathbf{a}}_x + \sin\left(\frac{2\pi z}{\Lambda_0}\right)\hat{\mathbf{a}}_y\right\} \tag{10.9.5}$$

interacting with a circularly polarized optical electromagnetic field given by

$$\mathbf{E}_{\text{opt}} = E_0\left\{\cos(\omega t - kz - \phi)\hat{\mathbf{a}}_x + \sin(\omega t - kz - \phi)\hat{\mathbf{a}}_y\right\} \tag{10.9.6}$$

The phase angle ϕ is a parameter that identifies (or tags) the electrons as they enter the interaction region. Ultimately, we will need to average our results over all possible angles to account for the random arrival times (according to the clock kept by the electromagnetic field). The problem facing us is to solve the Lorentz force equation.*

$$\frac{d\mathbf{p}}{dt} = \frac{d[\gamma m_0 \mathbf{v}]}{dt} = (-e)[\mathbf{E}_{\text{opt}} + \mathbf{v} \times \mathbf{B}] \tag{10.9.7}$$

Various approximations will be made to simplify the mathematics, the first one being the neglect of the electric field in the transverse part of the Lorentz force equation. This is a temporary assumption used to relate $\beta_\perp^2$ to the wiggler characteristics, and its justification is based upon relative magnitudes of the transverse force terms in (10.9.7).

Compare: $E_0 \leftrightarrow \beta_z(cB_0)$ for typical values.

For $B_0 = 1.0$ Tesla (10,000 Gauss), $\beta_z \sim 1$, E_0 would have to be 3×10^8 volts/meter for the two transverse forces to be equal. This corresponds to a running wave intensity of

*Newton's law when written in a correct relativistic format.

1.2×10^{14} watts/m^2, a fairly large intensity. Hence, for a first-order theory of a low power FEL, we are justified in neglecting the electric field in the Lorentz force equation. If we restrict our attention to the *small* signal gain, then it is surely justified.

Since the magnetic field never does any work on the electron and we have just neglected the term that could change the electron energy, we can also treat γ as a constant in the force equation, and thus the x and y components become*

$$\frac{d\beta_x}{dt} = +\frac{e(cB_0)\beta_z}{\gamma m_0 c} \sin\frac{2\pi z}{\Lambda_0} \tag{10.9.8a}$$

$$\frac{d\beta_y}{dt} = -\frac{e(cB_0)\beta_z}{\gamma m_0 c} \cos\frac{2\pi z}{\Lambda_0} \tag{10.9.8b}$$

If we make the substitution, $z = \beta_z ct$ and define a mechanical angular frequency by $\omega_m = 2\pi\beta_z c/\Lambda_0$ then (10.9.8) can be integrated easily to find

$$\beta_x = \frac{-e(cB_0)\Lambda_0}{\gamma(2\pi m_0 c^2)} \cos(\omega_m t) = \frac{-a_w}{\gamma} \cos(\omega_m t) \tag{10.9.9a}$$

$$\beta_y = \frac{-e(cB_0)\Lambda_0}{\gamma(2\pi m_0 c^2} \sin(\omega_m t) = \frac{-a_w}{\gamma} \sin(\omega_m t) \tag{10.9.9b}$$

These equations provide the relationship between $\beta_\perp^2$ and the "wiggler parameter," which is abbreviated by the symbol a_w.

$$\boxed{\beta_\perp^2 = \beta_x^2 + \beta_y^2 \triangleq \frac{a_w^2}{\gamma^2}} \tag{10.9.10}$$

where

$$\boxed{a_w = \frac{e(cB_0)\Lambda_0}{2\pi m_0 c^2} = \text{wiggler parameter}} \tag{10.9.11}$$

[If we had a linear array of magnets—oriented, say, in the y direction as shown in Fig. 10.35(a)—then the electron would undulate in only the x direction and only (10.9.9a) is applicable. The average value of $\beta_\perp^2$ would be 1/2 of (10.9.10).] We can now modify (10.9.4) for a more explicit expression for the tuning of a FEL:

$$\boxed{\lambda_0 = \frac{\Lambda_0}{2\gamma^2}[1 + a_w^2]} \tag{10.9.12}$$

We now have to consider the phase of the electrons entering the interaction zone, and we quickly acknowledge that they are equally likely to arrive such at $\mathbf{E} \cdot \mathbf{v}$ is negative just as well as being positive as has been assumed in Fig. 10.36. Thus, if we assume that the axial velocity is a constant and average the Lorentz force equation over all possible entrance

*Notice that all equations are expressed in MKS (SI) units rather than the cgs versions found in much of the literature.

phases ϕ, we find, again, that the net interaction of the field with the electron beam is zero (which is a third reason why an FEL will not work).

However, the axial velocity is not a constant if we consider the interaction with the optical wave. Some of the electrons gain a bit of energy from the field as they enter the interaction zone and thus speed up while others lose energy to the field and slow slightly. The net energy cost is zero, but this sequence of events "bunches" the electrons allowing the beam to amplify the optical field. While bunching does not cost the beam anything in energy, it would cost us a significant effort in arithmetic to describe it. That we will save for more advanced texts.

The free electron laser is not a small device, but it does have some unique and attractive features.

1. We can tune the wavelength of operation by the strength of the magnetic field, the periodicity of the magnets, or through the beam energy.
2. The same technology applies from microwave frequencies to the visible through UV portion of the spectrum.
3. The technology of the production of high current very high voltage electron beams has been around for quite a few years so one major hurdle has been solved.
4. If reasonable efficiencies are obtained—say, 10%—then very large optical powers can be generated. Consider the 100 MeV beam used in the prior example. If the average current of that beam were 100 mA, then we can expect 1 MW of continuous optical power to be generated with the 10% efficiency assumption. (Many free electron tube oscillators run at much higher efficiencies. For instance 60–80% for a magnetron is not unreasonable.)

While the CO_2 laser and some chemical lasers have achieved such power levels, the FEL promises to do so over a range of wavelengths. However, the FEL is a big and very expensive device.

PROBLEMS

10.1. While the time scale for interchange of populations between the **E** and the 2**A** states is short—1 ns or so—it is not instantaneous. Discuss how this fact would influence your modeling of a Q-switched ruby laser or a system used for amplification.

10.2. Use the data provided in Table 10.3 for the following questions. Present the answers in tabular form whenever possible.

(a) Identify the upper and lower laser levels for each of the wavelengths listed.

(b) Use the fact that the spontaneous lifetime of the $^4F_{3/2}$ level is 255 μs, the given branching ratios, and the line widths of the various transitions to compute the stimulated emission cross section for the various wavelengths. Assume a Lorentzian line shape for each.

(c) Assume that the atoms in the $^4F_{3/2}$ level are in local thermodynamic equilibrium with the lattice at 300 K. What are the percentages in the two levels of that state? (Ans.: 40.1% in R_2 and 59.9% in R_1.)

(d) If there are 5×10^{17} atoms/cm^3 excited to the $^4F_{3/2}$ manifold, compute the small-signal gain coefficients for each of the transitions. Assume the concentration of the $^4I_{11/2}$ to be zero.

(e) At what lattice temperature would the dominant transition be different from that of 1.06415 μm? Suppose the crystal were cooled to 77 K (liquid nitrogen). Ignore any change in the line widths (or that all change in a synchronous fashion) and assume 5×10^{17} atoms/cm^3 were excited to the $^4F_{3/2}$ manifold. Repeat (d). Which transition has the highest gain coefficient?

(f) Find the saturation intensity for each of the transitions (at 300 K).

10.3. Use the data of Fig. 10.5 to compute the wavelengths of the absorption between the $^4I_{9/2}$ and $^4F_{3/2}$ manifolds. Note that some of these transitions match the emission of GaAs or AlGaAs semiconductor lasers.

10.4. Assume a Lorentzian line shape for all of the transitions listed in Table 10.3. Compute the effective stimulated emission cross section at 1.06415 μm including the contribution from the wings of the 1.0644 μm transition. (Assume $T = 300$K)

10.5. Use the data given in Table 10.4 to compute the fluorescent power emitted by the glass laser systems when pumped to achieve a gain of 1%/cm. Assume a doping of 1 wt%.

10.6. Use the ideas discussed in Sec. 9.6 to estimate the maximum optical energy that can be extracted from a glass rod 10 cm long and 1 cm^2 in area with a doping density of 1 wt% Nd.

10.7. Assume that the glass laser system indicated in Table 10.4 were incorporated in a cavity 40 cm long, pumped to 1.5 $\times$ threshold, and modelocked by one means or another. Without being too complicated, estimate the pulse width.

10.8. Use the data of Table 10.6 to compute the following parameters for the helium-neon laser system. Assume low-pressure gas with Doppler broadening being dominant ($T = 400$°C) and a homogeneous linewidth of 50 MHz.

(a) Stimulated emission cross sections for the following transitions: $\lambda = 3.39$ μm, $\lambda = 1.1523$ μm, and $\lambda = 0.6328$ μm. Name the transitions.

(b) What are the radiative lifetimes of the $3s_2$ and $2s_2$ states of neon?

(c) What are the saturation intensities for these transitions? (Assume that $\tau_1 \ll \tau_2$.)

(d) Assume a laser tube 50 cm long with a population difference $[N_2 - (g_2/g_1)N_1]$ equal to 10^{10} cm^{-3} for each of the transitions listed above. What is the maximum intensity that can be extracted from this laser on these three transitions? [This part requires the solution to parts (a) and (c) and the theory from Sec. 9.2.1.]

(e) The laser sequence $3s_2 \rightarrow 3p_4 \rightarrow 2s_2 \rightarrow 2p_4$ is sometimes called the push-pull laser. Why? What wavelengths are generated?

(f) The transition at 6401 Å ($3s_2 \rightarrow 2p_2$) seldom lases. Why?

10.9. The gain on the $3s_2 \rightarrow 3p_4$ transition in the helium-neon laser at 3.39 μm is very large, say, 30 dB in a 1-m-long tube.

(a) Assume a low-pressure gas mixture at 400 K and use the data to compute the population inversion necessary to provide this gain. [Ans.: $N_2 - (g_2/g_1)N_1 = 2.04 \times 10^9$ cm^{-3}.]

(b) Suppose that this same population inversion was obtained on the common "red" line at 6328 Å ($3s_2 \rightarrow 2p_4$). What would be the small-signal gain in a 1-m tube? (Ans.: 0.452 dB/m.) Why is the gain so different?

(c) Use the data given in Table 10.6 and Fig. 10.28 to compute the saturation intensity for the two laser transitions. Assume a homogeneous line width of 20 MHz and that $\tau \ll \tau_2$. [Ans.: $I_s(3.39\ \mu\text{m}) = 2.5$ mW/cm^2; $I_s(0.6328\ \mu\text{m}) = 17.1$ W/cm^2).]

10.10. By providing sufficient wavelength discrimination, we can make all of the $3s \rightarrow 2p$ transitions in neon lase. Assume a simple cavity, without prisms, gratings, or etalons, and equal inversion densities on the various transitions. What must be the finesse of the cavity at the various wavelengths compared to that at 6328 Å to ensure lasing? (Hint: Compute the gain-to-loss ratios and compare to that at 6328 Å.)

10.11. Use the NBS tables to construct an energy level diagram for the Cd^+ (ion) laser in a manner similar to that shown in Fig. 10.28. Include the helium metastable state, 2^3S and 2^1S, and pick your scale so that the upper and lower levels of the 4415 and 3250 Å Cd^+ laser are shown.

10.12. A Penning reaction, which is exemplified by

$$\text{He}(2^1\text{S or } 2^3\text{S}) + \text{Cd} \rightarrow \text{He(ground)} + (\text{Cd}^+)^* + e(\text{kinetic energy})$$

is thought to be responsible for some part of the pumping of the (Cd^+) laser. Use the data found in Problem 10.11 to compute the excess energy of this reaction (which is usually transferred to the free electron).

10.13. Compute the quantum efficiency of the argon-ion laser at 4880 Å.

10.14. Construct a graph showing the amount of energy stored in the vibrational levels ($v = 1$ to 8) of N_2 per liter-atm of gas as a function of temperature between 500 and 1500°C. (Use Table 10.7 for data on N_2.)

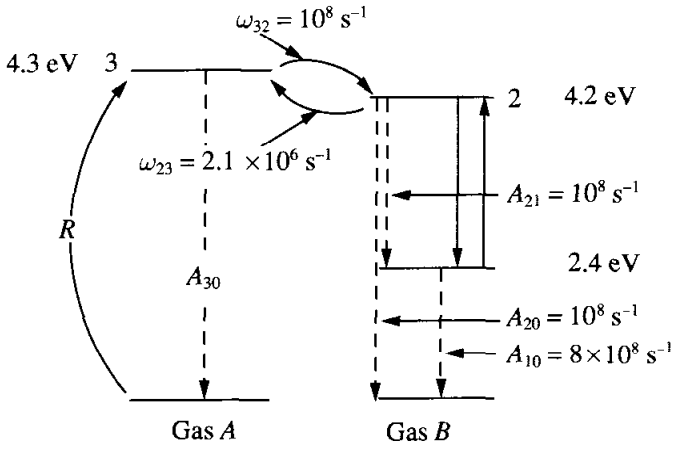
10.15. Evaluate the small-signal gain coefficient for the P(22) and R(20) transitions of the CO_2 laser using the A coefficients given in Table 10.7, the rotational constant (pick a mean value for the upper and lower states), a rotational temperature of 500 K, and a ratio of population in the 001 to 100 states of 1.1. Assume a density in the 100 state to be 10^{15} cm^{-3} and a homogeneous line width of 1 GHz.

10.16. What follows is a model for an excitation transfer laser. State 3 in gas A is excited by an external source at a rate R (cm^{-3}/sec), which can decay back to the ground state at a rate $A_{30}N_3$ or transfer its excitation to state 2 of gas B at a rate $\omega_{32}N_3$. The fate of state 2 follows the usual routes (i.e., spontaneous and stimulated decay) but can also transfer its energy back to state 3 of gas A at a rate $\omega_{23}N_2$.

(a) Show the rate equations for this laser.

(b) Find the value of R that enables this system to have a gain of 0.01 cm^{-1}. Assume a stimulated emission cross section of $\sigma = 10^{-16}$ cm^2. (Ans.: $R = 1.5 \times 10^{22}$ cm^{-3}/s^{-1}.)

(c) Assume that this laser is pumped far above threshold and its intensity is such that the populations N_2 and N_1 are equalized (for all practical purposes) by stimulated emission. What is the efficiency of this laser? (Ans.: 35%.)



10.17. The argon-ion laser is inhomogeneously (Doppler)-broadened with a width of 5.0 GHz and is contained in a cavity 100 cm long. The average power is 4 W when the small signal gain is three times the threshold value.

(a) How many $TEM_{0,0,q}$ modes are above threshold? (Assume that the central mode is at line center.)

(b) How much power is carried by the central mode? (Assume that the distribution of power among modes is Gaussian; that is, $I(\nu_q) = I(\nu_{q=1} = \nu_0) \exp\left\{-4 \ln 2[(\nu_q - \nu_0)/\Delta\nu_0]^2\right\}$.

(c) What is the peak power of the laser when it is mode-locked?

(d) What is the pulse width (FWHM) of the mode-locked output?

(e) If the spot size is 2 mm, what is the peak *density* (i.e., photons/volume) of the mode-locked pulse?

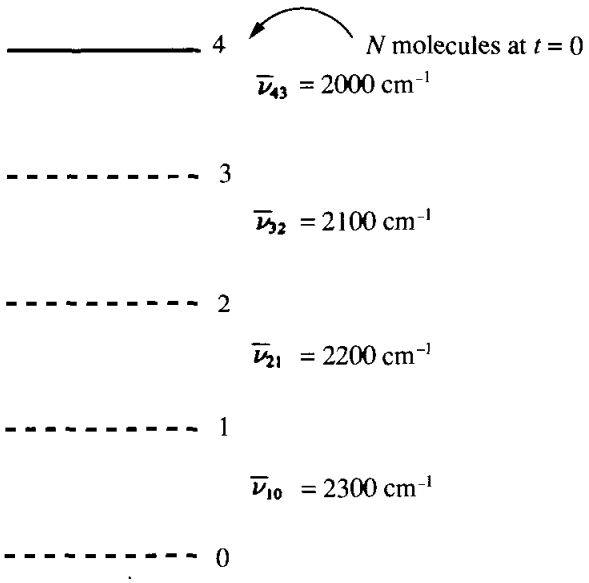
10.18. Consider the hypothetical molecular system shown, where it is assumed that the N molecules are formed in the ($\upsilon = 4$) vibrational state at $t = 0$. If we neglect all relaxation processes except stimulated emission for lowering the vibrational quantum number:

(a) What is the maximum energy that can be extracted from this system as optical power?

(b) What is the final distribution of vibrational states after stimulated emission ceases? [This problem can be approximated by a chemical laser in which the upper state is formed by a chemical reaction A + B $\rightarrow$ AB* ($\upsilon = n$). Note that there is no ground-state population initially.] To make the problem easier, neglect threshold requirements and assume that lasing occurs provided that

$N_{upper} > N_{lower}$. Also ignore the rotational structure, by computing the energy only in the vibrational bands.

(c) Evaluate the optical power assuming 1 atm of atoms in the $\upsilon = 4$ state and lasing occurring for 1 μs.



10.19. The helium-neon laser transition is Doppler broadened with a FWHM of 1.5 GHz. Assume the pumping and thus the small-signal gain coefficient are four times the threshold value, and the cavity is 100 cm long.

(a) Find the number of $TEM_{0,0}$ modes that can oscillate simultaneously.

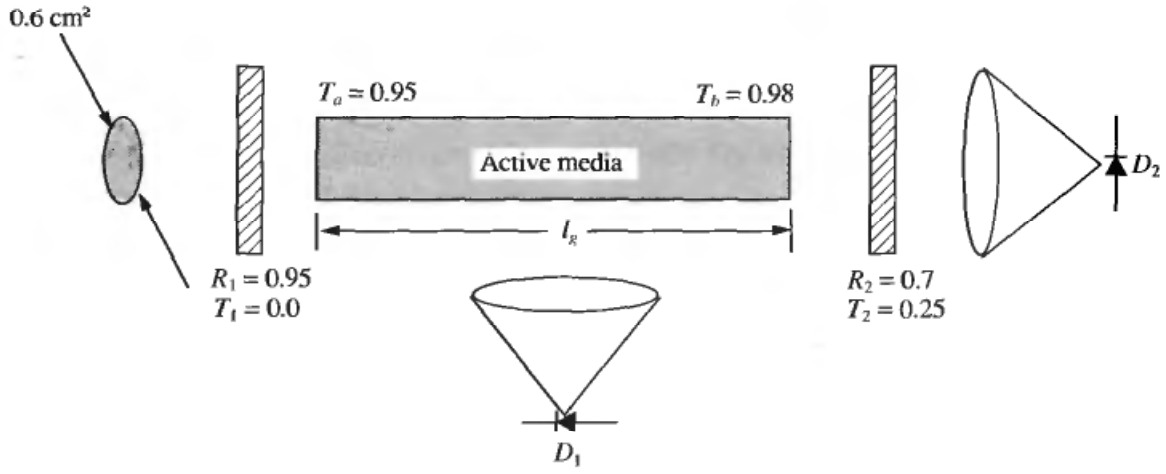
(b) Suppose all of the modes were locked together:

(1) What is the repetition frequency of the pulses?

(2) Estimate the pulse width.

10.20. The problem is to model the following "Gedanken" experiment. We have an ideal laser medium (i.e., $\tau_1 = 0$ and hence $N_1 = 0$ always) with a stimulated emission coefficient of 10^{-15} cm^2, an upper state lifetime of 50 ns, a branching ratio of 0.6, a length of 15 cm and an area of 0.6 cm^2, and the wavelength is $\lambda = 5000$ Å. It is placed in the cavity shown below and the excitation rate of state 2, R_2, is varied from zero to three times the threshold for oscillation. There are two detectors shown: One is used to measure the spontaneous emission from the upper state and emerging from the side of the cavity, and the other is to measure the laser output along the axis of the cavity. Assume that the output of each detector is a current that is proportional to power.

Construct a graph showing the variation of both signals as a function of a normalized pumping rate $R_2/R_{threshold}$. Plot the *total* spontaneous and laser power in watts from this system. (The side-light detector would only capture a fraction of the total spontaneous amount, but that fraction would not be a variable.)



10.21. Suppose the short-range nuclear potential repelling the krypton and fluorine atoms can be expressed as

$$V(r) = V_0 \left(\frac{a}{r}\right)^6 \text{ with } V_0 = 54.7 \text{ eV and } a = 1.5 \text{ Å}$$

The attractive force is due to the coulomb forces. Use the above and the IP of krypton in Table 10.8 to estimate the wavelength of the $(KrF)^*$ excimer radiation. The following steps may be helpful:

(a) Find the "equilibrium radius" r_{min} of the upper state. (It will be close to but not equal to the 2.3 Å specified in Table 10.9.)

(b) Find the total energy of the $(Kr^+F^-)^*$ state at r_{min}.

(c) Find the total energy of the Kr+F at r_{min} by evaluating $V(r)$ at r_{min}.

(d) Subtract the answer of (c) from that of (b). This will be the energy of the transition and will be close but not equal to that of Table 10.9.

(e) It helps to draw the interaction potentials and to label r_{min} on that sketch.

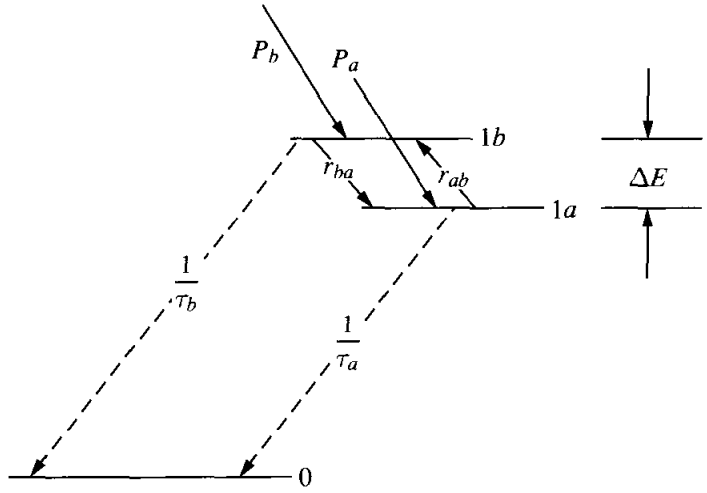
10.22. What follows are in reference to an FEL with a beam energy of 75 MeV.

(a) Evaluate the relativistic factor $\gamma = m/m_0 = [1 - (v/c)^2]^{-1/2}$. (Ans.: $\gamma = 146.8$.)

(b) Evaluate the quantity $\beta = v/c$ for this electron. (Ans.: $\beta = 0.999998$.)

(c) If the periodicity of the magnets in Fig. 10.35 were 3 cm and $B_0 = 1T$ what would be the wiggler parameter and the wavelength of the laser? (Ans.: $a_w = 2.8$ and $\lambda = 6.17\ \mu$m.)

10.23. Consider the system shown below depicting two states ($1a$, $1b$), which are closely coupled by very fast thermal processes shown on the diagram and which are related by $r_{ab} = r_{ba} \exp[-\Delta E/kT]$, where $r_{ba} = 10^{12}\ \text{sec}^{-1}$. This coupling rate is much faster than the natural decay rate of each state: $\tau_a^{-1} = 10^4 s^{-1}$ and $\tau_b^{-1} = 10^3 s^{-1}$, and thus the sum of the populations decays with a time constant that is different from either. Assume that τ_a, τ_b are independent of temperature.



(a) If the Boltzmann factor $\exp[-\Delta E/kT] = 1/9$ at room temperature ($T = 300$ K), what is the decay rate of the *manifold* ($1a + 1b$)?

(b) If the temperature were increased to 600 K, what is the decay rate of the populations in the manifold?

10.24. The following questions refer to the alexandrite laser (see Fig. 10.25) and to References 87 and 89.

(a) Let us focus on the dynamics of the ^2E doublet and assume that $\tau_b/\tau_a = 2.93$. Derive a formula that predicts the temperature dependence of the effective lifetime of the manifold ($N_a + N_b$) in terms of the lifetime τ_a, assuming a Boltzmann relationship between the two states is always maintained and that τ_a, τ_b are independent of T. Show that the lifetime increases with temperature between 4 and 100 K where the coupling to the $^4T_1$ and $^4T_2$ levels can be ignored.

(b) Figure 5 of Ref. [87] indicates a combined lifetime of ~1.58 ms at 100 K where the populations of $^4T_2$ and $^2T_1$ manifolds are negligible. Use that fact to evaluate the lifetimes τ_a, τ_b.

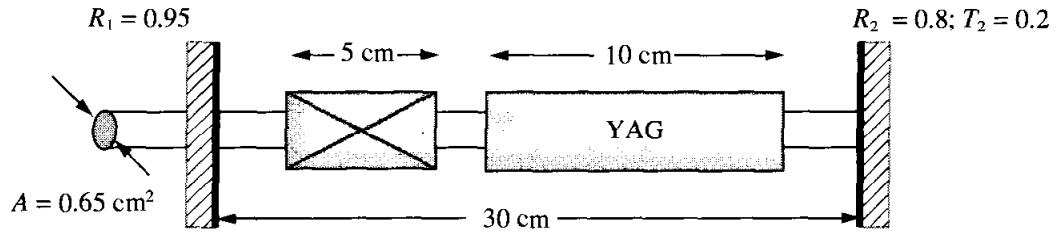
(c) Suppose optical pumping promotes N^* atoms/cm^3 into excited states to be shared by the ^2E, $^4T_2$, and $^2T_1$ manifolds. What fraction is in each state? Assume $T = 300$ K. Note that the answers are given in Table 3 of Ref. [87] but you must show the derivation.

(d) Assume $T = 300$ K, a lifetime of 10 μs for the $^4T_2$ manifold, and $\tau(^2T_1) \rightarrow$ very long. Use the values for τ_a and τ_b as computed in (a). What is a formula for the effective lifetime of the entire excited state population? Evaluate and compare in a graphical format to the data presented in Fig. 5 of [87].

(e) Assume a [Cr] doping density of 2.2×10^{19} cm^{-3}, $T = 300$ K, and evaluate the total density of excited states to obtain (1) optical transparency on the R_1 line and (2) optical transparency on the $2 \rightarrow 1$ transition assuming equal degeneracies between 2 and 1.

10.25. The YAG laser shown below is pumped to three times threshold before the electro-optic shutter is opened for Q switching. Assume (1) the characteristics of the YAG

rod are those specified in Tables 10.2 and 10.3 but ignore the scattering loss, (2) the transmission at all air-device interfaces is 0.99, (3) the index of refraction of the electro-optic switch is 2.3, (4) lasing at 1.0615 μm (where there is no significant overlap with any other transition), (5) the populations in the $^4F_{3/2}$ states are always distributed according to the Boltzmann relation with $kT = 208\ \text{cm}^{-1}$, even within the Q-switched pulse.



(a) Evaluate the following parameters to be used in the calculation: photon lifetime of passive cavity in ns, A_{21} coefficient for the transition in sec^{-1}, stimulated emission cross section in cm^2, initial density of atoms in $^4F_{3/2}$ manifold in cm^{-3}, and final density of atoms in $^4F_{3/2}$ manifold in cm^{-3}.

(b) Compute the peak power in watts, the energy in joules, and determine the pulse width using the theory of Sec. 9.4.

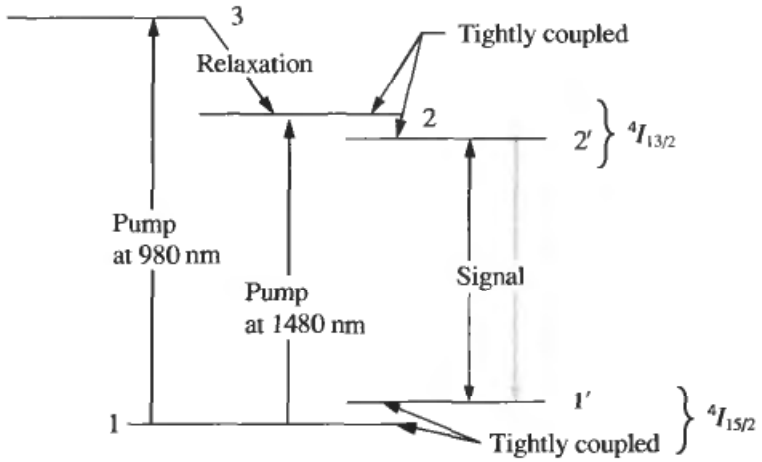
(c) The theory of Sec. 9.4 is not quite applicable to this problem since the lifetime of the lower state is only 30 ns and the time scale for the establishment of a Boltzmann factor among the levels of the $^4I_{11/2}$ manifold is even shorter. Redo the calculations of (a) assuming that $N_1 = 0$, which is a bit different from the analysis of Sec. 9.4.

10.26. The erbium-doped fiber amplifiers present a new set of problems that have not been encountered in other systems, not the least of which are the physical extremes encountered. The amplifiers are long (meters to hundreds of meters), optically pumped from one end or another or both (at $\lambda \sim 980$ or 1480 nm), the active atoms are embedded in the center of a small ($\sim$ 10–50 μm) diameter glass (SiO_2) fiber, and thus these atoms are excited by a spatially nonuniform pump mode and then amplify a characteristic mode at the signal frequency. Obviously, various approximations have become popular so as to identify some of the major features without being stymied by all of the interacting issues. The following problem was inspired by parts of the paper by K. Nakagawa et al. [54], and is intended to illustrate some of the unique problems encountered.

(a) To consider dynamics of all of the Stark levels of the $^4I_{15/2}$, $^4I_{13/2}$ and $^4I_{11/2}$ manifolds of Fig. 10.11 would require a major effort. Probably the simplest approximate model assumes just two levels (1, 2) representing the $^4I_{15/2}$ and $^4I_{13/2}$ manifolds respectively and presumes that $^4I_{11/2} = 3$ decays so fast to $^4I_{13/2}$ that the density of state 3 is zero. Hence $N_1 + N_2 = N$, the density of active atoms. To accommodate the manifold character of (1, 2), one assigns: σ_p = absorption cross section of the pump at ν_p; σ_s = stimulated absorption

emission cross-section at ν_s; τ = spontaneous lifetime of state 2; and neglects stimulated emission by the pump from 2 to 1.

(1) Examine this model in terms of the energy level diagram shown below and identify circumstances which justify any assignments and the approximations suggested.



(2) Use this model (with all of its faults) to show that the signal and the pump propagate according to:

$$\frac{\partial S}{\partial z} = \frac{P - 1}{1 + 2S + P} \cdot \alpha_s S \tag{1}$$

$$\frac{dP}{dz} = \mp \frac{S + 1}{1 + 2S + P} \cdot \alpha_p P \tag{2}$$

where

$$S = I_s/I_{ss}; \quad I_s = \text{signal intensity (W/cm}^2); \quad I_{ss} = [h\nu_s/\sigma_s\tau]$$

$$P = I_p/I_{sp}; \quad I_p = \text{pump intensity (W/cm}^2); \quad I_{sp} = [h\nu_p/\sigma_p\tau]$$

and the sign choice is dependent upon whether the pump is propagating along z (copropagation for $-$ sign) or along $-z$ (use $+$ sign in 2). I_{ss} and I_{sp} are characteristic saturation intensities.

(b) Assume copropagation along z. Divide (1) by (2); separate the variables, integrate between the input (1) and output (2) of the amplifier, and solve for the saturated gain $G = S_2/S_1$ in terms of the corresponding values of the pump.

$$\ln G + S_1(G - 1) = \frac{\alpha_s}{\alpha_p}\left[(P_1 - P_2) - \ln \frac{P_1}{P_2}\right] \tag{3}$$

or

$$\ln G + S_1(G - 1) = \frac{\alpha_s}{\alpha_p}\left[P_1(1 - T) - \ln \frac{1}{T}\right] \tag{4}$$

where T is the transmission of the pump and G is the gain with saturation of both being a possibility.

(c) Evaluate Eq. 3 for a "small" signal $S_{1,2} \ll 1$ (but still a significant pump), and assume this condition for the remainder of the problem.

$$\ln G_0 = \frac{\alpha_s}{\alpha_p}\left[(P_1 - P_2) - \ln \frac{P_1}{P_2}\right] \tag{5}$$

(d) Check your calculations by choosing $P_1 \to 0$ and insuring that parts (b) and (c) yield consistent and logical answers.

(e) As the pumping power increases from zero, the absorption at the signal frequency decreases until Optical Transparency ($G_0 = 1$) is reached and life becomes interesting. Show that this minimum (normalized) pump intensity P_{10} is given by:

$$P_{10} - P_{20} = \ln \frac{P_{10}}{P_{20}} \qquad \text{for O.T. or } G_0 = 1 \tag{6}$$

(f) Eq. 6 has 2 unknowns P_{10} and P_{20}, the normalized pumping powers at optical transparency. A second equation can be obtained by integrating Eq. (2) under the assumption of small signal intensity ($S_{1,2} \ll 1$). Show that:

$$P_1 - P_2 + \ln \frac{P_1}{P_2} = \alpha_p L \tag{7}$$

which applies under all cases for which $S_{1,2} \ll 1$ and not just for OT.

(g) Eq. 7 can be applied for the condition of optical transparency. Show that:

$$P_{10} - P_{20} = \frac{\alpha_p L}{2} = \ln \frac{P_{10}}{P_{20}} \tag{8}$$

(h) Combine the above so as to show:

$$\ln G_0 = 2\frac{\alpha_s}{\alpha_p}\left[(P_1 - P_2) - \frac{\alpha_p L}{2}\right] = 2\frac{\alpha_s}{\alpha_p}[(P_1 - P_2) - (P_{10} - P_{20})] \tag{9}$$

(i) **(1)** Sketch the variation of G_0 with input pump power above that required for optical transparency assuming a fixed $\alpha_p L$.

(2) Sketch the variation of G_0 with the physical length L (thus $\alpha_p L$) for a fixed pump power. (Compare your sketches to Figs. 7 and 8 of cited paper.)

For the numerical answers requested below, assume pumping at $\lambda_p = 1.48\mu$m; with $\sigma_p = 3 \times 10^{-21}$cm^2; $\tau = 10.2$ms; $\lambda_s = 1.55\mu$m; $\sigma_s = 5.0 \times 10^{-21}$cm^2; [Er] $= 5.6 \times 10^{17}$cm^{-3}. Assume a Gaussian mode with $w_0 = 5\ \mu$m.

(j) If the pumping were 10× that required for optical transparency, then maximum small signal gain is obtained when $\alpha_p L \approx 5.5$ (see Fig. 8 of cited paper). What is the length of this amplifier? (Ans.: 32.7 meters)

(k) Evaluate incident pumping power which brings the amplifier to optical transparency. How much power is absorbed? (Ans.: $P_{10} = 2.9378$; $I_{sp} = 4.38\ \text{kW/cm}^2$; $p_{10} = 5.06\text{mW}$; $p_{abs} = p_{10}(1 - T_{p0}) = 4.73\text{mW}$)

(l) Now let the incident pump power be 50.6 mW and thus there is even more saturation by the pump and thus a greater fraction is transmitted. How much of incident power is absorbed? (Ans.: $T_p = 0.82$; $P_{abs} = 9.12\text{mW}$)

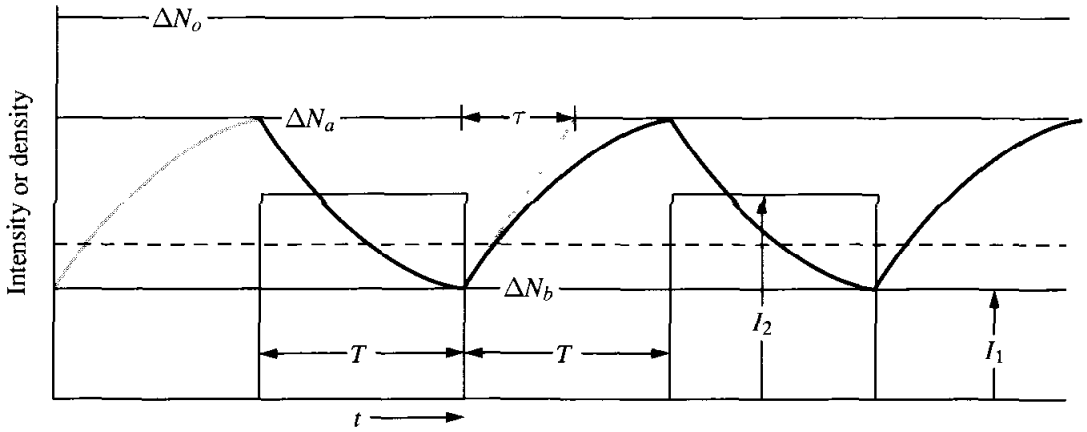
(m) What is the small signal gain (as a numeric and in dB)? (Ans.: $G_0 = 4.93 \times 10^3 \Rightarrow 36.9\text{dB}$)

(n) If the signal I_s (or amplified spontaneous emission) becomes significant, one can no longer use the above calculations. Present an argument showing that larger signal leads to more absorption of the pump which tends to maintain the gain. (NOTE: The Feb. 1991 issue of *J. Lightwave Technology* has many articles with much more sophisticated models. It is highly recommended.)

10.27. The purpose of this exercise is to illustrate the problem of "cross-talk" between two (or more) wavelengths using a common optical amplifier—say a EDFA or a semiconductor amplifier. Consider the sketch shown below in which the "signal" (it may be just amplified spontaneous emission) in channel 1 is a constant in time whereas that in 2 has a square-wave envelope. Let's assume a simple model of the amplifier whose population inversion obeys:

$$\frac{d\Delta N}{dt} = \frac{\Delta N_0 - \Delta N}{\tau} - \frac{\sigma[I_1(t) + I_2(t)}{h\nu}\Delta N \tag{1}$$

where ΔN_0 = inversion for $I_1 = I_2 = 0$; τ = time constant for the recovery of the gain coefficient; $(2T)^{-1}$ = bit rate; and the wavelengths of the two signals are close enough that we can use a common stimulated emission cross-section and optical frequency for the rate equation.



Shown also in the sketch is the temporal behavior of the inversion (and thus the gain coefficient) fluctuating between ΔN_a and ΔN_b as predicted by (1), which, in turn, causes a temporal change in the amplified value of I_1, which will now contain the same digital information as that on channel 2; i.e., there is cross talk.

(This negates the assumption of $I_1 \neq f(t)$, but accounting for that fact makes the problem much more involved.)

(a) Use Eq. 1 to evaluate the saturated inversion values (ΔN_a, ΔN_b) in terms of ΔN_0, τ and $I_{1,2}/I_s$. (Hint: Use the periodic boundary conditions and abbreviate by defining $\epsilon_1 = \exp[-(1 + I_1/I_s)T/\tau]$; $\epsilon_2 = \exp[-(I_2/I_s)T/\tau]$; $a = I_1/I_s$; $b = I_2/I_s$).

(b) Solve for $\Delta N_a - \Delta N_b$ which will be proportional to the change in the gain. Assume $I_1/I_s = I_2/I_s = 1$ and evaluate under two different circumstances:

(1) Let $\tau \gg T$, so that $\exp[-mx] \simeq 1 - mx$, corresponding to our EDFA.

(2) Let $\tau \leq T$, so that $\exp[-T/\tau] \ll 1$, corresponding to a semiconductor amplifier.

(c) Which case minimizes the cross talk?

REFERENCES AND SUGGESTED READINGS

1. T.A. Hänsch, M. Pernier, and A.L. Schalow, "Laser Action of Dyes in Gelatin," *IEEE J. Quant. Electron. QE – 7*, 47, Jan. 1971.
2. (a) T.H. Maiman, "Stimulated Optical Radiation in Ruby Masers," *Nature 187*, 493, 1960.
   (b) T.H. Maiman, "Optical and Microwave-Optical Experiments in Ruby," *Phys. Rev. Lett. 4*, 564, 1960.
   (c) T. Maiman, "Stimulated Optical Emission in Fluorescent Solids, I, Theoretical Considerations," *Phys. Rev. 123*, 1145–1150, Aug. 15, 1961.
   (d) T H. Maiman, R.H. Hoskins, I.J. D'Haenens, C.K. Asawa, and V. Evtuhov, "Stimulated Optical Emission in Fluorescent Solids, II, Spectroscopy and Stimulated Emission in Ruby," *Phys. Rev. 123*, 1151–1157, Aug. 15, 1967.
3. D.C. Cronemeyer, "Optical Absorption Characteristics of Pink Ruby," *J. Opt. Soc. Am. 56*, 1703, 1966.
4. W. Koechner, L.C. DeBenedictis, E. Matovich, and G.E. Meyer, "Characteristics and Performance of High Power CW Krypton Arc Lamps for Nd:YAG Laser Pumping," *IEEE J. Quant. Electron. QE-8*, 310, 1972.
5. E. Snitzer and C.G. Young, "Glass Lasers," in *Lasers*, Vol. 2, Ed. A.K. Levine (New York: Marcel Dekker, 1968), p. 199.
6. H.G. Danielmey, "Progress in YAG Lasers," in *Lasers*, Vol. 4, Eds. A.K. Levine and A.J. DeMaria (New York: Marcel Dekker, 1976), p. 8.
7. T. Kushida, 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, 1969.
8. B.B. Snavely, "Flash Lamp Pumped Dye Lasers," *Proc. IEEE 57*, 1374, 1969.
9. M. Bass, T.F. Deutsch, and M.J. Weber, "Dye Lasers," in *Lasers*, Vol. 3, Eds. A.K. Levine and A. DeMaria (New York: Marcel Dekker, 1971), p. 275.
10. A. Javan, W.R. Bennett, Jr., and D.R. Herriott, "Population Inversion and Continuous Optical Maser Oscillation in a Gas Discharge Containing He-Ne Mixtures," *Phys. Rev. 6*, 106, 1961.
11. W.R. Bennett, Jr., "Inversion Mechanisms in Gas Lasers," *Appl. Opt., Suppl. 2 Chem. Laser*, 3–33, 1965.
12. E.F. Labuda and E.I. Gordon, *J. Appl. Phys. 35*, 1647, 1964.

13. P.K. Cheo, "CO_2 Lasers," in *Lasers*, Vol. 3, Eds. A.K. Levine and A. DeMaria (New York: Marcel Dekker, 1971).

14. The even-odd sequence is due to the symmetry of the CO_2 molecule (see G. Herzberg, *Infrared and Raman Spectra* (Princeton, N.J.: D. Van Nostrand Company, 1966), Chap. 1.

15. G.J. Schultz, "Vibrational Excitation of N_2, CO, and H_2 by Electron Impact," *Phys. Rev. 135*, 988–994, 1964.

16. (a) E.L. Patterson and R.A. Gerber, "Characteristics of a High Energy Hydrogen Fluoride (HF) Laser Initiated by an Intense Electron Beam," *IEEE J. Quant. Electron. QE-11*, 642–647, Aug. 1975.
    (b) G.N. Hays, J.M. Hoffman, and G.C. Tisone, "Phoenix 2: Sandia's 1 kJ HF Laser System," Paper WE 1 Topical Meeting on Inertial Confinement Fusion, San Diego, Calif., Feb. 26–28, 1980.

17. (a) A. Yariv, *Introduction to Optical Electronics*, 2nd ed. (New York: Holt, Rinehart and Winston, 1971), Chap. 7.
    (b) A. Yariv, *Quantum Electronics*, 2nd ed. (New York: John Wiley & Sons, 1975), Chap. 10.

18. B.G. Streetman, *Solid State Electronic Devices*, 2nd ed. (Englewood Cliffs, N.J.: Prentice Hall, 1980), Chap. 10.

19. W.R. Bennett, Jr., "Gaseous Optical Masers," *Appl. Opt., Suppl. Opt. Masers*, 24–64, 1962.

20. C.K.N. Patel, "High Power Carbon Dioxide Laser," *Sc. Am. 219*, 22–23, 1968.

21. D.R. Herriott, "Applications of Laser Light," *Sci. Am. 219*, 141–156, 1968.

22. A.B. Siegman, *An Introduction to Lasers and Masers* (New York: McGraw-Hill, 1971).

23. A.E. Siegman, *Lasers* (Mill Valley, Calif.: University Science Books, 1986).

24. W. Koechner, *Solid State Engineering* (New York: Springer-Verlag, 1976).

25. A.A. Kaminski, *Laser Crystals*, in *Springer Series in Optical Sciences* (New York: Springer-Verlag, 1981).

26. H. Brown, *High Power Glass Lasers* (New York: Springer-Verlag, 1976).

27. See "Third Special Issue on Free-electron Lasers," *IEEE J. Quant. Electron. QE-21*, No. 7, 1985. This contains 40 papers on free-electron lasers.

28. Thomas C. Marshall, *Free Electron Lasers* (New York: Macmillan 1985). This has a comprehensive bibliography of 206 references and a very readable introduction. See also *IEEE J. Quant. Electron. QE-23*, No. 9, 1987, for additional 27 articles.

29. (a) J.E. Velazco and D.W. Setser, "Quenching Studies of Xe ($^3P_2$) Metastable Atoms," *IEEE J. Quant. Electron. QE-11*, 708–709, Aug. 1975.
    (b) L.G. Piper, J.E. Velazco, and D.W. Setser, *J. Chem. Phys. 59*, 3323, 1973.
    (c) J.E. Velazco and D.W. Setser, *Chem. Phys. Lett. 25*, 197, 1974.

30. M. Rokni, J.A. Mangano, J.H. Jacob, and J.C. Hsia, "Rare Gas Fluoride Lasers," IEEE *J. Quant. Electron. QE-14*, 464–481, July 1978.

31. C.A. Brau, "Rare Gas Halogen Excimers," in *Excimer Lasers*, Ed. C.K. Rhodes, Topics in Applied Physics, Vol. 30 (New York: Springer Verlag, 1979.)

32. John. M. Madey, "Stimulated Emission of Bremsstrahlung in a Periodic Magnetic Field," *J. Appl. Phys. 42*, 1906, 1971.

33. H. Motz, "Application of the Radiation from Fast Electron Beams," *J. Appl. Phys. 22*, 527, 1951.

34. J.A. Pasour, "Free Electron Lasers," *IEEE Circuits and Devices Magazine, 55–63*, 1987.

35. J. Madey, "Stimulated Emission of Radiation in Periodically Deflected Electron Beam," U.S. Patent 3,822,410, July 2, 1974.

36. J.J. Ewing, "Excimer Lasers," in *The Laser Handbook*, Vol. 3, Article A4, Ed. M. L. Stitch, (New York: North Holland Publishing Company, 1979).

37. C.J. Ultee, "Chemical and Gas Dynamic Lasers," in *The Laser Handbook*, Vol. 3, Article A5, Ed. M. L. Stitch, (New York: North Holland Publishing Company, 1979).

38. T.Y. Fan and R.L. Byer, "Diode Laser-Pumped Solid-State Lasers," *IEEE J. Quant. Electron. 24*, 895–912, 1988.

39. A.A. Kaminski, *Laser Crystals*, Vol. 14 in Springer Series in Optical Sciences (New York: Springer-Verlag, 1981).

40. E. Desurvire and J.R. Simpson, "Evaluation of $^4I_{15/2}$ and $^4I_{13/2}$ Stark-Level Energies in Erbium-Doped Aluminosilicate Glass Fibers," *Opt. Lett. 15*, 547–549, 1990.

41. W.L. Barnes, R.I. Laming, E.J. Tarbox, and P.R. Morkel, "Absorption and Emission Cross Section of Er^{3+} Doped Silica Fibers," *IEEE J. Quant. Electron. QE-27*, 1004–1009, 1991.

42. E. Desurvire, "Study of the Complex Atomic Susceptibility of Erbium-Doped Fiber Amplifiers," *J. Lightwave Technol. 8*, 1517–1527, 1990.

43. E. Desurvire, "Analysis of Distributed Erbium-Doped Fiber Amplifiers With Fiber Background Loss," *IEEE Photon. Technol. Lett. 3*, 625–628, 1991.

44. E. Desurvire, "Design Optimization for Efficient Erbium-Doped Fiber Amplifiers," *J. Lightwave Technol. 8*, 1730–1741, 1990.

45. A. Lidgard, D.J. DiGiovanni, and P.C. Becker, "A Comparative Study of an Erbium Doped Fiber Amplifier Pumped at 811 and 980 nm," *J. Quant. Electron. QE-28*, 43–47, 1992.

46. P.C. Becker, "Erbium-Doped Fiber Makes Promising Amplifiers," *Laser Focus World*, 197–203, 1990.

47. B. Pedersen, A. Bjarkslev, J. Hedegaard Povlsen, K. Dybdal, and C.C. Larsen, "The Design of Erbium-Doped Fiber Amplifiers," *J. Lightwave Technol. 9*, 1105–1112, 1991.

48. S.G. Grubb, W.F. Humer, R.S. Cannon, T.H. Windhorn, S.W. Vendetta, K.L. Sweeney, P.A. Leilabady, W.L. Barnes, K.P. Jedrzejewski, and J.E. Townsend, "+21 dBm Erbium Power Amplifier Pumped by a Diode-Pumped Nd:YAG Laser," *IEEE Photon. Technol. Lett. 4*, 553–555, 1992.

49. B. Pedersen, J. Chirravuri, and W.J. Miniscalco, "Gain and Noise Properties of Small-Signal Erbium-Doped Fiber Amplifiers Pumped in the 980-nm Band," *IEEE Photon. Technol. Lett. 4*, 556–558, 1992.

50. E. Desurvire, "Analysis of Erbium-Doped Fiber Amplifiers Pumped in the $^4I_{15/2} - {}^4I_{13/2}$ Band," *IEEE Photon. Technol. Lett. 1*, 293–296, 1989.

51. C.R. Giles and E. Desurvire, "Propagation of Signal and Noise in Concatenated Erbium-Doped Fiber Optical Amplifiers," *J. Lightwave Technol. 9*, 147–154, 1991.

52. W.J. Miniscalco, "Erbium-Doped Glasses for Fiber Amplifiers at 1500 nm," *J. Lightwave Technol. 9*, 234–250, 1991.

53. C.R. Giles and E. Desurvire, "Modeling Erbium-Doped Fiber Amplifiers," *J. Lightwave Technol. 9*, 271–283, 1991.

54. K. Nakagawa, S. Nishi, K. Aida, and E. Yoneda, "Trunk and Distribution Network Applications of Erbium-Doped Fibers for Optical Amplifiers," *J. Lightwave Technol. 9*, 198–207, 1991.

55. D.E. McCumber, "Theory of Phonon-Terminated Optical Masers," *Phys. Rev. 134*, A299–A307, 1964.

56. D.E. McCumber, "Einstein Relations Connecting Broadband Emission and Absorption Spectra," *Phys. Rev. 136*, A954–A957, 1964.
57. W. Beall Fowler and D.L. Dexter, "Relation Between Absorption and Emission Probabilities in Luminescent Centers in Ionic Solids, *Phys. Rev. 128*, 2154–2165, 1962.
58. L.F. Johnson, H.J. Guggenheim, and R.A. Thomas, "Phonon-Terminated Optical Masers," *Phys. Rev. 149*, 179–185, 1966.
59. L.F. Johnson, R.E. Dietz, and H.J. Guggenheim, "Optical Maser Oscillation from Ni^{2+} in MgF_2 Involving Simultaneous Emission of Phonons," *Phys. Rev. Lett. 11*, 318–320, 1963.
60. S.A. Payne, L.L. Chase, L.K. Smith, W. Kwang, and W.F. Krupke, "Infrared Cross-Section Measurements for Crystals Doped with Er^{3+}, Tm^{3+}, and Ho^{3+}," *IEEE J. Quant. Electron. QE-28*, 2619–2630, 1992.
61. As advertised in the Exciton, Inc., Catalog, page 4, 1992.
62. R.C. Powell, J.L. Caslavsky, Z. AlShaieb, and J.M. Bowen, "Growth, Characterization, and Optical Spectroscopy of Al_2O_3:Ti^{3+}," *J. Appl. Phys. 58*, 2331–2336, 1985.
63. C.H. Muller III, D.D. Lowenthal, K.W. Kangas, R.A. Hamil, and G.C. Tisone, "2.0-J Ti:Sapphire Laser Oscillator," *Opt. Lett. 13*, 380–382, 1988.
64. A. Sanchez, A.J. Strauss, R.L. Aggarwal, and R.E. Fahey, "Crystal Growth, Spectroscopy, and Laser Characteristics of Ti:Al_2O_3," *IEEE J. Quant. Electron. QE-24*, 995–1002, 1988.
65. R.L. Aggarwal, A. Sanchez, M.M. Stuppi, R.E. Fahey, A.J. Strauss, W.R. Rapoport, and C.P. Khattak, "Residual Infrared Absorption in As-Grown and Annealed Crystals of Ti:Al_2O_3," *IEEE J. Quant. Electron. QE-24*, 1003–1008, 1988.
66. J.M. Eggleston, L.G. DeShazer, and K.W. Kangas, "Characteristics and Kinetics of Laser-Pumped Ti:Sapphire Oscillators, *IEEE J. Quant. Electron. QE-24*, 1009–1015, 1988.
67. K.F. Wall, R.L. Aggarwal, R.E. Fahey, and A.J. Strauss, "Small-Signal Gain Measurements in a Ti:Al_2O_3 Amplifier," *IEEE J. Quant. Electron. QE-24*, 1016–1020, 1988.
68. N.P. Barnes, J.A. Williams, J.C. Barnes, and G.E. Lockard, "A Self-Injection Locked, Q-Switched, Line-Narrowed Ti:Al_2O_3 Laser," *IEEE J. Quant. Electron. QE-24*, 1021–1028, 1988.
69. J.C. Barnes, N.P. Barnes, and G.E. Miller, "Master Oscillator Power Amplifier Performance of Ti:Al_2O_3," *IEEE J. Quant. Electron. QE-24*. 1029–1038, 1988.
70. P.A. Schulz, "Single-Frequency Ti:Al_2O_3 Ring Laser," *IEEE J. Quant. Electron. QE-24*, 1039–1048, 1988.
71. R. Moncorgé, G. Boulon, D. Vivien, A.M. Lejus, R. Collongues, V. Djévahirdjian, K. Djévahirdjian, and R. Cagnard, "Optical Properties and Tunable Laser Action of Verneuil-Grown Single Crystals of Al_2O_3:Ti^{3+}," *IEEE J. Quant. Electron. QE-24*, 1049–1051, 1988.
72. P.A. Schulz, K.F. Wall, and R.L. Aggarwal, "Simple Model for Amplified Spontaneous Emission in a Ti:Al_2O_3 amplifier," *Opt. Lett. 13*, 1081–1083, 1988.
73. K.W. Kangas, D.D. Lowenthal, and C.H. Muller III, "Single-Longitudinal-Mode, Tunable, Pulsed Ti:Sapphire Laser Oscillator," *Opt. Lett. 14*, 21–23, 1989.
74. L.S. Wu, H. Looser, and P. Günter, "High Efficiency Intracavity Frequency Doubling of Ti:Al_2O_3 Lasers with $KNbO_3$ Crystals," *Appl. Phys. Lett. 56*, 2163–2165, 1990.
75. N. Sarukura, Y. Ishida, T. Yanagawa, and H. Nakano, "All Solid-State CW Passively Mode-Locked Ti:Sapphire Laser Using a Colored Glass Filter," *Appl. Phys. Lett. 57*, 229–230, 1990.
76. R. Scheps, J.F. Myers, and S.A. Payne, "CW and *Q*-Switched Operation of a Low Threshold Cr^{+3}:$LiCaAlF_6$ Laser," *IEEE Photon. Tech. Lett. 2*, 626–628, 1990.

77. R. Scheps and J.F. Myers, "Doubly Resonant Ti:Sapphire Laser," *IEEE Photon. Lett. 4*, 1–3, 1992.

78. P.F. Moulton, "Spectroscopic and Laser Characteristics of Ti:Al_2O_3," *J. Opt. Soc. Am. B. 3*, 125–132, 1986.

79. C.E. Byvik and A.M. Buoncristiani, "Analysis of Vibronic Transitions in Titanium Doped Sapphire Using the Temperature of the Fluorescence Spectra," *IEEE J. Quant. Electron, QE-16*, 85–87, 1991.

80. D.E. Spence and P.N. Kean, "60-fsec Pulse Generation From a Self-Mode-Locked Ti:Sapphire Laser," *Opt. Lett. 16*, 42–44, 1991.

81. J. Squier, F. Salin, G. Mourou, and D. Harter, "100 fs Pulse Generation and Amplification in Ti:Al_2AlO_3," *Opt. Lett.*, 324–326, 1991.

82. J. Squier, F. Salin, S. Coe, P. Bado, and G. Mourou, "Characteristics of an Active Mode-Locked 2 ps Ti:Sapphire Laser Operating in the 1 μm Wavelength Region," *Opt. Lett. 16*, 85–87, 1991.

83. C. Spielmann, F. Krausz, T. Brabes, E. Wintner, and A.J. Schmidt, "Femtosecond Pulse Generation From a Synchronously Pumped Ti:Sapphire Laser," *Opt. Lett. 16*, 1180–1182, 1991.

84. C.H. Bair, P. Brockman, R.V. Hess, and E.A. Modlin, "Demonstration of Frequency Control and CW Diode Laser Injection Control of a Titanium-Doped Sapphire Ring Laser With No Internal Optical Elements," *IEEE J. Quant. Electron. QE-24*, 1045–1048, 1988.

85. B.F. Gächter and J.A. Koningstein, "Zero Phonon Transitions and Interacting Jahn-Teller Phonon Energies From the Fluorescence Spectrum of α-Al_2O_3:Ti^{3+}," *J. Chem. Phys. 60*, 2003–2006, 1974.

86. J.C. Walling, "Tunable Lasers," in *Tunable Lasers*, Ed. L.F. Mollenauer and J.C. White (New York: Springer-Verlag, 1988) pp. 331–352.

87. J.C. Walling, O.G. Peterson, J.P. Jennsen, Robert C. Morris, and E. W. O'Dell, "Tunable Alexandrite Lasers," *IEEE J. Quant. Electron. QE-16*, 1302–1315, 1980.

88. M.L. Shand, J.C. Walling, and R.C. Morris, "Excited-State Absorption in the Pump Region of Alexandrite," *J. Appl. Phys. 52*, 953–955, 1981.

89. J.C. Walling, "Alexandrite Lasers: Physics and Performance," *Laser Focus*, 45–50, 1982.

90. M.L. Shand, J.C. Walling, and H.P. Jennsen, "Ground State Absorption in the Lasing Wavelength Region of Alexandrite: Theory and Experiment," *IEEE J. Quant. Electron. QE-18*, 167–169, 1982.

91. M.L. Shand and J.C. Walling, "A Tunable Emerald Laser," *IEEE J. Quant. Electron. QE-18*, 1829–1830, Nov. 1982.

92. J.C. Walling, "Recent Advances in Alexandrite Laser Technology," *Laser Focus/Electro-Opt.* 213–215, Sept. 1983.

93. S.T. Lai and M.L. Shand, "High Efficiency CW Laser-Pumped Tunable Alexandrite Laser," *J. Appl. Phys. 54*, 5642–5644, 1983.

94. J. Hecht "Vibronic Solid State Lasers: Review and Update," *Lasers and Applications*, 77–82, Sept. 1984.

95. M.D. Perry and R. Olson, "LiCAF and LiSAF Amplify Tunable Laser Pulses," *Laser Focus World*, 69–72, Nov. 1991.

96. R. Scheps, J. Myer, and S. Payne, "CW and Q-Switched Operation of a Low Threshold Cr^{+3}:$LiCaAlF_6$ Laser," *IEEE Photon, Lett. 2*, 626–628, 1990.

97. J. Hecht, "Tunability Makes Vibronic Lasers Versatile Tools," *Laser Focus World*, 93–103, Oct. 1992.

98. L.K. Smith, S.A. Payne, W.L. Kway, L.L. Chase, and B.H.T. Chai, "Investigation of the Laser Properties of Cr^{3+}:$LiSrGaF_6$," *IEEE J. Quant. Electron, QE-28*, 2612–2618, 1992.

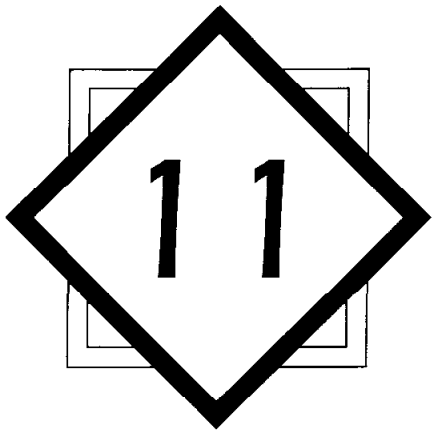
99. C. Deka, B.H.T. Chai, Y. Shimony, X.X. Zhang, E. Munin, and M. Bass, "Laser Performance of Cr^{4+}:Y_2SiO_5, *Appl. Phys. Lett, 61*, 2141–2143, 1992.

100. Michael G. Littman, "Single-Mode Operation of Grazing-Incidence Pulsed Dye Laser," *Opt. Lett. 3*, 138–140, 1978.

101. Karen Liu and Michael G. Littman, "Novel Geometry for Single-Mode Scanning of Tunable Lasers," *Opt. Lett. 6*, 117–120, 1981.

102. K.W. Kangas, D.D. Lowenthal, and C.H. Muller III, "Single-Longitudinal-Mode, Tunable, Pulse Ti:Sapphire Laser Oscillator," *Opt. Lett. 14*, 21–24, 1989.

103. W.S.C. Chang, *Principles of Quantum Electronics* (Reading, Mass.: Addison-Wesley, 1969), p. 155.



Semiconductor Lasers

11.1 INTRODUCTION

11.1.1 Overview

Semiconductor lasers are the last class of optical oscillators to be considered and are the most important for communications and control applications because of their convenience, efficiency, and natural compatibility with the rest of modern electronics. They have the simplest structure imaginable: Two pigtail leads to be connected to a power supply, a p-n junction where most of the physics resides, and cavity mirrors formed by cleaving (i.e., breaking) the crystal. A typical laser is shown in Fig. 11.1 with some artistic artifacts added to identify some of the problems to be addressed in this chapter.

As implied by Fig. 11.1, it is the injected current density that provides the optical gain to a wave whose spatial extent may be much larger than the active region. This is an electromagnetic problem that will be addressed in Chapter 12. Note also the small dimensions indicated in Fig. 11.1 and think about the implications based on our prior discussion of the larger gas or solid-state lasers. In order for the round trip gain (RTG) to exceed 1, the gain coefficient must be quite large. For instance, let the facet reflectivity be 0.3 (typical) and apply the RTG equation to compute $\gamma = 48.2\ \text{cm}^{-1}$, a value unheard of in our previous work. The injected current spreads in the plane of the junction so that the effective area is larger than the $10 \times 250\ \mu\text{m}^2$ of the contact. Let us arbitrarily pick a

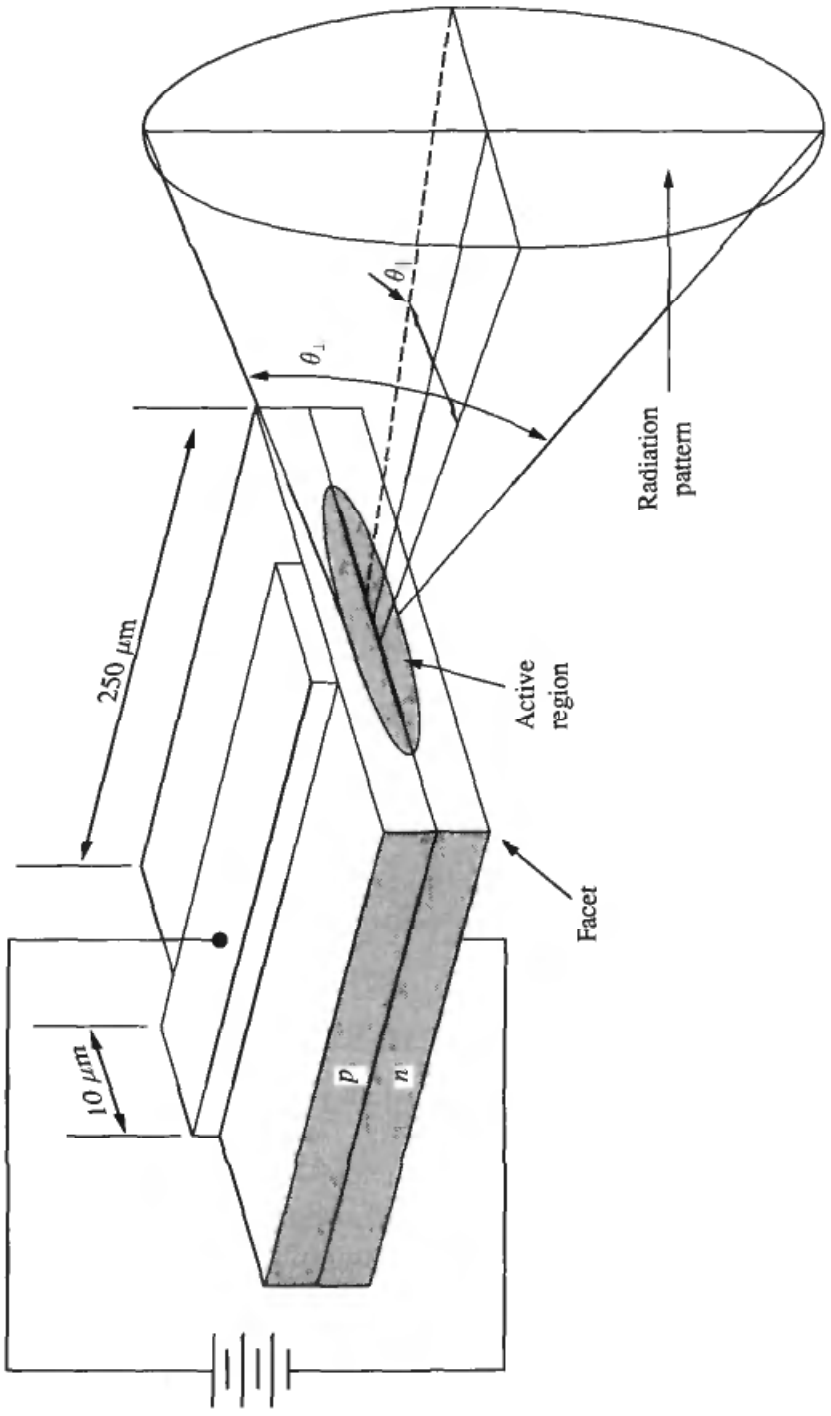


FIGURE 11.1. A generic *p-n* junction laser. The active region is at the junction of *p* and *n* regions where the current is flowing and is shown as solid black. The mode inside the semiconductor is more or less like an ellipsoid with a spatial extent much larger than the active region and is different perpendicular and parallel to the junction. This results in a radiation pattern with different spreading angles.

factor of 4 increase to account for this spreading. For a drive current of 100 mA, the current density is a rather startling 1 kA/cm^2. Such a laser might well operate at $> 50\%$ efficiency. The voltage source has to be some bigger than the band gap $\sim$ 1.43 V, so pick 2.0 V for this estimation. The electrical power used to drive this generic device is 200 mW with 100 mW emerging as optical radiation from a device smaller than this exclamation mark! No wonder there is so much excitement.

Some of their performance characteristics are considerably less than what can be achieved with other types of lasers. For instance, the output radiation is *not* a simple $TEM_{0,0}$, mode as it can be with a gas or solid-state laser; rather, it is usually a field whose dimensions along the plane of the junction are much larger than those perpendicular to it, and in either case, is on the order of a free space wavelength or so. Because of the small dimensions the beam diverges with angles on the order of a few degrees, and, because of the unequal dimensions, the beam spreads unequally in the two directions. The optical bandwidth over which oscillation is possible is enormous by gas or solid-state laser standards, and thus multimode oscillation is the norm unless special precautions are taken. Further, the spectral purity of the oscillating modes tends to be much worse than the corresponding values for the larger gas or solid-state types (but still better than the resolution of most optical instruments).

11.1.2 Populations in Semiconductor Laser

There are many phenomena that a semiconductor laser has in common with the larger ones that have been used in most of our analysis. For instance, both use an *inverted* population; hence our first task is to decide what constitutes such an inversion, or even more basic, what are the "populations"?

Let us establish an answer to this last question before anything else. We are interested in the stimulated recombination of holes and electrons generating a photon according to the following "chemical" equation.

$$e + h \longrightarrow (h\nu) > E_g \qquad (11.1.1)$$

and thus the electron and hole populations are involved. Only direct band-gap semiconductors permit this reaction with high probability for transitions between the lowest allowed states in the respective bands. The situation is shown in Fig. 11.2 where we plot the number of available states, shaded as being occupied, white as being empty, for the two types of semiconductors as a function of the momentum of a carrier. In that figure, we presume a few electron states in the conduction band are occupied and there are a few unoccupied states (i.e., holes) in the valence band. Note that in either case, the electrons will give up an energy roughly equal to the band-gap value in making the transition from a state *up* in the conduction band to one *down* in the valence band. For the most part, it does so by emitting a photon (11.1.1) for a direct material but finds some nonradiative process for the indirect material that competes for the electron population, and thus few photons are emitted. Why?

The reason is contained in the fact that both energy and momentum must be conserved in any reaction. Now conservation of energy is usually the easiest and most transparent. Initially, the electron is at or above the conduction edge, and it makes a transition to a

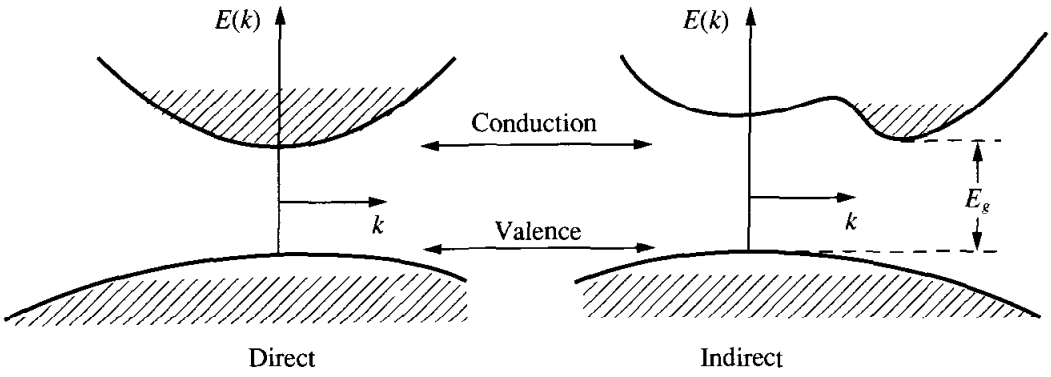


FIGURE 11.2. $E(k)$ vs. k (momentum) for direct and indirect semiconductors. The shaded areas are intended to imply states filled with an electron, and the white areas are empty.

state at or below the valence band edge. Hence, it yields an energy somewhat greater than $E_c - E_v = E_g$ to either a photon or to this other unnamed process.

Conservation of momentum requires a more careful scrutiny of the process since we require the vector sum of the final momentum of all partners to be equal to the initial value. Let us make this accounting for the average electron in the direct band-gap case. As we proceed, the numbers will point out the difficulty for the indirect case:

$$\hbar\mathbf{k}_f + \hbar\mathbf{k}_p = \hbar\mathbf{k}_i \tag{11.1.2}$$

where $\mathbf{k}_{f,p,i}$ equals the wave vector of the final electron state, the photon, and the initial state. Now $|\hbar\mathbf{k}|$ is the momentum of the electron, and we can approximate it by the effective mass, m_e^*, times its mean kinetic velocity found by equating its kinetic energy *in the band*, $m^*v^2/2$, to $3kT_e/2$ and letting the electron temperature equal the lattice temperature. For $m_e^* = 0.067m_0$ (typical of GaAs) and $kT/e = 0.026$ V, we find the random velocity for an electron in the conduction band to be

$$v_e = (3kT/m_e^*)^{1/2} = 2.61 \times 10^5 \text{ m/ sec} \tag{11.1.3a}$$

Thus

$$|m_e^* v_e| = \hbar k_i = 1.59 \times 10^{-26} \text{ kg} - \text{m/ sec} \tag{11.1.3b}$$

Consider the momentum of a photon with a wavelength of 8400 Å: $k_p = 2\pi/\lambda_0$, and

$$\hbar k_p = h/\lambda_0 = 7.89 \times 10^{-28} \text{ kg} - m/\text{ sec} \tag{11.1.4}$$

Note that the momentum of the photon is very small compared to that of the average electron, and thus the momentum of the hole or final state must be just about equal to the starting value. In other words, *the transition must be vertical* and

$$\Delta\mathbf{k} = \mathbf{k_i} - \mathbf{k_f} \approx 0 \tag{11.1.5}$$

This is usually referred to as the "**k** conservation" rule.

Usually there are enough scattering events in a direct material to relax the **k** conservation rule somewhat, but the problem becomes much worse for the indirect materials where

TABLE 11.1 Properties of Common Semiconducting Materials

Material	Band Gap (eV)	ϵ_r	μ_e	μ_h	m_e^*/m_0	Lattice (Å)
			(cm^2/V-sec)			
C(i)	5.47	5.7	1800	1200	0.2	3.5668
GaP(i)	2.26	11.1	1600	100	0.82	5.4512
AlAs(i)	2.16	10.9	180	—	—	5.6605
GaAs(d)	1.43	13.2	8500	400	0.067	5.6533
InP(d)	1.35	12.4	4600	150	0.077	5.8686
Si(i)	1.12	11.9	1500	450	1.1	5.4309
GaSb(d)	0.72	15.7	5000	850	0.042	6.0957
InAs(d)	0.36	14.6	33, 000	460	0.023	6.0584
InSb(d)	0.17	17.7	80, 000	1250	0.0145	6.4794

the electrons at the indirect minima of Fig. 11.2(b) have a **k** vector on the order of π/a (a is lattice constant). For a typical distance—say, 5.45 Å(GaP)—the momentum is enormous compared to the previous numbers: $\hbar k = 6.08 \times 10^{-25}$ kg-m/sec. This is 40 times the kinetic momentum of the average electron, which in turn was 20 times that of the photon.

Optical transitions can and do take place in indirect materials—the "green" LED using GaP is an experimental existence proof—but the oscillator strength for radiation is much weaker than that of a direct material. The amount of radiation that a group of electron-hole pairs produces is really a matter of competition between radiative and nonradiative rates. The latter tends to dominate in indirect materials, especially in the presence of impurities or defects. We will concentrate on the direct materials listed in Table 11.1 in which the radiative rates are quite fast.

There are many similarities between semiconductor lasers and the much larger gas and solid-state ones, but there are distinct differences also. Hence our first job is to reestablish our connection with elementary semiconductor theory such as that used to describe a *p-n* junction, transistor, or photodiode so as to identify an "inverted" population.

11.2 REVIEW OF ELEMENTARY SEMICONDUCTOR THEORY

It is assumed that most of the readers are familiar with many of the issues associated with semiconductors, and many of the elementary ideas that have been used again and again in elementary electronic courses will be skipped. However, there a few ideas that are crucial to laser understanding, and these will be discussed in detail. The first task is to determine how many electronic states are available to be filled by an electron in the conduction band or emptied in the valence band. We proceed in two steps: first to identify the number available and then to compute the probability that the state is filled (or empty). In a loose sort of way, it is the filled states in the conduction band minus the filled states in the valence band which correspond to $N_2 - N_1$ of our earlier description of lasers.

11.2.1 Density of States

The fact that elementary particles, such as electrons, can be described by a wave function is an idea that has been established again and again in elementary courses in physics, chemistry, and electronics. We need the idea here also. It should also come as no surprise that much of the notation used to describe these electronic wave functions in a crystal is the same as that used in electromagnetic theory. After all, both are wave phenomena and both have quantization forced on the wave functions for the same reasons, namely, boundary conditions.

For a semiconductor of size L_x, L_y, L_z, we invoke the most rudimentary and reasonable type of requirement; namely, the electron (or hole) should be found somewhere inside the volume (and not outside). We assume that the wave function for an electron varies as

$$\psi(\mathbf{r}) = u(r) \exp(-j\mathbf{k} \cdot \mathbf{r}) \tag{11.2.1}$$

where $u(r)$ reflects the effect of the lattice through which the electron wave is propagating and is the function that generates the familiar valence and conduction bands. Let us concentrate on the physics of the propagating factor. For a rectangular box of semiconductor with sides L_x, L_y, and L_z, we must have standing waves between the extremes, and thus the components of k must be related to these dimensions by

$$k_x = m\pi/L_x \qquad k_y = p\pi/L_y \qquad k_z = q\pi/L_z \tag{11.2.2}$$

for

$$\mathbf{k} = k_x\mathbf{a}_x + k_y\mathbf{a}_y + k_z\mathbf{a}_z \tag{11.2.3}$$

To compute the density of states in the bands we need to count (yes, count) the number of combinations of quantum numbers m, p, and q that are permitted for various energies with respect to the band edges. The drudgery of this seemingly uninspiring calculation can be minimized by constructing a "mode" diagram similar to that used in Chapter 7 for electromagnetic modes in a blackbody cavity. However, let us proceed independently of that prior development and attempt to describe the quantization requirement (11.2.2) by the "k-space" diagram of Fig. 11.3. The axes are k_x, k_y, k_z, and an allowed value of $\mathbf{k}$ is the vector from the origin to one of the triplets of points labeled by the quantum numbers (m, p, q) with the appropriate multiplicative factors $\pi/L_{x,y,z}$ understood. There is one last item to be noted about the construction of Fig. 11.3: There is exactly one set of quantum numbers (or modes) per elementary volume in k space: $V_k = (\Delta k_x = \pi/L_x)(\Delta k_y = \pi/L_y)(\Delta k_z = \pi/L_z)$. Thus each state occupies a volume $\pi^3/L_xL_yL_z$ in k space, or

$$\begin{aligned}\text{density of modes in } k\text{-space} &= (1 \text{ mode})/(\text{volume of each mode}) \\ &= L_xL_yL_z/\pi^3\end{aligned} \tag{11.2.4}$$

Now comes the step that saves us the drudgery of counting. We recognize that these points are very close together for a macroscopic piece of semiconductor. For instance, let $L_x = L_y = L_z = 1 \text{ mm} = 10^{-3}$ m. Choose some multiplicative factor—say, $(m^2 + p^2 + q^2)^{1/2} = 100$—and multiply it by (π/L_x) to obtain a value of $k \approx 3.14 \times 10^5\ m^{-1}$ as a test case. Let us translate this value into a more comfortable one by computing the momentum and kinetic energy of the electron with this value of k. The momentum is trivial,

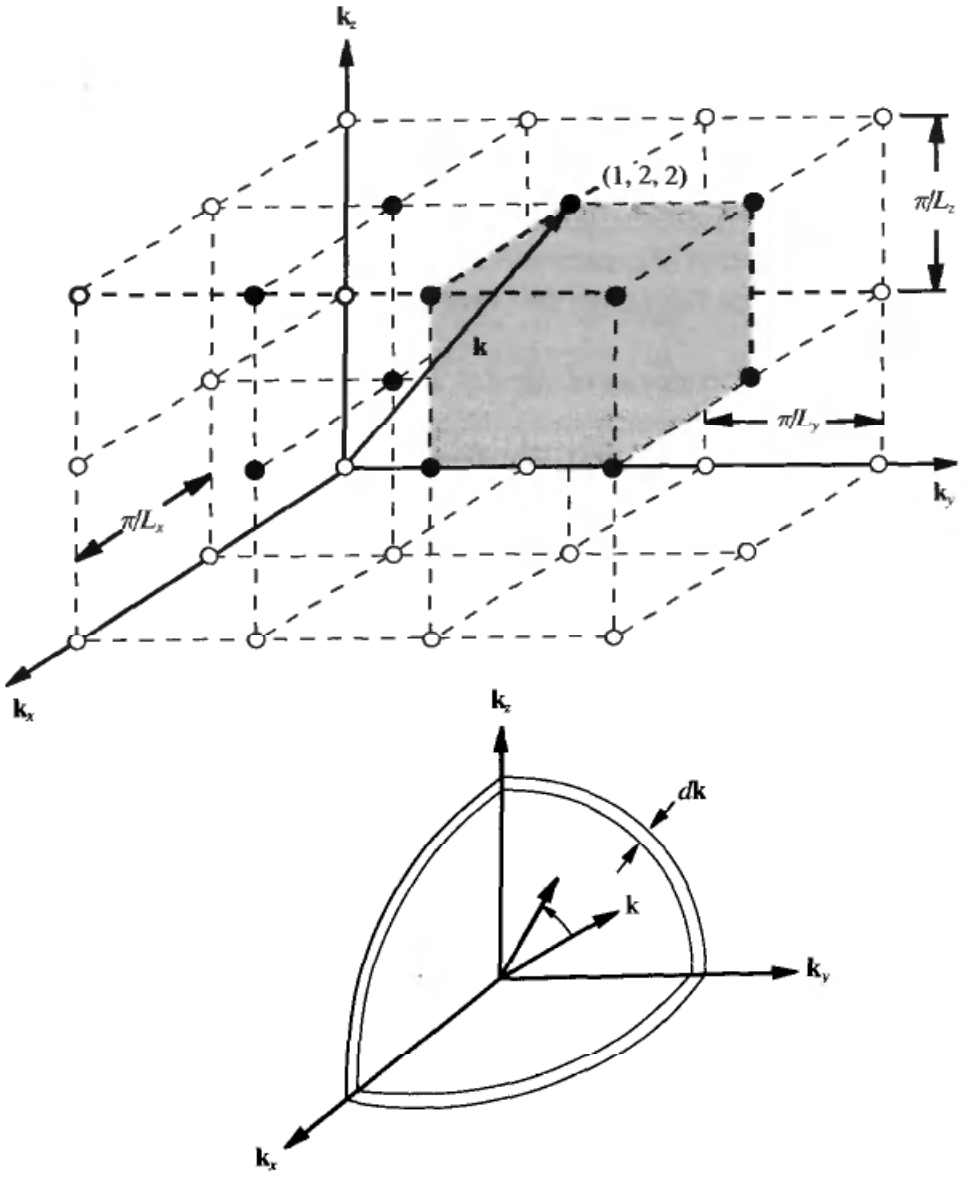


FIGURE 11.3. Mode diagram for a "large" semiconductor. The top sketch shows each individual allowed value of **k** whereas the insert implies that the points are very close together so that they could be considered as a continuum.

$p = \hbar k = 3.31 \times 10^{-29}$ kg $-$ $m/$ sec. If we assume an effective mass $m^* = 0.067\, m_0$, we find an equivalent speed of this electron equal to 543 m/sec, hardly blinding. The kinetic energy is also unimpressive, $E = (\hbar k)^2/2m^* = 8.99 \times 10^{-27}$ joules $= 5.6 \times 10^{-8}$ eV. (This energy is to be added to the conduction band value or subtracted from the valence band one.)

The point to be drawn from these numbers is that the mode numbers must be huge, $(m^2 + p^2 + q^2)^{1/2} \approx 10^5$ or so, before the energy is comparable to the thermal value $kT/e = 0.0259$ eV, or the spacing between them must be very small for ranges of **k** of interest. The smaller scaled version shown as an insert in Fig. 11.3 illustrates the simple computation involved in counting the states. The wave vector can range from along k_z, k_y,

or k_x or any combination. For $|\mathbf{k}|$ = constant, this describes one eighth of a sphere, and if we allow $\mathbf{k}$ to expand to $\mathbf{k} + d\mathbf{k}$, this puts a shell thickness dk on this fraction of that sphere. The volume of that shell, $(4\pi k^2 dk/8)$ times the density of points in k-space is the number of states, $N_k dk$, encompassed by allowing k to vary from k to $k + dk$.

$$N_k dk = \left\{L_x L_y L_z/\pi^3\right\}\left\{1/8\right\}\left\{4\pi k^2 dk\right\} \tag{11.2.5}$$

Remember that we are dealing with momentum states of an electron that also has two possible spin orientations. Hence, (11.2.5) must be multiplied by a factor of 2 and the density of states in k-space per unit of volume becomes

$$\rho_k dk \cdot V = 2 \cdot N_k dk$$

or

$$\rho_k dk = 2N_k dk/L_x L_y L_z = k^2 dk/\pi^2 \tag{11.2.6}$$

Most feel more comfortable expressing the density of states in terms of the energy of the carrier (electron or hole) beyond the band edge (i.e., above the conduction band or below the valence band edges). The energy and momentum *in a band* are related by

$$E = (\hbar k)^2/2m^* = (\epsilon - E_c) \qquad \text{or} \qquad (E_v - \epsilon) \tag{11.2.7}$$

where ϵ is the total energy and E is an energy measured from the edge of a band. Thus

$$k = (2m^* E/\hbar^2)^{1/2}$$

$$dk = \left(\frac{2m^*}{\hbar^2}\right)^{1/2} \frac{dE}{2E^{1/2}}$$

Substituting these relations into (11.2.6), we obtain the density of states in either the conduction or valence band.

$$\rho_{c,v}(E)\,dE = \left\{\frac{1}{2\pi^2}\right\}\left\{\frac{2m^*}{\hbar^2}\right\}^{3/2}\left\{E^{1/2}dE\right\} \quad (\text{in m}^{-3}) \tag{11.2.8}$$

where E is measured from the band edge, that is, up into the conduction band or down into the valence band, and m^* is the effective mass of each of the carriers. This is sketched in Fig. 11.4 for a semiconductor with an effective mass of the electron equal to 0.067 m_0 and that of the "heavy or normal" hole* of $\sim$ 4 times that value. [For GaAs, the actual value of the mass of the heavy hole is 0.55 m_0. It should be apparent from (11.2.8) and Fig. 11.4 that the curvature of the band is related to the effective mass in that band, and if the actual value of m_h had been used, the curvature of the valence band would be 1/23.5 of that of the conduction and the lower solid curve of Fig. 11.4 would have appeared as a straight line.]

There are a couple of points to be noted. First, the density of states for holes per unit of energy dE is much greater than that for electrons, a fact that is a consequence of the

*In the valence band of III-V semiconductors, there is a light hole band that has a maximum energy identical to that of the normal or heavy hole at $k = 0$. However, because of its smaller mass (more or less that of an electron), the density of states of the light hole is much less than that of the normal hole, and thus most holes at a given energy in that band are "heavy" holes.

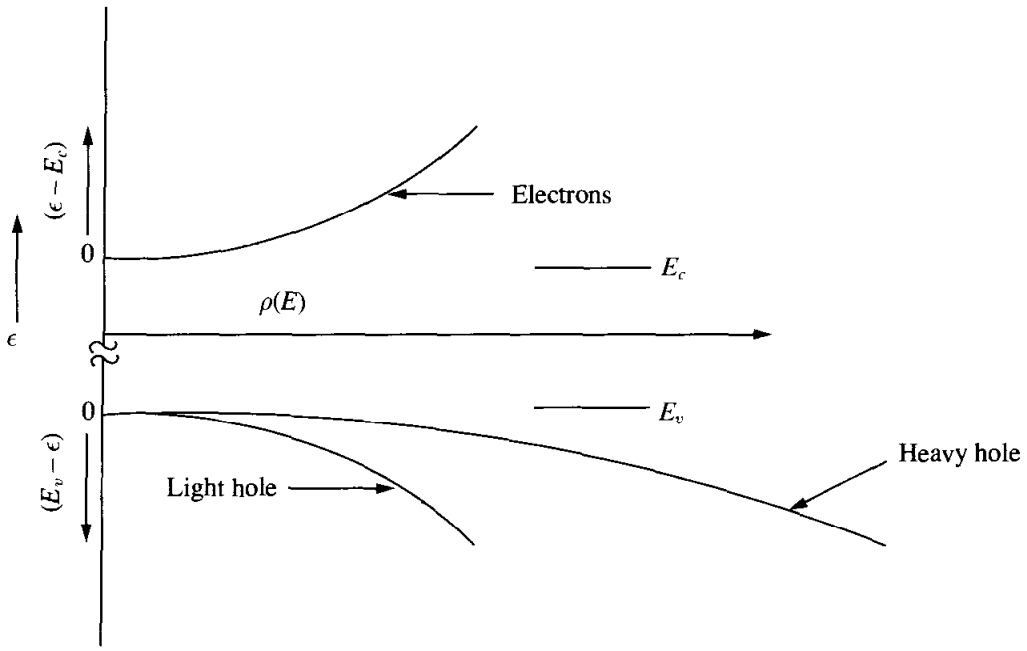


FIGURE 11.4. Density of states in a semiconductor with $m_{hh}^* = 4 \cdot m_e^*$ and $m_{lh} = m_e^*$.

disparity in effective masses. The second point is that this represents a fairly large number of quantum states to be filled. For instance, suppose we wished to fill every available state in the conduction band from E_c to $E_c + \Delta E$ with the rest being empty. Thus the density of electrons that must be created in this sample, either by injection or by some other means, is found by integrating (11.2.8) from 0 to ΔE. This presumes that every state is filled or emptied with equal probability (an issue that will be addressed more precisely in the next section).

$$N_e = \int_0^{\Delta E} \rho_c(E)\, dE = \frac{1}{3\pi^2}\left[\frac{2m_e^* \Delta E}{\hbar^2}\right]^{3/2} \quad (\text{in m}^{-3}) \tag{11.2.9}$$

Suppose $\Delta E = 0.1\text{ V} = 1.6 \times 10^{-20}$ joules and $m_e^* = 0.067\, m_0$ using GaAs as an example. Then we would need 2.5×10^{24} electrons/m^3 or 2.5×10^{18} cm^{-3} to fill every state that is less than 0.1 V above the conduction band edge. If we simultaneously created the same number of holes in the valence band (so that the electrons could make an optical transition to an empty state) then we can use (11.2.9) to find the empty energy interval occupied by the holes. Performing the same type of arithmetic we find that the empty interval of the valence band is only $(m_e^*/m_h^*)\Delta E_c = 0.0122$ V. Of course, the above calculations presume a complete filling (or emptying) of one of the bands up to the levels involved, which does not happen at a finite temperature, but it does illustrate the numerical factors involved.

Let us use the example considered above to estimate the distribution of photon energies resulting from the recombination of electrons within 0.1 eV of E_c to empty states, of the holes in the valence band that are 0.0122 eV below E_v. Let us arbitrarily pick the center of the recombination spectra to be midway between the allowed extremes, $E_g < h\nu <$

$E_g + \Delta E_c + \Delta E_v$. For $E_g = 1.43$ eV, we obtain

$$h\nu/e = E_g + (0.1 + 0.0122)/2 = 1.486 \text{ V (i.e., } \lambda_0 = 0.834\ \mu\text{m)}$$

A reasonable estimate for the spectral width is one fourth of the energy spread within the two bands:

$$h\Delta\nu/e = (0.1 + 0.0122)/4 = 0.02805 \text{ V}$$

Now using $\Delta\nu/\nu = \Delta E/E$ we obtain $\Delta\nu = 6.788$ THz. This is a rough estimation for the bandwidth over which oscillation might be possible; it is enormous by gas or solid-state laser standards.

Of course, the numbers are only approximate because of the assumption of completely filled (or emptied) states. We need to fold in the probability of a state being occupied, a task to which we now turn.

11.3 OCCUPATION PROBABILITY: QUASI-FERMI LEVELS

In the previous section, we merely identified the states and blindly assumed all states filled to a certain level with equal probability. If the lattice temperature were 0 K, this is legitimate, but at a finite temperature, we must fold in the probability of a state being occupied. The Fermi function

$$f(\epsilon) = \frac{1}{\exp[(\epsilon - E_f)/kT] + 1} \tag{11.3.1}$$

describes the probability of a state of energy ϵ, if it exists, being occupied and thus $1 - f(\epsilon)$ is the probability that it is empty. Obviously, at $\epsilon = E_f$, the probability is 1/2 (at all temperatures). If the Fermi level is in the forbidden gap, $E_v < E_f < E_c$, and if the difference $(\epsilon - E_f) \gg kT$, then the Fermi function can be approximated by the Boltzmann factor and the probability of an electron state being occupied at an energy ϵ above E_f is given by

$$f(E) \approx \exp -[(\epsilon - E_f)/kT]$$

Such an approximation is quite often used for conventional semiconductor electronics. It is not valid for the active region of a semiconductor laser because, as we shall see, the Fermi level must be *within a band* and the inequality is not obeyed. Hence we will always use the complete expression given by (11.3.1).

We have another much more serious deviation from semiconductor electronics; that is we need *electrons and holes* to exist simultaneously at the same place. You should recognize that (11.3.1) presents a problem (before we try to solve it). If the Fermi level moves up to the conduction band, then (11.3.1) indicates that most of the states below E_f are *filled*! Thus the electron cannot make a transition to the valence band because all of the available states are taken. The problem is that (11.3.1) deals with an equilibrium situation, whereas a laser uses an *inverted* system that is far from thermodynamic equilibrium. To handle this situation we borrow Shockley's concept of a quasi-Fermi level F_n to describe the probability of a state

being filled in the conduction band and a level F_p to describe the probability that a state in the valence band is filled. At equilibrium $F_n = F_p = E_f$.

The idea is based on the fact that intraband relaxation occurs on a time scale of 10^{-12} sec, which is much less than the interband (recombination) rate. Hence, electrons achieve a quasiequilibrium situation in each band occupying the lowest available states according to the Fermi function with the appropriate quasi-Fermi level for that function. The probability of an occupied state in the conduction band is

$$f_c(\epsilon) = \frac{1}{\exp[(\epsilon - F_n)/kT] + 1} \tag{11.3.2}$$

and the probability of an occupied state in the valence bond is

$$f_v(\epsilon) = \frac{1}{\exp[(\epsilon - F_p)/kT] + 1} \tag{11.3.3}$$

The probability of an empty state (i.e., a hole) is, of course, $1 - f_{c,v}$ for each band.

To obtain the density of electrons in the conduction band in the energy interval $d\epsilon$, multiply the density of states of that band (11.2.8) by the Fermi function (11.3.2), to obtain

$$n_c(\epsilon)\, d\epsilon = \frac{1}{2\pi^2}\left[\frac{2m_e^*}{\hbar^2}\right]^{3/2} \frac{(\epsilon - E_c)^{1/2} d\epsilon}{\exp[(\epsilon - F_n)/kT] + 1} \qquad \epsilon > E_c \tag{11.3.4}$$

To obtain the number of holes in the interval $d\epsilon$ below E_v, use the density of states for the valence bond and multiply by $1 - f_v(\epsilon)$, which is the probability that the state is empty.

$$p_v(\epsilon)\, d\epsilon = \frac{1}{2\pi^2}\left[\frac{2m_h^*}{\hbar^2}\right]^{3/2} \frac{(E_v - \epsilon)^{1/2} d\epsilon}{\exp[(F_p - \epsilon)/kT] + 1} \qquad \epsilon < E_v \tag{11.3.5}$$

where the inequalities merely reflect the fact that there are no states to be filled (or emptied) for $E_v < \epsilon < E_c$ (with no doping).

Figure 11.5 is a sketch of the distribution of electrons for the case considered previously for the two different values of temperature, $kT/e = 0$ (0 K) and $kT/e = 0.0259$ V, with $F_n = E_c + 0.1$ eV. At 0 K, all of the states up to F_n in the conduction band are filled with all of the states above F_n being empty.

11.4 OPTICAL ABSORPTION AND GAIN IN A SEMICONDUCTOR

We are now in the position to consider situations that lead to gain in a semiconductor, by recognizing that this is just the opposite to those leading to absorption. For instance, consider the "Gedanken" or (thought) experiment shown in Fig. 11.6 where a small-signal variable frequency source irradiates a slab of semiconductor, and the log of $I_{\text{out}}(\nu)/I_{\text{in}}(\nu)$ is plotted as a function of frequency (or wavelength) as data. To make our initial task simple, let us assume $T = 0$ K and consider an intrinsic material with $F_n = F_p = E_f$ at mid-gap.

If $h\nu_s < E_g$, then there are plenty of occupied states in the valence band (all of them!) but no state for the electron to occupy if it is promoted to a higher energy state. Hence there

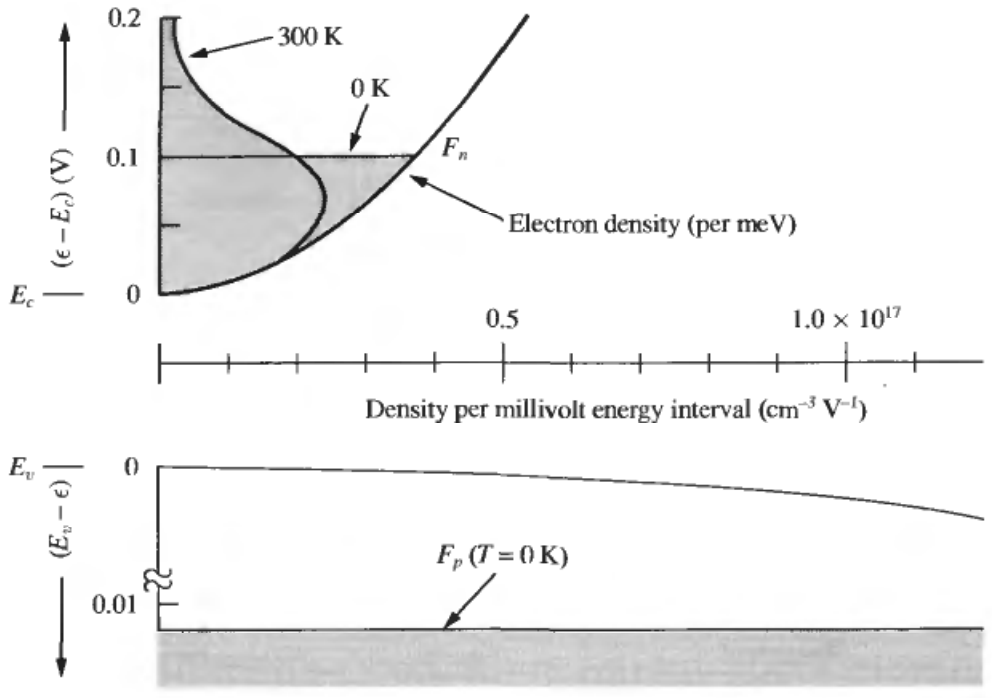
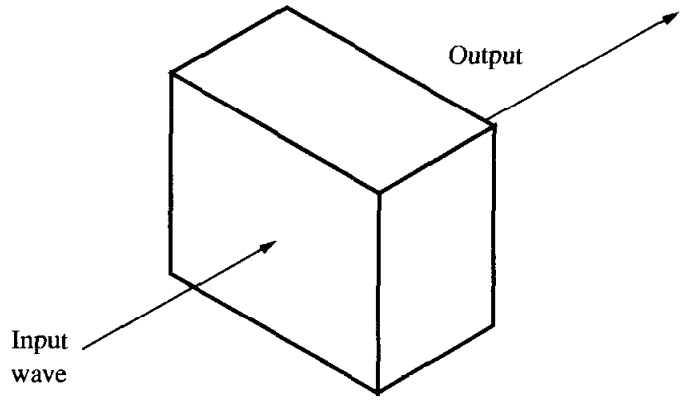


FIGURE 11.5. Distribution of electron and hole populations at $T = 0$ K and $T = 300$ K. Note that the energy scales are different for the conduction and valence bands. The sketch is shown for GaAs with $m_e^* = 0.067\ m_0$, $m_{hh} = 0.55\ m_0$, and $kT/e = 0.0259$ V. The presence of light holes has been ignored.

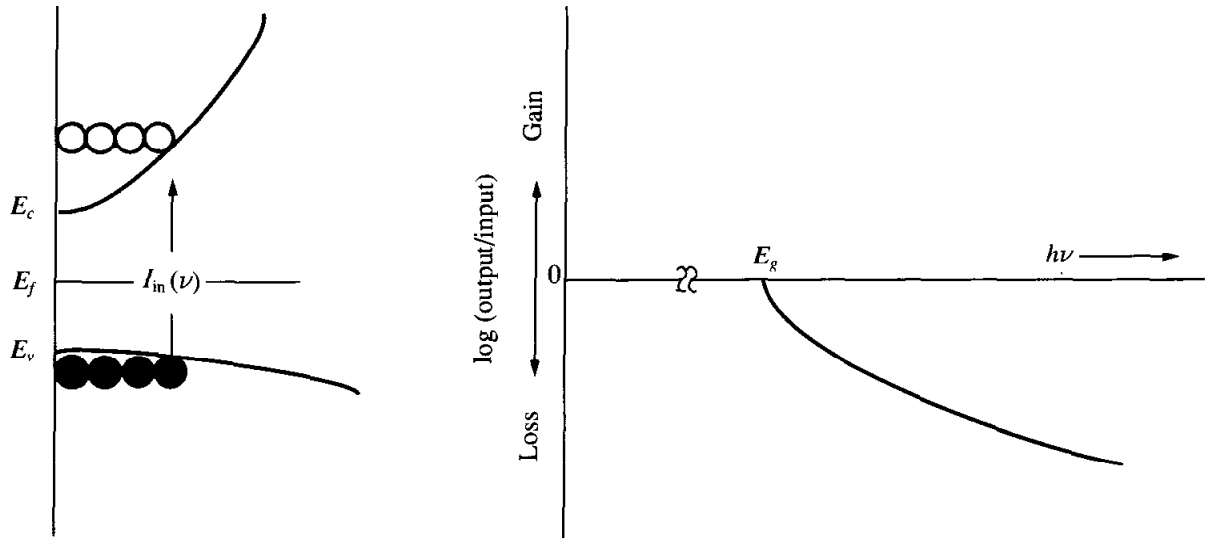
is no absorption. When $h\nu_s > E_g$ true absorption starts. We promote an electron from a filled state in the valence band to an empty state in the conduction band. Without being complicated and esoteric, we would guess that this absorption coefficient is proportional to the number of states available for absorption provided there is the appropriate number of empty states to receive the electrons. Obviously, as $h\nu$ becomes bigger than E_g, both numbers become larger, varying as the density of states in each band and hence the absorption coefficient also gets larger. The "data" are sketched in Fig. 11.6(c).

Now let us force the system away from equilibrium by using a very strong optical pump I_p whose path overlaps the probing wave in the manner shown in Fig. 11.7. We presume that this pump wave is strong enough to maintain a significant number of electrons in the conduction band. Because intraband relaxation to the extremities of each is very fast, the states from E_c to F_n fill and the states from E_v to F_p empty. Because of our assumption of 0 K, all states $E_g < E < F_n$ are filled and all states $E > F_n$ are empty. (Similar comments apply to the valence band.)

The exact position of $F_n - F_p$ relative to the pump photon energy, $h\nu_p$, depends on the intensity I_p, which must maintain the electrons in the conduction band against their "leaking" out downward to the hole states by recombination. It is similar to our attempting to fill a bucket with some holes in it. If the water flow is small, little water accumulates in the bucket. If the water flow from the hose (read the pump intensity) is large enough, then the bucket (i.e., conduction band) can fill up to a level such that the "leaks" equal the input. More about the kinetics of these leaks later.



(a) The experiment

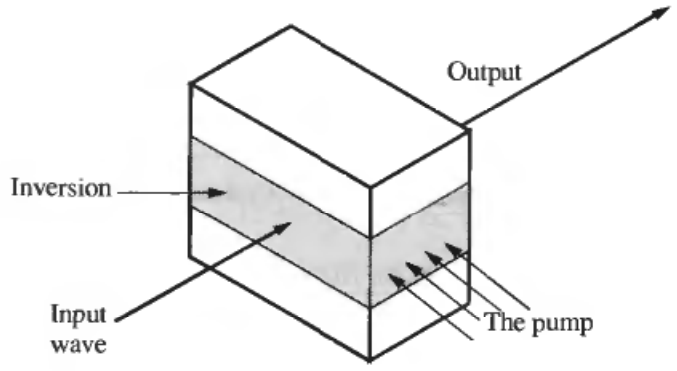


(b) The band diagram

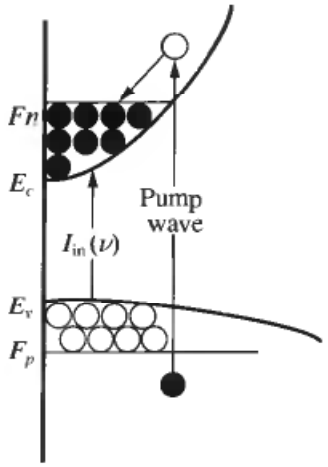
(c) The "data"

FIGURE 11.6. Optical absorption experiment: (a) indicates a wave propagating through a slab of semiconductor, (b) shows the band diagram of a "normal" semiconductor, and (c) indicates the data.

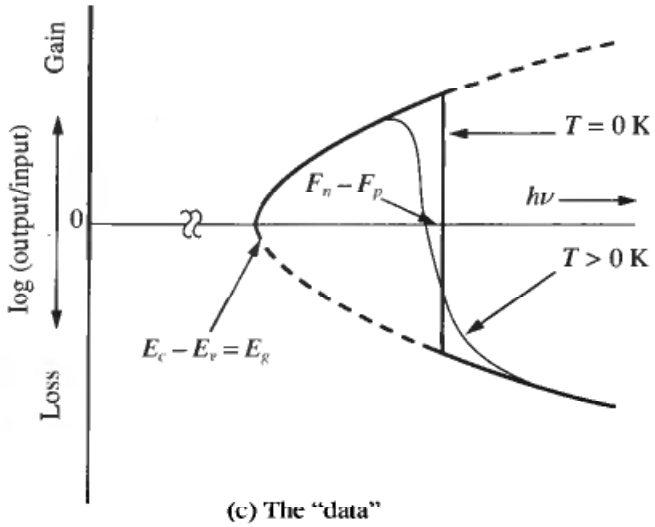
Let us tum our attention to the small-signal, variable-frequency probing wave and talk our way through its interaction with the crystal as $h\nu_s$ is varied. As before, there is no absorption or gain if $h\nu_s < E_g$ because there are not states in the forbidden gap. If $h\nu_s > E_g$, there are states separated by the photon energy but *inverted* from that considered previously. Now we have occupied states *up*, and unoccupied states *down*. The effect on the probing wave is just the opposite to that found for the equilibrium case: *It experiences gain in the same proportion as it had experienced absorption.*



(a) The experiment



(b) The band diagram



(c) The "data"

FIGURE 11.7. An optically pumped amplifier with the same conditions as were presumed for Fig. 11.6 with the addition of a strong pump creating an inversion by moving the quasi-Fermi levels into the bands.

Thus the gain coefficient is the mirror image about the horizontal (or frequency) axis of the absorption coefficient as shown in Fig. 11.7. If you recall from Chapter 7 that stimulated emission and absorption are just reverse processes, then this change of absorption to amplification is just a natural result of the inversion with filled states up (i.e., electrons) and empty states (holes) down. This recognition rescues us from the task of performing detailed quantum calculations to obtain the photon-electron-hole interaction rates. The gain does *not* persist for all frequencies. If $h\nu_s > F_n - F_p$, then we have a normal situation with filled states down and empty states up and there is absorption. Consequently the necessary condition for amplification in a semiconductor is that a pumping mechanism creates an "*inversion*" expressed by

$$F_n - F_p > h\nu > E_g \tag{11.4.1}$$

This was first pointed out by Bernard and Duraffourg [16] and was used by Dumke [13] in estimating the possibility of amplification in a semiconductor.

A finite temperature $T > 0$ K rounds off the sharp points and makes the transition from gain to absorption in a smooth fashion as is shown in the development of the next section.

11.4.1 Gain Coefficient in a Semiconductor

Let us consider optical transitions from a small band dE_2 centered at E_2 that is ΔE_c above E_c to a band dE_1 centered at E_1 located at ΔE_v below the valence band as is shown in Fig. 11.8. The notation, E_2 for the upper and E_1 for lower, has been chosen deliberately to emphasize a connection to other lasers.

The picture presented by Fig. 11.8 is deceptively simple. It appears that one merely adjusts the position of E_2 and E_1 such that the difference is the photon energy under consideration (either amplification or absorption). Can it be any combination of positions such that $E_2 - E_1 = h\nu_s$ with E_1, E_2 in the respective bands? No, these are optical transitions and k_c, measured in the conduction band, and k_v, measured in the valence band, must be equal before and after the absorption or emission of a photon (for a "pure" material). Thus the first question is; given $h\nu$, where are E_2 and E_1 with respect to E_c and E_v?

To answer this we return to the relationship between the allowed values of electron momentum, $\hbar k_{c,v}$, and energy in the respective bands:

$$\hbar k_c = [2m_e^*(E_2 - E_c)]^{1/2} \tag{11.4.2a}$$

$$\hbar k_v = [2m_h^*(E_v - E_1)]^{1/2} \tag{11.4.2b}$$

Since the transitions must conserve momentum, $k_c - k_v = k_{\text{opt}} \approx 0$, we obtain

$$\Delta E_c = (E_2 - E_c) = \frac{m_h^*}{m_e^*}(E_v - E_1) = \frac{m_h^*}{m_e^*}\Delta E_v \tag{11.4.3}$$

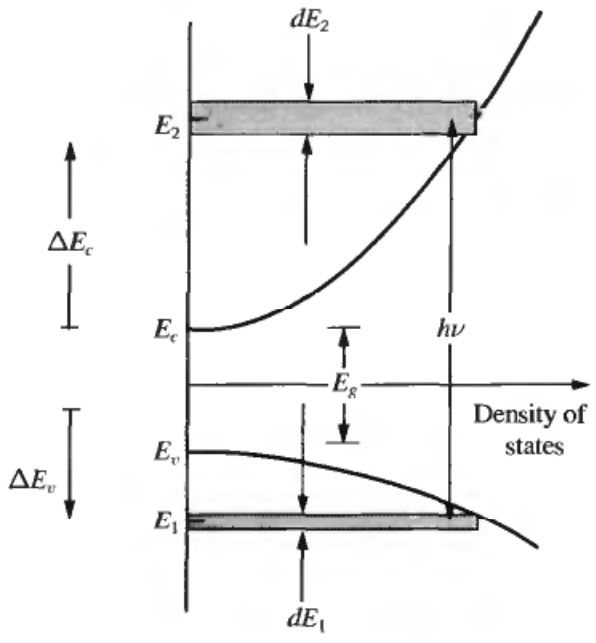


FIGURE 11.8. Optical transitions in a semiconductor.

which again points out that most of the "action" takes place in the conduction band because of the heavy mass of the holes.

The photon energy is $h\nu = E_2 - E_1$, hence

$$E_g + (E_2 - E_c) + (E_v - E_1) = h\nu = E_g + \Delta E_c + \Delta E_v \tag{11.4.4}$$

After combining (11.4.4) with (11.4.3) we obtain

$$E_2 - E_c = \frac{m_h^*}{m_h^* + m_e^*}(h\nu - E_g) \tag{11.4.5a}$$

and

$$E_v - E_1 = \frac{m_e^*}{m_h^* + m_e^*}(h\nu - E_g) \tag{11.4.5b}$$

If we combine (11.4.5a) and (11.4.5b), we obtain the obvious from Fig. 11.8; that is, $(E_2 - E_c) + (E_v - E_1) = h\nu - E_g$.

Now we ask the second question: Are the differential energy spreads shown in Fig. 11.8 equal? The answer is again no. From (11.4.3) we obtain

$$dE_2 = \frac{m_h^*}{m_e^*} d(-E_1) \tag{11.4.6}$$

where the minus sign reflects the fact that hole state energy increases downward in Fig. 11.8.

Not all states in the energy interval dE_2 and dE_1 can participate in the optical transition. The spin of the state must also be conserved, and this fact divides the total density of states in dE_2 by a factor of two. The conservation of momentum (i.e., $\hbar k$) puts an additional restriction on the number of states available, for not all states in dE_2 can terminate anywhere in dE_1. To examine this issue, we return to (11.2.6) and divide it by 2 to find the density of states in the interval k to $k + dk$.

$$\rho_{c,v}(k)\, dk = \frac{1}{2\pi^2} k^2 dk \text{ (one spin state)} \tag{11.4.7}$$

This formula applies to either or both bands, and thus the subscripts (c, v) are not used to identify k. However, most are more comfortable with an energy level diagram, such as Fig. 11.8, rather than a plot of $\rho_c(k)$ versus k, so we proceed along that line instead. The *reduced (or joint) density* of states is defined to reflect the number of states at E_2 and E_1, within dE_2 and dE_1 which *can* participate in the transition at $h\nu$, and which do conserve spin and momentum. We note that

$$\rho_{\text{jnt}}(h\nu)\, dE = \rho(k)\, dk = \frac{1}{2\pi^2} k^2 dk \tag{11.4.8}$$

Hence

$$\rho_{\text{jnt}}(h\nu) = \frac{1}{2\pi^2} k^2 \frac{dk}{dE} \qquad \text{(evaluated at } h\nu = E_2 - E_1) \tag{11.4.9}$$

and thus $\rho_{\text{jnt}}(h\nu)$ is the density of states (L^{-3}) per energy interval between $h\nu$ and $h\nu + d(h\nu)$ that can participate in the photonic process while conserving momentum. (Whether the participation is by absorption or stimulated emission comes later.)

We add the differential forms (11.4.2a) and (11.4.2b) to obtain an expression for dE:

$$dE = dE_2 - dE_1 = \hbar^2 \left[\frac{1}{m_e^*} + \frac{1}{m_h^*} \right] k dk$$

Solving for dk/dE and substituting the result into (11.4.9), one obtains

$$\rho_{\text{jnt}}(h\nu) = \frac{1}{2\pi^2\hbar^2} \frac{m_e^* m_h^*}{m_e^* + m_h^*} k \tag{11.4.10}$$

Now we invert (11.4.2a) to find k in terms of $E_2 - E_c$ and use (11.4.5a) to express that difference in terms of $(h\nu - E_g)$.

$$\hbar k = \left\{ \left[\frac{2m_e^* m_h^*}{m_e^* + m_h^*} \right] (h\nu - E_g) \right\}^{1/2} \tag{11.4.11}$$

and finally

$$\rho_{\text{jnt}}(h\nu) = \frac{1}{4\pi^2} \left(\frac{2m_r}{\hbar^2} \right)^{3/2} (h\nu - E_g)^{1/2} \tag{11.4.12a}$$

where

$$m_r = \frac{m_e^* m_h^*}{m_e^* + m_h^*} = \text{the reduced mass} \tag{11.4.12b}$$

Some prefer to write (11.4.12a) as

$$\rho_{\text{jnt}}(h\nu) = \frac{1}{2} \left[\frac{1}{\rho_c(E_2)} + \frac{1}{\rho_v(E_1)} \right]^{-1} \tag{11.4.12c}$$

which is an obvious problem "saved" for the student.

Having identified where and what states *can* be involved, we can now ask how many transitions ($L^{-3}s^{-1}$) *do* take place. We proceed in an identical fashion to that used in Chapter 7; the density of transitions from 2 $\rightarrow$ 1 (or from 1 $\rightarrow$ 2) caused by photons between ν and $\nu + d\nu$ is proportional to

1. An Einstein B coefficient just as in the rate equations for atomic lasers. ($L^{-3} - T^{-1} - E^{-1}$)
2. The energy density of the photons between ν and $\nu + d\nu$ causing these optical transitions, i.e. $\rho(\nu)\, d\nu$ ($E - L^{-3}$)
3. The joint density of states (14.4.12) but expressed as a density per unit of frequency: $\rho_{\text{jnt}}(\nu) = h\rho_{\text{jnt}}(h\nu)$ ($L^{-3} - T$).
4. The probability that the state from which the transition originates is filled multiplied by the probability that the terminal state is empty.

The only "new" issue in this list is item 4. In an atomic laser, the atom is in 2 or 1—period—there is no probability involved on a per atom basis.

Following the prescription given above, we obtain the following for the number of transitions $R_{1\to 2}$ per unit of volume caused by a wave with photon energy $h\nu$.

$$R_{1\to 2} = [B_{12}] \cdot [\rho(\nu)\, d\nu] \cdot [\rho_{\text{jnt}}(\nu)] \cdot [f_v(E_1)(1 - f_c(E_2))] \tag{11.4.13a}$$
$$\quad\; |\!\leftarrow\!1\!\rightarrow\!| \;\; |\!\leftarrow\!\!-2\!\!-\!\rightarrow\!| \;\; |\!\leftarrow\!3\!\rightarrow\!| \;\; |\!\leftarrow\!\!-\!\!-\!\!-\!\!-4\!\!-\!\!-\!\!-\!\!-\!\rightarrow\!|$$

and for the stimulated rate

$$R_{2\to 1} = [B_{21}] \cdot [\rho(\nu)\, d\nu] \cdot [\rho_{\text{jnt}}(\nu)] \cdot [f_c(E_2)(1 - f_v(E_1))] \tag{11.4.13b}$$
$$\quad\; |\!\leftarrow\!1\!\rightarrow\!| \;\; |\!\leftarrow\!\!-2\!\!-\!\rightarrow\!| \;\; |\!\leftarrow\!3\!\rightarrow\!| \;\; |\!\leftarrow\!\!-\!\!-\!\!-\!\!-4\!\!-\!\!-\!\!-\!\!-\!\rightarrow\!|$$

where the numbered factors correspond to the list above.

The net downward rate is the difference and is given by

$$R_{2\to 1} - R_{1\to 2} = [B_{21}] \cdot [\rho(\nu)\, d\nu] \cdot [\rho_{\text{jnt}}(\nu)] \left\{ \begin{array}{c} f_2(1 - f_1) - f_1(1 - f_2) \\ \text{or} \\ (f_2 - f_1) \end{array} \right\} \tag{11.4.14}$$

where $f_{2,1}$ is a convenient shorthand for $f_c(E_2)$ or $f_v(E_1)$, respectively. Equation (11.4.14) also used the fact that $B_{21} = B_{12}$ since each state has the same degeneracy (2 spin states).

The gain coefficient can be obtained directly from (11.4.14) by utilizing the word interpretation of the defining differential equation.

$$\gamma(\nu) \triangleq \frac{1}{I(\nu)} \cdot \frac{dI(\nu)}{dz} \triangleq \frac{dI(\nu)/dz}{I(\nu)} = \frac{\text{net power emitted per unit of volume}}{\text{power per unit of area} = I(\nu)}$$

The power emitted is just the energy per transition, $h\nu$, times the net downward transition rate given by (11.4.14). The optical power per unit of area causing these transitions is

$$I(\nu) = [\rho(\nu)\, d\nu] \cdot [v_g = c/n_g]$$

where v_g is the group velocity of the wave and n_g is the group index.* Thus

$$\gamma(\nu) = \frac{h\nu \cdot [R_{2\to 1} - R_{1\to 2}]}{[\rho(\nu)\, d\nu] \cdot c/n_g}$$

$$= B_{21} \cdot h\nu \cdot \frac{n_g}{c} \cdot [\rho_{\text{jnt}}(\nu)] \left\{ \begin{array}{c} f_2(1 - f_1) - f_1(1 - f_2) \\ \text{or} \\ f_2 - f_1 \end{array} \right\} \tag{11.4.15a}$$

This can be converted into a more familiar form by using the relationship between A_{21} and B_{21}

$$\frac{A_{21}}{B_{21}} = \frac{8\pi h\nu^3 n^2 n_g}{c^3}$$

*The group index is given by $n_g = n(\lambda) - \lambda(dn/d\lambda)$.

and the appropriate Fermi functions.

$$\gamma(\nu) = A_{21}[f_2(1 - f_1)]\frac{\lambda_0^2}{8\pi n^2}\rho_{\text{jnt}}(\nu)\left\{1 - \frac{f_1(1 - f_2)}{f_2(1 - f_1)}\right\} \tag{11.4.15b}$$

With a bit of work, we find that the last brace can be expressed in terms of the quasi-Fermi levels and the photon energy leading to

$$\boxed{\gamma(\nu) = A_{21}[f_2(1 - f_1)]\frac{\lambda_0^2}{8\pi n^2}\rho_{\text{jnt}}(\nu)\left\{1 - \exp\left[\frac{h\nu - (F_n - F_p)}{kT}\right]\right\}} \tag{11.4.15c}$$

If $F_n = F_p$—an equilibrium situation—then $\exp[h\nu/kT] > 1$, and the "gain" coefficient is negative signifying absorption.

We immediately see a familiar friend in the expression for the gain coefficient in terms of the Einstein A coefficient, $\lambda^2/8\pi n^2$, and a density difference. [Compare (11.4.15b) to (7.5.2).]

The format of (11.4.15b) presents a bit of a worry when f_2 or $1 - f_1 = 0$ (at $T = 0$ K). Another way of expressing the same gain expression is to use the bottom line of (11.4.15a) and allow the products of the Fermi functions to cancel.

$$\boxed{\gamma(\nu) = A_{21}\frac{\lambda_0^2}{8\pi n^2}\rho_{\text{jnt}}(\nu)[f_c(E_2) - f_v(E_1)]} \tag{11.4.15d}$$

Thus the products of the joint density of states and the difference in the Fermi functions replace the line shape function and the density difference of atomic lasers.

The format of (11.4.15d) indicates that the gain coefficient must be between two limits.

> If $f_c(E_2) = 0$ and $f_v(E_1) = 1$, then $\gamma(\nu) = -\alpha(\nu)$, the absorption coefficient without pumping. If $f_c(E_2) = 1$ and $f_v(E_1) = 0$, then $\gamma(\nu) = +\alpha(\nu)$, a completely inverted semiconductor. What was absorption becomes gain.

These factors were anticipated in Fig. 11.7.

Finally, recall that the joint density of states contains one part of the frequency dependence with the other given by the difference in the Fermi functions. Hence, we can lump many of the terms into an ever-present K and write

$$\gamma(\nu) = K(h\nu - E_g)^{1/2}[f_c(E_2) - f_v(E_1)] \tag{11.4.15e}$$

where

$$E_2 = E_c + \frac{m_h^*}{m_e^* + m_h^*}(h\nu - E_g)$$

$$E_1 = E_v - \frac{m_e^*}{m_e^* + m_h^*}(h\nu - E_g)$$

and the constant K can be determined from experiment thereby avoiding (once again) detailed quantum calculations of A_{21} or B_{21}.

An assumption of $T = 0$ makes the interpretation of (11.4.15e) particularly simple since the Fermi functions become 1 or 0. But whatever the temperature, the last bracket must be positive for optical gain. It is left for a problem to show that this occurs when

$$E_g < [h\nu = E_2 - E_1] < F_n - F_p \qquad \text{for gain}$$

a statement obvious from the form of (11.4.15c).

11.4.2 Spontaneous Emission Profile

In addition to relating the absorption to the gain coefficient, we can also relate it to the spectral distribution of the spontaneous emission resulting from recombination of the electrons and holes. To obtain this connection, we use an argument identical to that presented by Einstein in his explanation of the blackbody spectrum and covered in Sec. 7.3. Consider a semiconductor contained within a heated cavity, both at a temperature T and both in thermodynamic equilibrium. (Why we would ever do such an experiment is somewhat farfetched, but in any case the theory should handle the problem.) The semiconductor will absorb some of the blackbody radiation emitted by the atoms in the cavity, and then re-emit its own spontaneous profile because of electron-hole recombination. The key point is that the combination of absorption of the familiar blackbody radiation emitted by the atoms in the cavity wall and the production of new photons by electron-hole recombination must *still* reproduce the original blackbody spectrum. This is an example of the use of the principle of *detailed balancing*, a topic covered in more detail in Appendix II.

Thus we require a detailed balance between the generation of the carriers by the absorption of photons between ν and $\nu + d\nu$ from the ambient blackbody radiation and the radiative recombination rate $R(\nu)\,d\nu$, which emits photons into this same interval. At thermodynamic equilibrium this balancing must take place in every frequency interval $d\nu$, even though other, possibly more important, processes are taking place simultaneously. The rate of absorption of photons in this frequency interval is the spatial coefficient $\alpha(\nu)$, times the group velocity c/n_g, times the energy density at the frequency ν in the interval $d\nu$. At thermodynamic equilibrium, the rate of recombination of electron-hole pairs yielding a photon between ν and $\nu + d\nu$ must be *exactly* equal to the absorption rate of photons in this same interval from the ambient (blackbody) background radiation. Hence

$$\begin{aligned} h\nu R(\nu)\,d(\nu) &= \text{energy per unit of volume radiated spontaneously by recombination} \\ &\equiv \text{energy per unit of volume absorbed from the blackbody spectrum} \\ &= \alpha(\nu)\left\{\left[v_g = \frac{c}{n_g}\right]\cdot\left[\frac{8\pi\nu^2 n^2 n_g}{c^3}\cdot\frac{(h\nu)\,d\nu}{\exp[h\nu/kT]-1}\right]\right\} \end{aligned} \tag{11.4.16}$$

where the second bracket is the energy density of the blackbody spectrum given by (7.2.10), which is converted to intensity (watts/area) by multiplying by the group velocity, which is, the first bracket. The attenuation coefficient $\alpha(h\nu)$ is just the negative of the gain coefficient, (11.4.15e), evaluated under the conditions specified by our "Gedanken experiment." The

thermodynamic environment (LTE) specifies that $F_n = F_p = E_f$ and thus $\alpha(h\nu)$ becomes

$$\alpha(\nu)\Big|_{\text{LTE}} = [-\gamma(\nu)]\Big|_{F_n=F_p} = [A_{21}f_2(1-f_1)]\cdot\frac{\lambda_0^2}{8\pi n^2}\rho_{\text{jnt}}(\nu)\left\{\exp[h\nu/kT)-1\right\}$$

Use this expression in (11.4.16) to find

$$\boxed{R(\nu) = A_{21}[f_2(1-f_1)]\rho_{\text{jnt}}(\nu)} \tag{11.4.17}$$

While the thermodynamic environment has been used to derive (11.4.17), it is a perfectly general expression for the spontaneous recombination spectrum in any case. Quite often, this expression is abbreviated by an obvious substitution

$$R(\nu) \triangleq \frac{n_e}{\tau_r}g(\nu) \tag{11.4.18}$$

where N_e is the density of electrons in the conduction band, τ_r^{-1} is the spontaneous recombination rate, and $g(\nu)$ is the line shape for spontaneous emission. With this abbreviation, the semiconductor laser gain coefficient becomes

$$\gamma(\nu) = \frac{\lambda_0^2}{8\pi n^2}R(\nu)\left\{1-\exp\left[\frac{h\nu-(F_n-F_p)}{kT}\right]\right\} \tag{11.4.19a}$$

$$\gamma(\nu) = \frac{1}{\tau_r}\frac{\lambda_0^2}{8\pi n^2}n_e g(\nu)\left\{1-\exp\left[\frac{h\nu-(F_n-F_p)}{kT}\right]\right\} \tag{11.4.19b}$$

With this substitution and abbreviation the connection with atomic laser theory is complete.

Fig. 11.9 illustrates the wavelength dependence of the spontaneous emission from a p-type GaAs sample that was computed from the measured absorption coefficient.

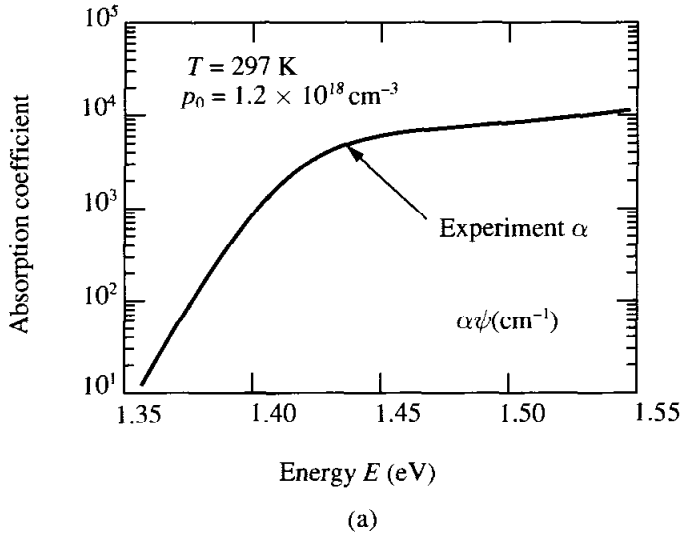
11.4.3 An Example of an Inverted Semiconductor

It is appropriate to take a break in the theoretical development and compute the density of electrons and holes required to obtain an inverted population at a finite temperature. For this example we will choose intrinsic GaAs and ignore the light holes. Because the whole question of an inverted population revolves around the position of the Fermi levels, we compute the density of carriers in each band as a function of these levels. This computation will enable us to estimate the injection current required to maintain this population of electron-hole pairs against recombination. The density of electrons in the conduction band is the integral of the density of states times the probability that the state will be filled; that is, the Fermi function.

$$n = \frac{1}{2\pi^2}\left[\frac{2m_e^*}{\hbar^2}\right]^{3/2}\int_{E_c}^{\infty}\frac{(\epsilon-E_c)^{1/2}\,d\epsilon}{\exp[(\epsilon-F_n)/kT]+1} \tag{11.4.20a}$$

The number of holes is a corresponding integral over empty states in the valence band:

$$p = \frac{1}{2\pi^2}\left[\frac{2m_h^*}{\hbar^2}\right]^{3/2}\int_{0}^{E_v}\frac{(E_v-\epsilon)^{1/2}\,d\epsilon}{\exp[(F_p-\epsilon)/kT]+1} \tag{11.4.20b}$$



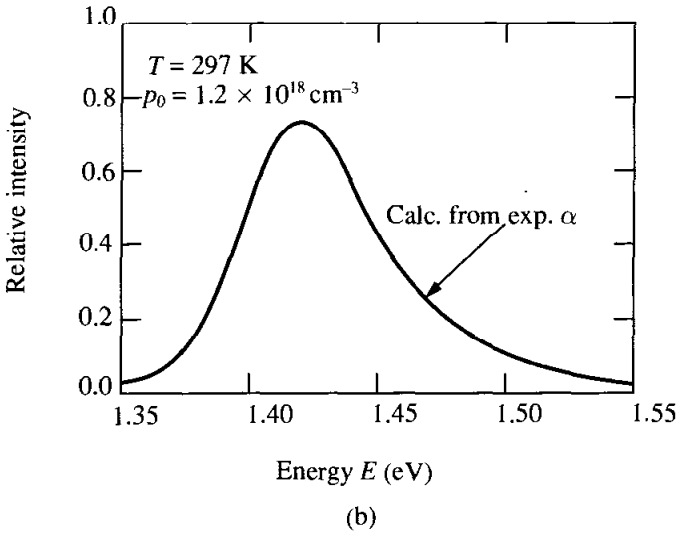


FIGURE 11.9. (a) Measured absorption coefficient in GaAs doped with an acceptor concentration of 1.2×10^{18} cm^3. (b) Spontaneous emission profile calculated from the absorption. (Data from Casey and Stern [19].)

If we make the substitution $\epsilon = E_c + u$ in (11.4.20a) and $\epsilon = E_v - u$ in (11.4.20b) and identify

$$x = u/kT \qquad a = \exp\{[F_p - E_v]/kT\} \qquad b = \exp\{[E_c - F_n]/kT\}$$

Then (11.4.20a) and (11.4.20b) become

$$n = \frac{1}{2\pi^2}\left[\frac{2m_e^* kT}{\hbar^2}\right]^{3/2} \cdot \left[\exp - \left(\frac{E_c - F_n}{kT}\right)\right] \int_0^\infty \frac{x^{1/2}\,dx}{e^x + 1/b} \quad (11.4.20c)$$

$$p = \frac{1}{2\pi^2}\left[\frac{2m_h^* kT}{\hbar^2}\right]^{3/2} \cdot \left[\exp - \left(\frac{F_p - E_v}{kT}\right)\right] \int_0^\infty \frac{x^{1/2}\,dx}{e^x + 1/a} \quad (11.4.20d)$$

Now in conventional semiconductor electronics (transistors, diodes, FETs, etc), the Fermi levels are in the gap and thus factors (a or b)$\gg$1. Hence the integrals in (11.4.20c) and

(11.4.20d) can be evaluated in closed form.

$$I = \int_o^\infty \frac{x^{1/2}\,dx}{e^x + 1/(a,b)} \rightarrow \int_0^\infty x^{1/2}e^{-x}\,dx = \frac{\sqrt{\pi}}{2} \qquad [\text{for } (a,b) \gg 1] \quad (11.4.21)$$

If the semiconductor were in equilibrium so that $F_n = F_p = E_f$, then the product of (11.4.20c) and (11.4.20d) yields a constant that depends only on the gap, $E_c - E_v = E_g$, the temperature, and the effective masses in the bands, but not on the position of the Fermi level. That constant is, of course, the square of the intrinsic carrier density and the product yields the familiar relation $n \cdot p = n_i^2$.

A semiconductor laser is not a system in equilibrium, and furthermore, the replacement of the denominators of (11.4.20c) and (11.4.20d) by the exponential is not valid because at least one of the quasi-Fermi levels must be *in* a band for an inversion and thus (a, b) may be less than 1. Thus (11.4.20) is multiplied and divided by $\pi^{1/2}/2$, and the integral evaluated numerically:

$$I_2 = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{x^{1/2}\,dx}{e^x + 1/(a,b)} \quad (11.4.22)$$

Values for the integral and the electron hole populations in an intrinsic semiconductor are given in Table 11.2. The evaluation of n, p used $m_e^* = 0.067\ m_0$, $m_h^* = 0.55\ m_0$, and $kT/e = 0.0259$ V in (11.4.20). Note that the "sacred" relationship $n \cdot p = n_i^2$ does *not* hold when the Fermi levels are within the bands or when the system is not in equilibrium.

Values for the electron and hole densities for the various positions of the Fermi levels are also given in Table 11.2 in addition to the evaluation of the integral. There are important points to be gained from the study of those numerical values.

TABLE 11.2 Carrier Densities in an Intrinsic Semiconductor as a Function of the Position of the Quasi-Fermi Levels

(a, b)	$(E_c - F_n)/kT$ $(F_p - E_v)/kT$	I_2	n (cm^{-3})	p (cm^{-3})	Comment
10^3	6.91	1.000	4.36^{14}	1.02^{16}	
10^2	4.61	0.996	4.34^{15}	1.02^{17}	
50	3.91	0.993	8.66^{15}	2.04^{17}	
20	3.00	0.983	2.14^{16}	5.04^{17}	
10	2.30	0.967	4.22^{16}	9.92^{17}	
7.75	2.05	0.957	5.40^{16}	$\mathbf{1.27^{18}}$	Fermi
5	1.61	0.936	8.16^{16}	1.92^{18}	levels
2	0.69	0.860	1.87^{17}	4.41^{18}	in gap ↑
1	0.00	0.765	3.34^{17}	7.84^{18}	—
0.5	−0.69	0.641	5.59^{17}	1.31^{19}	Fermi ↓
0.2	−1.61	0.457	9.96^{17}	2.34^{19}	levels
0.129	−2.05	0.373	$\mathbf{1.27^{18}}$	2.99^{19}	within
0.1	−2.30	0.329	1.43^{18}	3.37^{19}	bands

1. It only takes about 3.34×10^{17} electron/cm^3 to move the quasi-Fermi level for electrons to the edge of the conduction band, but it takes 23.5 times as many holes to make $F_p = E_v$. Hence it is advantageous to create (or inject) electrons into a region that is heavily doped p type. (However, if a material is so heavily doped, the impurity levels merge with the bands and the gap is no longer a sharp cutoff for allowed states. Furthermore, the k selection rule is also relaxed. We leave those considerations for more advanced texts.)
2. Note that for equal electron and hole densities of 1.27×10^{18} cm^{-3}, the quasi-Fermi level for the electrons is inside the conduction band, $(E_c - F_n)/kT = -2.05$, whereas the quasi-Fermi level for the holes is still in the forbidden gap, $(F_p - E_v)/kT = 2.05$. However, the difference $F_n - F_p$ *equals* the band gap.

If we were to pump harder and create more electron-hole pairs, then we would obtain net optical gain.

This critical density can be used to estimate the pumping rate required to reach threshold by some sort of a pumping scheme, say by optical pumping with another laser or flash lamp, E-beam excitation, or carrier injection. The excess electrons (or holes) created by this pumping scheme recombine according to the following kinetics:

$$\frac{dn}{dt} = -\beta \cdot n \cdot p + G \tag{11.4.23}$$

where β is the recombination rate and is equal to 2×10^{-10} cm^3/ sec (by experiment) for GaAs and G is the generation (or pumping) rate to be discussed later. To sustain a steady state electron-hole pair density of $n = p = 1.27 \times 10^{18}$ cm^{-3}, we require a generation (or pumping) rate of

$$\begin{aligned} G = \beta \cdot n \cdot p &= 2 \times 10^{-10} \cdot 1.27 \times 10^{18} \cdot 1.27 \times 10^{18} \\ &= 3.23 \times 10^{26} \frac{e - h \text{ pairs}}{\text{cm}^3 - \text{sec}} \end{aligned} \tag{11.4.24}$$

If these carriers are supplied by current injection and if they recombine in a width $d = 1\ \mu$m, then the current density required (in amp/cm^2) is

$$\begin{aligned} J = ed \cdot \left[\frac{dn}{dt}\right]_{\text{recomb}} &= 1.6 \times 10^{-19} \cdot 1 \times 10^{-4} \cdot 3.23 \times 10^{26} \\ &= 5.17 \text{ kA/cm}^2 \end{aligned} \tag{11.4.25}$$

This is a threshold type of calculation based on the numerics of Table 11.2. If the current is bigger than this value, then we can anticipate gain and oscillation with proper feedback. As indicated previously, band-tailing effects from heavy doping introduce modifications, but the numbers are typical.

It should also be obvious from (11.4.25) that there is considerable virtue in reducing the width of the recombination region. For instance, if the width over which recombination occurred were reduced to 100 Å (0.01 μm), then the current would only be 51.7 amp/cm^2. If the plane of the junction were 10 μm wide and 500 μm long (see Fig. 11.11), then the

circuit current would only be 2.6 mA, small enough for most flashlight batteries. This is one of the major advantages of heterostructures (to be discussed in the next section).

The densities in Table 11.2 are a very strong function of lattice temperature varying as

$$T^{3/2} \exp -[\Delta E/kT]$$

For instance, let $T = 77$ K (liquid N_2 temperatures). Then the densities required to obtain $(F_n - E_c)/kT = -2.05$ and $(F_p - E_v)/kT = +2.05$, that is, $(F_n - F_p)/kT = E_g/kT$), would be reduced to $n = 1.27 \times 10^{18}\ (77/300)^{3/2} = 1.65 \times 10^{17}\ \text{cm}^{-3}$; and the threshold current density of (11.4.25) becomes (assuming β is a constant, which it is not) $J = 87.3\ \text{amp/cm}^2$ (at 77 K). The exponential factor is the controlling one, and thus many researchers report the temperature dependence of the threshold current as

$$J = J_0 \exp\left[\frac{T - T_0}{T_0}\right] \tag{11.4.26}$$

where a good laser device has T_0 of 400 K (i.e., 100 C) or so.

11.5 DIODE LASER

Recall that the necessary condition for optical gain as given by (11.4.1) is

$$F_n - F_p > h\nu > E_g \tag{11.4.1}$$

Our task is to realize this condition by the injection of carriers into the region where they recombine and generate photons.

11.5.1 Homojunction Laser

Figure 11.10 illustrates the geometry employed in the earliest type of semiconductor lasers. It used heavily doped p and n regions on either side of a junction leading to the energy level diagram along the $(y = 0, z)$ plane as shown below the sketch of the metallurgical junction. As is the case for every equilibrium p-n junction, the Fermi level is constant throughout the device with no current flowing.

When the junction is biased in the forward direction, as in Fig. 11.10(c), the Fermi level splits because of the injection of minority carriers (electrons into the p region, holes into the n region) and there exists a region near the junction where there is *simultaneously* a high density of electrons and a high density of holes. Because of the much higher mobility of electrons compared to that of holes, most of the injection is by electrons into the p region. These electrons recombine with the majority holes after diffusing a distance d which is more or less the diffusion length $d = L_n = (D\tau)^{1/2}$. If we take GaAs as our example, some typical parameters pertinent for heavily doped materials are

$$\mu_n = 600\ \text{cm}^2/\text{V} - \text{sec} \qquad D_n = 15.15\ \text{cm}^2/\text{sec}$$
$$\mu_p = 30\ \text{cm}^2/\text{V} - \text{sec} \qquad D_p = 0.78\ \text{cm}^2/\text{sec} \tag{11.5.1}$$

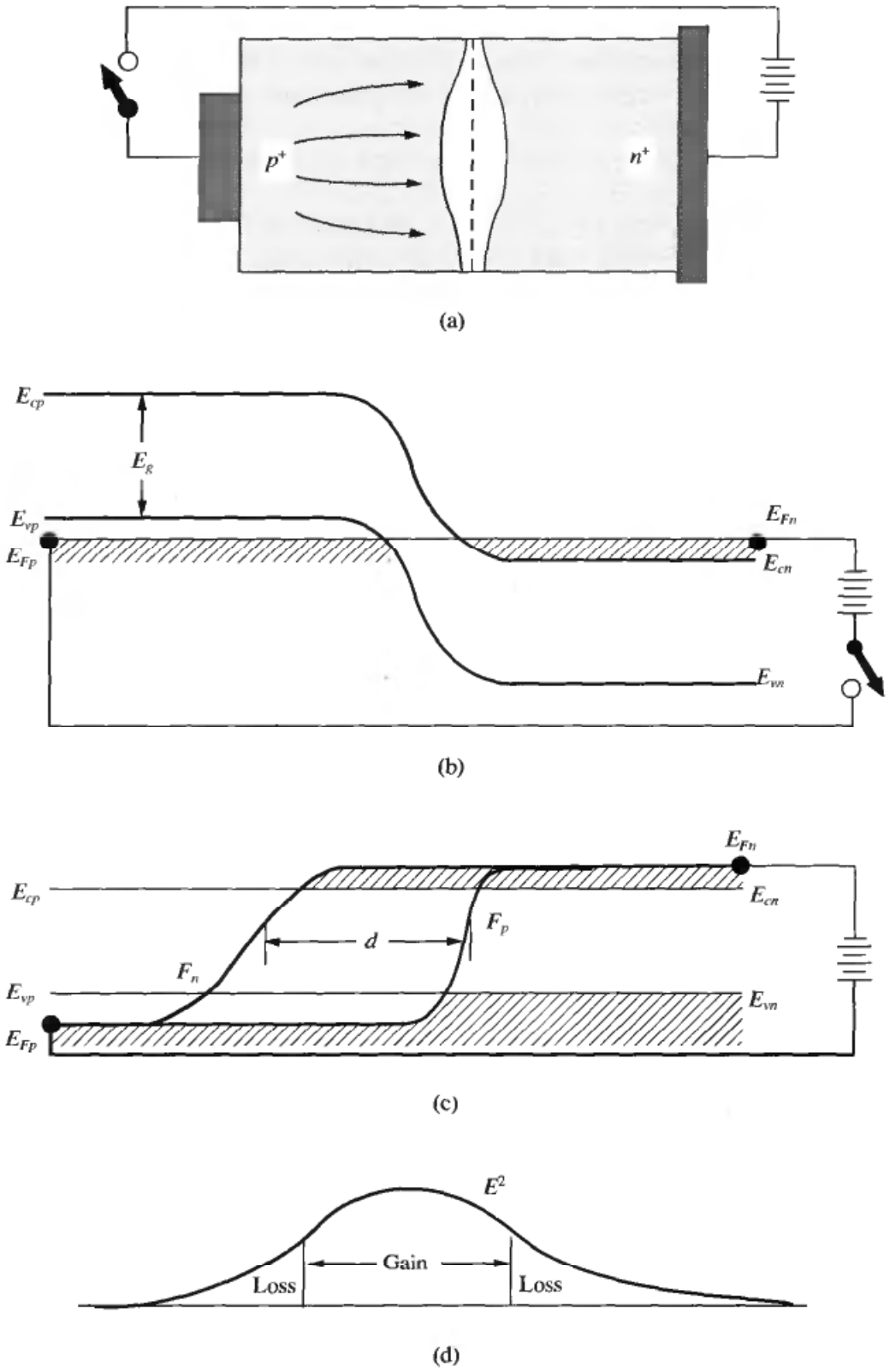


FIGURE 11.10. The homojunction laser: (a) shows a cross section of the junction with the bowed area being due to current spreading; (b) and (c) show the band diagram in equilibrium and with injected current; (d) illustrates the electromagnetic mode experiencing gain and loss.

where the Einstein relation $D/\mu = kT/e$ has been applied.

The minority carriers, the electrons, recombine with the "sea of holes" made available through doping, according to the following kinetics:

$$\frac{dn_p}{dt} = -(\beta p_p) \cdot n_p = -\frac{n_p}{\tau_r} \tag{11.5.2}$$

The numbers of Sec. 11.4.3 (Table 11.2) indicate that the acceptor doping density of $\approx 9 \times 10^{18}$ cm^{-3} would make the p side degenerate (E_f in the valence band). Thus, for $\beta = 2 \times 10^{-10}$ cm^3/ sec, we obtain a recombination lifetime for the electron injected into the p side of

$$\frac{1}{n_p}\frac{dn_p}{dt} = -\frac{1}{\tau_r} = \beta p_p = -1.8 \times 10^9 \text{ sec}^{-1}$$

or

$$\tau_r = 0.56 \text{ ns}$$

The diffusion length and thus the cross sectional distance d over which gain is possible can be estimated to be

$$d \sim L_n = (D_n \tau_n)^{1/2} = 0.93 \ \mu\text{m}$$

We can make a rough estimate of the current density (amp/cm^2) necessary to create this population inversion by a few simple calculations. Let us assume that the quasi-Fermi level for electrons is 0.050 eV above E_c and that we can treat their distribution in energy as if $T = 0$ K (to make the arithmetic easy). Thus

$$n = \frac{1}{3\pi^2} \cdot \left\{ \frac{2m_e^* \Delta E}{h^2} \right\}^{3/2} = 8.79 \times 10^{23} \text{ m}^{-3} = 8.79 \times 10^{17} \text{ cm}^{-3}$$

Now these carriers are going to recombine in a distance of roughly 0.93 μm at a rate of 1.8×10^9/ sec (i.e., $1/\tau_r$). Hence the current density required to maintain this nonequilibrium condition is

$$J = \frac{ned}{\tau_r} = 23.5 \text{ kA/cm}^2 \tag{11.5.3}$$

That is a fair amount of current and is typical of homojunction lasers. [It differs from (11.4.25) because of the assumptions, but it is close.] There are a few more issues to be noted about this homojunction laser.

1. The distance d over which carrier inversion is achieved (i.e., $F_n - F_p > h\nu$) is *not* a controlled parameter, and there is very little "design" involved.
2. The distance d is on the order of the wavelength being amplified. Although it is not appropriate now to become involved in the electromagnetics of the laser mode, it is reasonable to recognize that the mode might extend over even a larger distance as shown in Fig. 11.10(d). Shown here is the case where the central part of the electromagnetic mode experiences gain, whereas the edges experience loss. It is a nontrivial task to address this problem and will be discussed in Chapter 13 on advanced electromagnetics.

3. Finally, we can guess that the spatial extent of the field in the y direction, which is controlled by the width of the contact ($\sim$ 10 μm), is considerably larger than in the x direction, which in turn is controlled by the diffusion length. This fact has a profound effect on the free-space radiation pattern of the laser. As shown in Fig. 11.11, the small width in the direction perpendicular to the junction implies a much larger radiation angle $\theta_\perp$ compared to that measured in the plane of the junction. We do not obtain a nice cylindrically symmetric $TEM_{0,0}$ mode from an injection laser.

Comments 1 and 2 are significantly modified for heterojunction lasers (covered in the next section), but much of comment 3 applies to all semiconductor lasers. It may be the one instance in which a semiconductor laser is at a disadvantage compared to the larger gas or solid-state laser.

11.5.2 Heterojunction Lasers

There are many problems associated with the simple p-n junction lasers described in the previous section, which can be attributed to the fact that we are using the same material (i.e., GaAs) for both the p and n region. Two of the critical ones are

1. The injected minority carriers are "free" to diffuse where they will, a fact that dilutes the spatial distribution of recombination and thus the gain.
2. There is very little guiding and confinement of the electromagnetic wave being amplified. There is a small bit of wave guiding caused by the slight decrease of refractive index on the n side (due to the free electrons) and on the p side due to the small

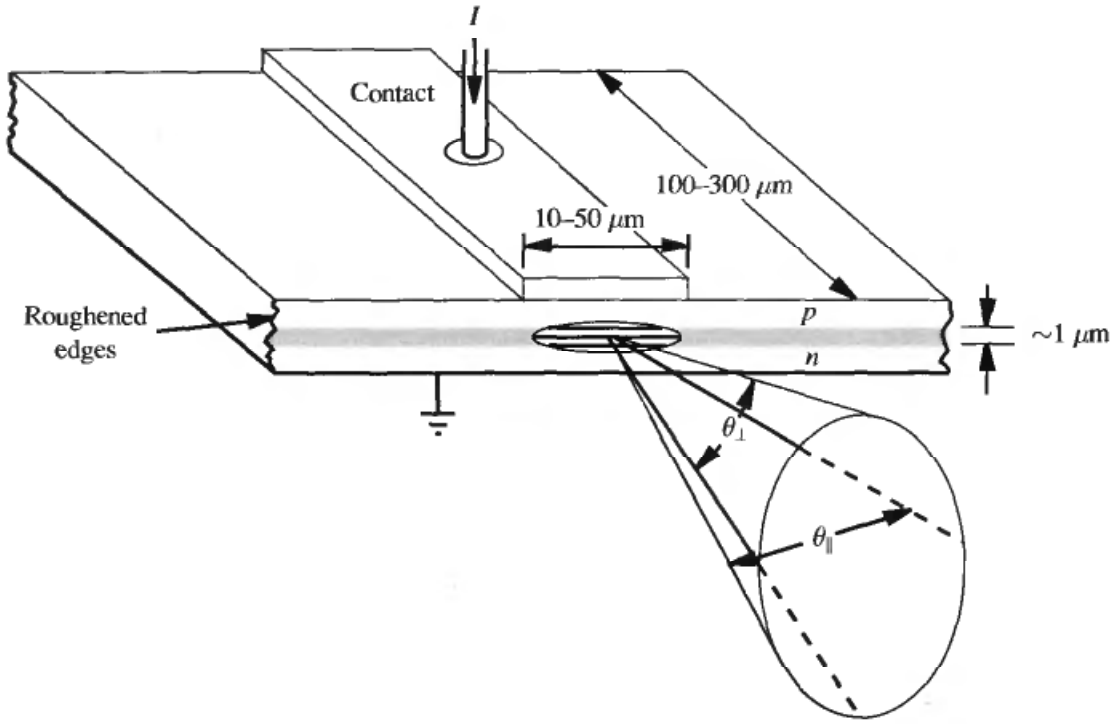


FIGURE 11.11. The radiation field of a semiconductor laser.

change in E_g with acceptor doping. However, these changes are very small, and we face the unpleasant fact that the central part of the wave may be amplified with the tails, which extend into the noninverted regions, being attenuated.

Both of these problems can be ameliorated by the use of *heterostructures* to form the active portion of the laser. These are junctions between two dissimilar materials such as GaAs with $Al_xGa_{1-x}As$, with x being the fraction of gallium being replaced by aluminum. It is a fortuitous fact of nature that GaAs and AlAs semiconductors have almost identical lattice constants (see Table 11.1) and thus can be mixed and can be grown on top of each other with little strain involved and a very small density of traps at the interface.* This metallurgical fact is critical to the success of *making* the junction.

Two other physical factors play a critical role. As the percentage of aluminum is increased (x ↑), the band gap *increases* and the index of refraction goes *down*, and this asynchronous behavior is true for quaternary alloy combinations also. This fact is truly God's gift to the semiconductor laser field, for it greatly alleviates both of the above problems.

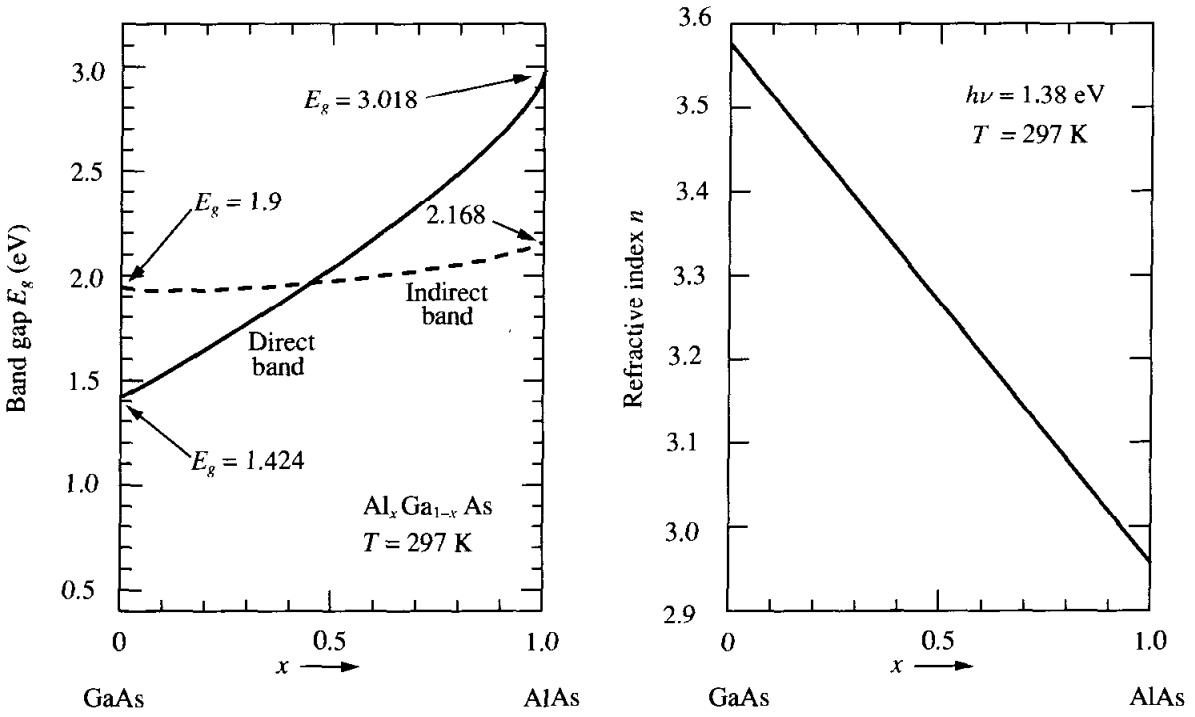


FIGURE 11.12. Dependence of the bandgap and index of refraction on the fraction, x, of aluminum in the composition. These graphs are plots of the analytic expressions for E_g and the index given by reference [22].

$$E_g(\text{direct}) = 1.424 + 1.247x + [1.147(x - 0.45)^2]u(x - 0.45)$$

$$E_g(\text{indirect}) = 1.900 + 0.125x + 0.143x^2$$

$$n(x) = 3.590 - 0.710x + 0.091x^2$$

*The GaAs-$Al_xGa_{1-x}As$ system is the best known case involving three different atoms and is the consequence of the nearly equal sizes of Al and Ga. By substituting a bigger and a smaller one from the Col. III and V elements in Table 11.1, other quaternary lattice matches can be achieved such as (GaIn) (AsP) and (GaIn)(AsSb).

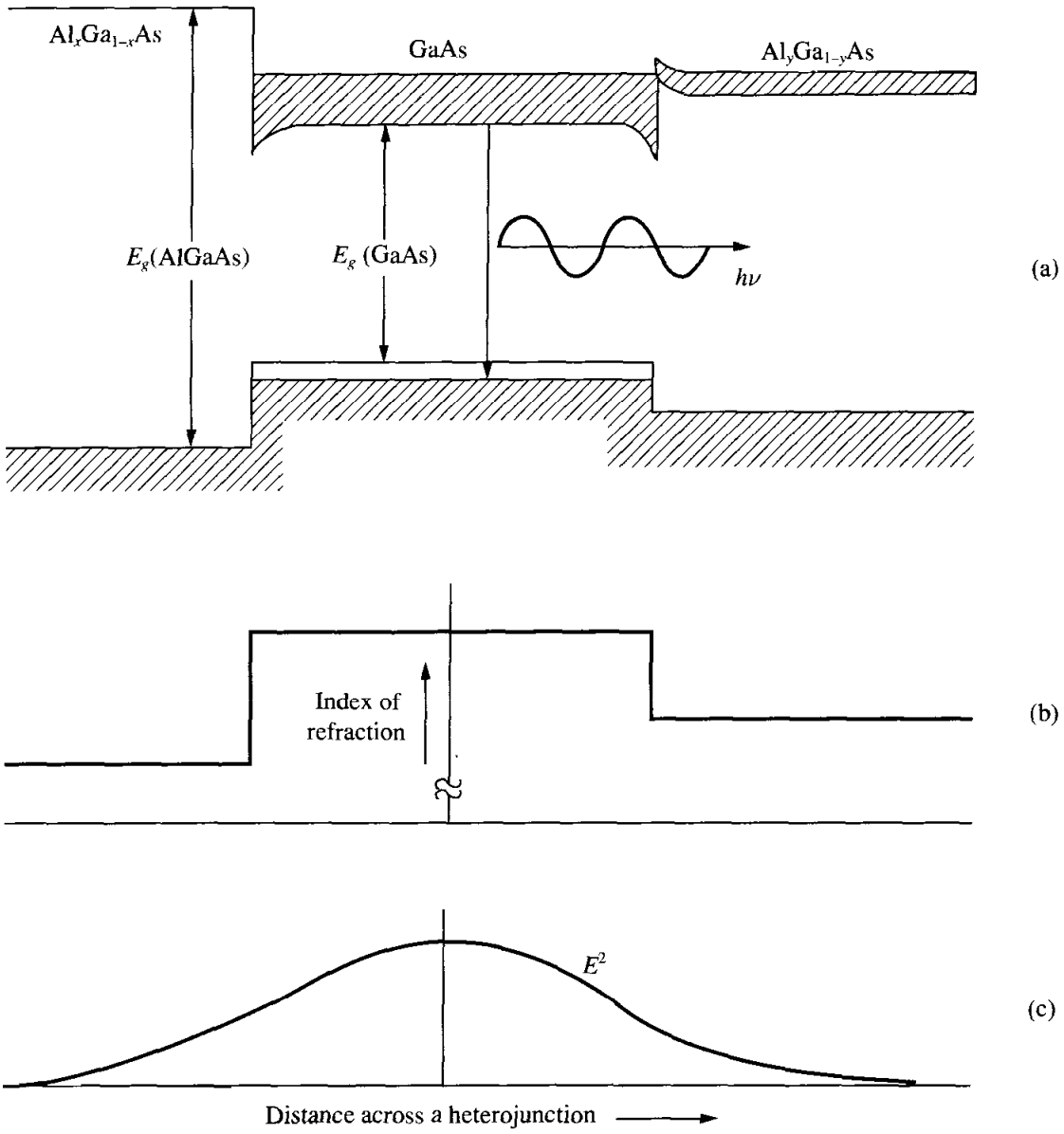


FIGURE 11.13. (a) The band diagram for a forward-biased heterostructure, (b) the refractive index, and (c) a sketch of the light intensity in the vicinity of the active region.

Fig. 11.12 illustrates the dependence of the band gap and the index of refraction on the mole fraction of Al substituted for gallium.

Fig. 11.13 illustrates a laser using a double heterostructure geometry and is shown biased in the forward direction. Shown also is the variation of the refraction index and the light intensity in a plane perpendicular to the junction. Note that the electrons injected from the n-type $Al_xGa_{1-x}As$ material are confined to and recombine in lower band-gap p-type GaAs. Furthermore, note that now there is a well-defined dielectric slab waveguide formed similar to that analyzed in Chapter 4. This causes the electromagnetic intensity to be maximized in the same region as the inversion, a condition that also maximizes the stimulated emission rate. In addition, the photon energy of the "tails" of the electromagnetic

mode is *less* than the band gap of the confining layers. Hence, most of the built-in absorption problems have also disappeared. All of the above effects have caused a dramatic reduction in threshold currents and have caused a virtual explosion in the application of semiconductor lasers. For low power applications, such as communication and control, there is no other serious competitor.

Before we leave this section, we should point out that the construction of the energy level diagram of Fig. 11.13 involves considerable semiconductor physics. For instance, how much of the band discontinuity $E_g(Al_xGa_{1-x}As) - E_g(GaAs)$ should be apportioned to the conduction band and how much to the valence band? This is determined by the requirement that the vacuum level be *continuous* and *parallel* to the conduction and valence bands. After due allowance for the electron affinities, this implies 65% of ΔE_g being assigned to the conduction band discontinuity and 35% to the valence band.* To ensure continuity of the electric displacement flux and to ensure a constant Fermi level in equilibrium, one obtains the "spikes" at the interfaces (similar to the "spikes" found at a Schottky junction). The details of this are best left to texts on semiconductors.

11.6 QUANTUM SIZE EFFECTS

11.6.1 Infinite Barriers

The ability to grow very thin layers ($L \approx 40 - 100$ Å) of semiconductor materials has brought about a revolution in many phases of electronics and, in particular, in lasers incorporating these layers. The distances are so small, comparable to the deBroglie wavelength, that the *q*uantum *s*ize *e*ffects (QSE) become easily observable to us living in the macroscopic world. The ability to grow many layers of such different semiconductors on top of each other implies the ability to create artificial materials with a tailoring of characteristics. Indeed it is not clear where the revolution will stop—if ever; it appears to be limited only to the imagination and ingenuity of the researcher.

One of the major considerations in a semiconductor laser theory is the number of carriers occupying the available states in the various bands. This is determined by the pumping rate, the Fermi functions, *and* the density of states. In our initial walk through of semiconductor lasers, we assumed, implicitly, that all dimensions were huge compared to the deBroglie wavelength and thus were able to treat the allowed momentum states of the electrons (and holes) as continuous variables. This led to the density of states as a function of k being $\rho(k) = k^2 dk/\pi^2$ or, expressed in terms of energy (measured from the band edge), as

$$\rho(E)\,dE = \frac{1}{2\pi^2} \cdot \left\{ \frac{2m^*}{\hbar^2} \right\}^{3/2} E^{1/2} dE \tag{11.2.8}$$

When quantum size effects are important, as they are for these ultrathin layers, the density of states changes to reflect the quantization of momentum perpendicular to the thin layer.

*This has been and will continue to be debated in the literature.

As before, the allowed projections of the **k** vector are quantized† according to

$$k_x = \frac{m\pi}{L_x} \qquad k_y = \frac{p\pi}{L_y} \qquad k_z = \frac{q\pi}{L_z} \tag{11.2.2}$$

If we assumed one of these dimensions, say L_z, is much smaller than the other two, then the allowed quantum states are distributed in **k** space in the manner shown in Fig. 11.14.

Only the states in Fig. 11.14 indicated by the heavy dots with nonzero quantum (or "mode") numbers are allowed. It is apparent from this figure that the density of states in the (k_x, k_y) plane is much larger than the density in the k_z direction (where only one state is shown). As before, each state occupies a *volume in k space* of

$$V_k = \left(\frac{\pi}{L_x}\right) \cdot \left(\frac{\pi}{L_y}\right) \cdot \left(\frac{\pi}{L_z}\right) \tag{11.2.4}$$

but now we need to count the number of modes included as k increases by dk, keeping k_z equal to constant π/L_z (for the $q = 1$ mode). Because of the much larger density of points in the k_x, k_y plane, we treat those mode numbers as continuous variables as before and evaluate the number of allowed states in the interval $k_\parallel$ to $k_\parallel + dk_\parallel$ (i.e., motion parallel to

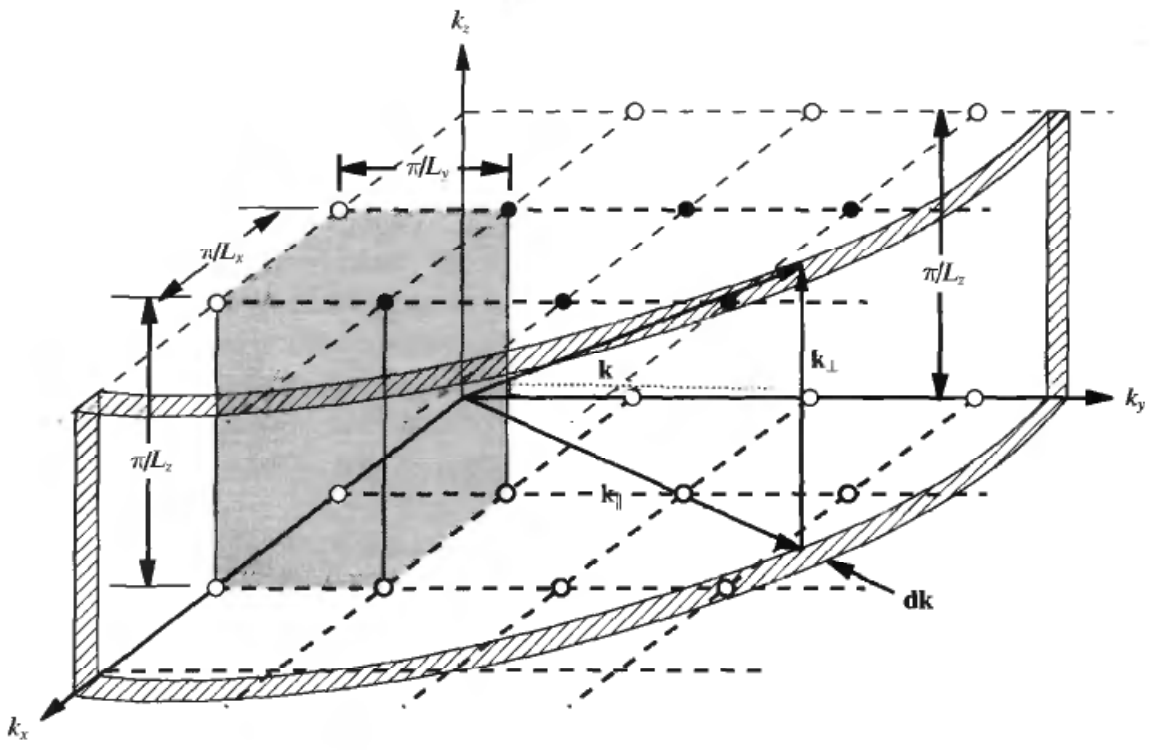


FIGURE 11.14. Allowed momentum vectors in a "thin" (i.e., $L_z < 200$ Å) semiconductor. The solid dots represent allowed states.

†The presence of confining layers modifies the expression for k_z and is addressed later. See the standard quantum problem of a "particle in a box." However, we will keep our simple quantization to illustrate the effect.

the thin surface) in the area dA given by:

$$dA = \frac{2\pi k_{\parallel} dk_{\parallel}}{4} \tag{11.6.1}$$

We first compute the volume of the thin surface on the "quarter-round" of Fig. 11.4, which is equal to the area given by (11.6.1) multiplied by the height π/L_z and also by 2 for the two spin states. Then we divide this cylindrical volume by that occupied by one state (the shaded volume $= V_k = \pi^3/L_xL_yL_x$) to obtain the number of states between k and $k + dk$.

$$N_k dk = \left(\frac{2\pi k_{\parallel} dk_{\parallel}}{4} = dA\right)\left(\frac{\pi}{L_z} = \text{height}\right)(2 \text{ spins})\left(\frac{L_xL_yL_z}{\pi^3}\right)$$

or

$$N_k dk = (L_xL_yL_z)\left[\frac{1}{\pi^2} k_{\parallel} dk_{\parallel}\right]\left[\frac{\pi}{L_z}\right] \tag{11.6.2}$$

This is fine as it stands, but we prefer to refer to the total vector $\mathbf{k}$ not $k_{\parallel}$: however, they are related by

$$k^2 = \left(\frac{\pi}{L_z}\right)^2 + k_{\parallel}^2 \tag{11.6.3a}$$

and thus

$$2k dk = 2k_{\parallel} dk_{\parallel} \tag{11.6.3b}$$

The mode density (per unit of k) for $k_z = \pi/L_z$ is

$$\frac{N_k dk}{(L_xL_yL_z)} = \rho_k dk = \frac{1}{\pi^2} k dk\left(\frac{\pi}{L_z}\right) \tag{11.6.4}$$

Now we convert to energy as before by

$$E = \frac{(\hbar k)^2}{2m^*} \quad \text{with } k > \pi/L_z \text{ and } E > E_1 = \frac{[\hbar(\pi/L_z)]^2}{2m^*} \text{ for } q = 1 \tag{11.6.5}$$

The density of states in the energy interval dE is

$$\rho(E)\, dE = \frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)\left(\frac{\pi}{L_z}\right) dE \qquad \text{provided} \quad E > E_1 \tag{11.6.6}$$

Thus the density of states is a *constant* independent of energy provided E is larger than the first allowed state E_1, which in turn must be larger than the normal band edge of the semiconductor. Hence by choosing the dimension L_z, we can "design" the energy state and thus "engineer" the band gap. A simple manipulation of the formulas for the density of states indicates that this *constant* value given by (11.6.6) is contained within the usual value given by (11.2.8). This is shown in Fig. 11.15.

Now, of course, if we pick the energy high enough, we can allow k to have a projection along k_z equal to $2\pi/L_z$. This merely requires repeating the prior development with identical

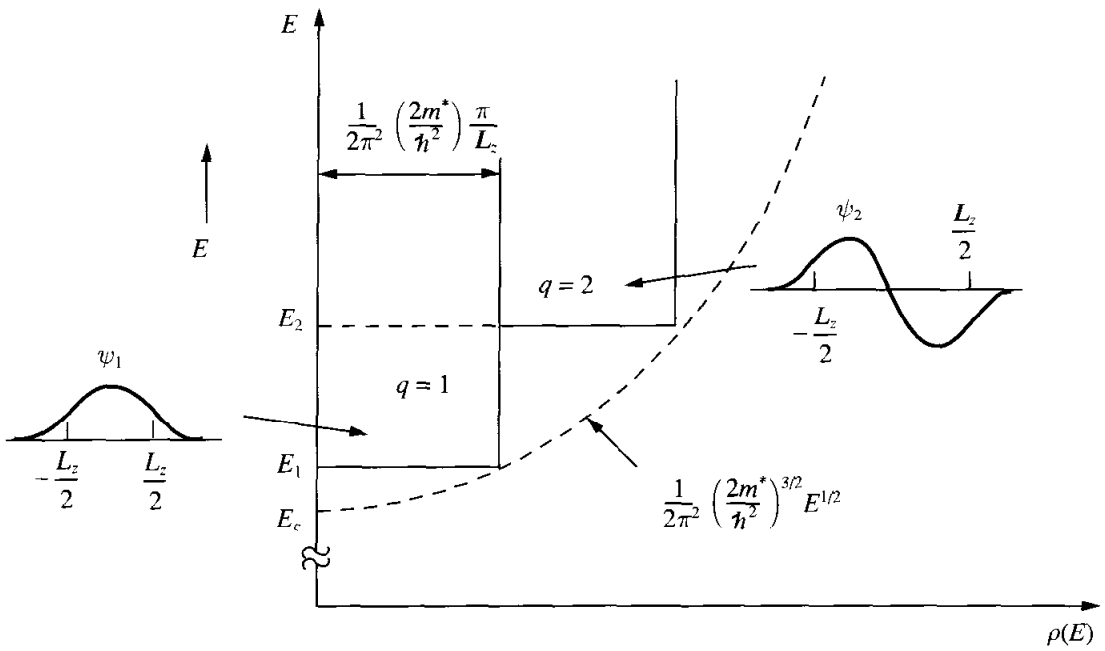


FIGURE 11.15. Density of states in a quantum well of thickness L_z. The lighter dashed curve is the normal density of states given by (11.2.8). The sketch indicates dependence of the wave function along z for the two subbands shown.

results except that it defines another subband that starts at $E = E_2 = \hbar^2(2\pi/L_z)^2/2m^*$ and proceeds upward at a constant value.

There are a couple of artistic issues added to Fig. 11.15 for clarity and to help understand the origin of the quantized states. For the first allowed band the ($q = 1$) wave function is just that of a simple standing ($\lambda/2$) wave, more or less going to zero at the boundaries. For $q = 2$ and the second mode, the wave function is antisymmetric and corresponds to fitting a full wavelength of the electron wave function between the boundaries. The constant density of states above the respective energies E_1 and E_2 merely reflects the idea that $k_\parallel$ and thus the momentum parallel to the interface can be just about anything provided L_x, L_y are sufficiently large. Note also that this stair-step-like density of states is contained within the density of states for a large semiconductor. This has an immediate consequence for a laser.

The gain coefficient of *any* laser, gas, solid state, or semiconductor, is always proportional to the inversion multiplied by how that inversion is smeared out among the various energy states (i.e., the line shape function). By restricting the number of states as a function of energy until $E > E_1$, a larger number of electrons can have the same energy within the band; and thus these inverted electrons can be stimulated much more effectively by an electromagnetic wave.

We should realize that the above development is incomplete. We cannot hold a 100-Å wafer with a tweezer and expect it to survive. It has to be grown on another substrate with still another layer grown on top of it. The beginning and terminating layers affect the wave functions of the electrons in the confining structure somewhat. Even so, there is not any law against repeating the growth procedure with some rather fantastic and complex structures

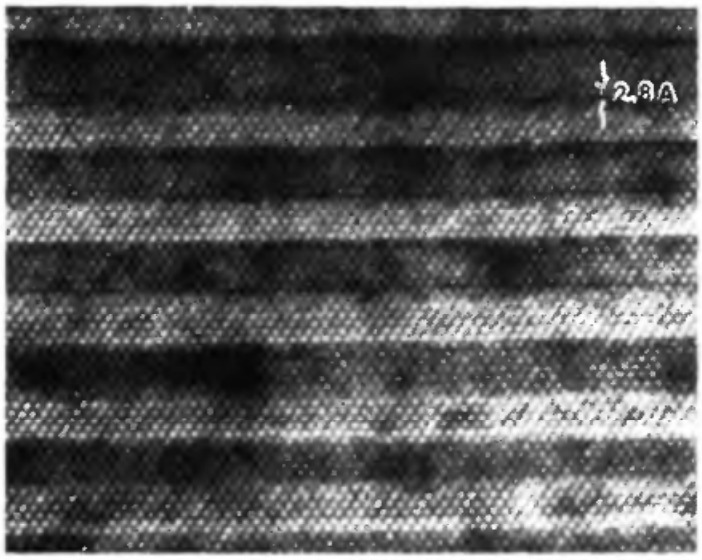


FIGURE 11.16. A multiple quantum well structure. This is a transmission electron microscope (TEM) display showing alternating layers of GaAs and $Al_xGa_{1-x}As$ materials. Note that the transition between the materials occurs abruptly on an atomic scale. (Data provided by J. J. Coleman.)

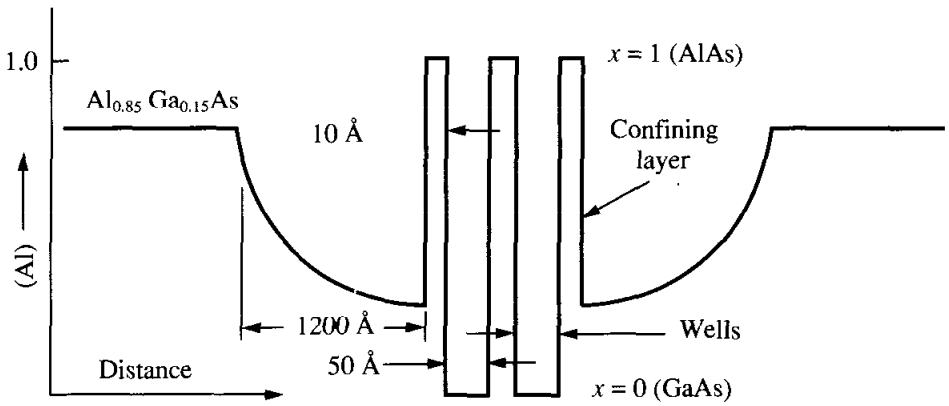
being achieved. We *can engineer a band gap* both in amplitude and in space. Figure 11.16 is one example of what has been done and points out that the combination is an artificial material with applications limited only by the imagination.

Although all of the above has focused on the conduction band, similar effects occur in the valence band but with some additional complications. Since it has both light and heavy holes (i.e., different effective masses) the positions of the subbands are also different. Transitions can occur between an electron state in the conduction band to either a light hole (lh) or a heavy hole (hh) state in the valence band with the details best left to more advanced texts.

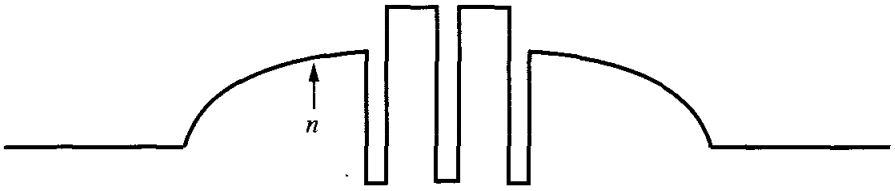
An example of the use of the quantum size effects discussed is illustrated in Fig. 11.17. This is intended to show the details of the active region of a diode laser with electrons being injected from the left and holes from the right. The quantum wells are so short that it is possible for the injected carrier to pass through the regions of GaAs because of the lack of collisions necessary to trap the carriers in the wells. Consequently, the aluminum concentration is slowly graded over much larger distances to aid in the trapping of the injected carriers in the central region where the recombination takes place by transitions between allowed states in the quantum wells formed by sandwiching thin layers of GaAs between even thinner layers of AlAs. These confining layers create a potential barrier to the wave function of the electron (or hole) trapped in the GaAs well and make the calculations given above approximate. This example also illustrates the use of the fact mentioned earlier, namely, that the band gap and index of refraction vary in opposite directions as the percentage of aluminum in the alloy is changed.

There are a couple of issues to be noted about the technology and the physics of such lasers:

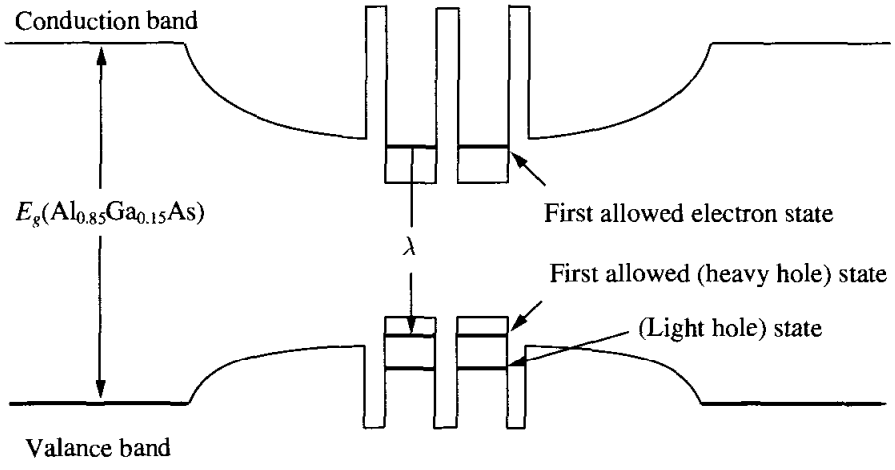
1. The distances are incredibly small. For instance, the 10-Å confining layers of AlAs, which provide the barrier to the electron wave function in the 50-Å GaAs layer is only $\sim$ 2-unit cell dimensions thick (see Table 11.1). Nevertheless, Fig. 11.16 indicates that the transition from one material to another can be sharp on an atomic



(a) Aluminum concentration at the junction



(b) Variation of the index of refraction



(c) Band diagram

FIGURE 11.17. An application of quantum size effects for junction lasers: (a) shows the typical variation of the aluminum concentration as a function of depth in the junction. The number of wells can be changed from one to many. (b) and (c) show the effect of the aluminum concentration on the index of refraction and the band gap.

scale. There is considerable technology invested in achieving that type of growth capability.

2. The graded layers provide a convenient structure to collect the carriers injected from either side. The size of the quantum wells is so small that the injected carriers will not scatter and collect into the wells if the graded layers are not present.
3. The first allowed band above E_c is not given by (11.6.5) because the confining layers do not present an infinitely high barrier as was assumed in the calculations of this section. However, the important point to remember is that the energy E_1 is a function of the layer thickness and thus *can be considered as a design variable*. Thus the wavelengths can also be *designed* (within limits).
4. The change in the aluminum concentration creates a highly desirable dielectric waveguide for the fields being amplified.

We should realize that an exact calculation of many of the physical features may not be possible, either because of incomplete knowledge of the growth parameters, the material properties, or a dependence of those properties on the size. However, this is offered as an indication of some of the possibilities and to convey some of the excitement that such devices generate. Imagination appears to be the most serious limitation.

11.6.2 Finite Barriers: An Example

The prior development suffers from the assumption that the barriers confining the electron to the smaller bandgap material was infinite. This is still a reasonable approximation for the semiconductor-air boundary involving eV's for the barrier, but it is a poor one for the heterointerface. Actual barrier heights are finite, and thus the electron wavefunction penetrates into the barrier region in the manner anticipated in Fig. 11.15 where, for instance, Ψ is drawn to be finite at $z = \pm L_2/2$ rather than zero as used for the boundary condition leading to (11.6.2). The formalism expressed in Sec. 11.6.1 is correct, but the allowed values of $k_\perp$ and hence sub-band energies are too large.

Let us illustrate, by way of the example shown in Fig. 11.18, the procedure for correcting the approach. Shown there is a practical situation of a 100 Å layer of GaAs ($E_g = 1.424$ eV) confined between two confining layers of 20% AlGaAs (i.e., $Al_{0.2}Ga_{0.8}As$) with $E_g = 1.6734$ eV where the following formula relating the experimental values [22] of E_g as a function of x has been used.

$$E_g = 1.424 + 1.247x \qquad (x < 0.45) \qquad \text{(direct band gap)} \tag{11.6.7}$$

One of the first conceptual problems encountered is the placement of the bandgap of the GaAs with respect to that of $Al_{0.2}Ga_{0.8}As$. Should it be in the middle with equal discontinuities in conduction and valence bands, should one of the bands line up, thereby assigning all of the $1.6734 - 1.424 = 0.2494$ eV discontinuity to the other, or should it be some other fraction? The answer is subject to considerable controversy with some of the reasoning best saved for specialized books [23]. Figure 11.18 was drawn assuming that

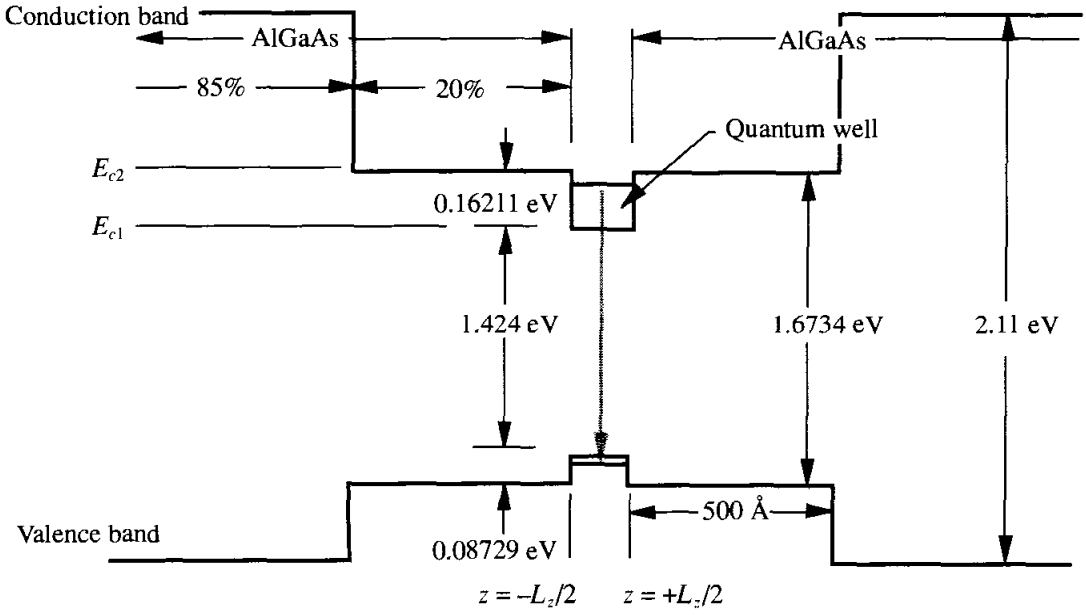


FIGURE 11.18. A 100 Å GaAs layer sandwiched between two 20% AlGaAs barrier layers that are ~ 500 Å thick. The 85% AlGaAs layer is to provide a dielectric waveguide for an electromagnetic mode. The indirect bandgap for $Al_{0.85}Ga_{0.15}As$ is given.

65% of bandgap discontinuity be assigned the conduction band.* (Casey and Panish [22] recommend a 85% assignment, illustrating the fact that the issue continues to be debated.)

To determine the allowed confined states, we must solve Schrödinger's equation for an electron propagating between the barriers but penetrating into the classically forbidden region of the 20% AlGaAs.

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dz^2} + (E_{c1} - \epsilon)\Psi = 0 \qquad |z| < L_z/2 \tag{11.6.8a}$$

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dz^2} + (E_{c2} - \epsilon)\Psi = 0 \qquad |z| > L_z/2 \tag{11.6.8b}$$

where $E_{c1} < \epsilon < E_{c2}$ for a bound state.

These equations are rewritten to emphasize the character of the solution and also to shift the energy scale by using $\epsilon = E + E_{c1}$, where E is the energy as measured (up) from band edge in the GaAs. Hence,

$$\frac{d^2\Psi}{dz^2} + \left\{\left[\frac{2m}{\hbar^2}\right]E\right\}\Psi = 0 \qquad E > 0 \qquad |z| < L_z/2 \tag{11.6.8c}$$

$$\frac{d^2\Psi}{dz^2} - \left\{\left[\frac{2m}{\hbar^2}\right](\Delta E_c - E)\right\}\Psi = 0 \qquad E < \Delta E_c \qquad |z| > L_z/2 \tag{11.6.8d}$$

where $\Delta E_c = E_{c2} - E_{c1}$ is conduction band difference in energies (= 0.16211 eV in Fig. 11.18).

*The numerical values for this example are carried to far greater precision than the data warrant. This is done so that the student can practice and obtain the same answers.

The form of (11.6.8c) and (11.6.8d) emphasizes the fact that the wavefunction must have a standing wave or trigonometric solution for the region within the well, and an exponentially decaying one for $|z| > L_z/2$. Define

$$k_z^2 = \frac{2m^*}{\hbar^2} E \qquad \text{(as usual)} \tag{11.6.9a}$$

$$h_z^2 = \frac{2m^*}{\hbar^2} (\Delta E_c - E) \tag{11.6.9b}$$

The solutions to these differential equations are very simple:

$$\psi(|z| < L_z/2) = A \cos k_z z + B \sin k_z z$$

$$\Psi(|z| > L_z/2) = C \exp(-h_z z) + D \exp(+h_z z)$$

The first allowed state is described by a wavefunction that is even with respect to z. Thus, $B = 0$ since the sine function is odd; the next confined state has an odd wavefunction, and $A = 0$ for it. The wave function must be bounded for $|z| > L_z/2$ hence, $D = 0$ for $z > L_z/2$ and $C = 0$ for $z < (-L_z/2)$. Once the even-odd character of the wavefunctions is recognized, it is sufficient to concentrate only on the region $z > 0$:

$$\Psi = \begin{bmatrix} A \cos k_z z & \text{(even)} \\ \text{or} & \\ B \sin k_z z & \text{(odd)} \end{bmatrix} \qquad \text{for} \quad z < L_z/2 \tag{11.6.10}$$

Schrödinger's equation is of second order, and the two boundary conditions needed are that Ψ and $d\Psi/dz$ are continuous at $z = L_z/2$.

$$\begin{bmatrix} A \cos \dfrac{k_z L_z}{2} \\ \text{or} \\ B \sin \dfrac{k_z L_z}{2} \end{bmatrix} = C \exp -\frac{h_z L_z}{2} \qquad \begin{bmatrix} -k_z A \sin \dfrac{k_z L_z}{2} \\ \text{or} \\ k_z B \cos \dfrac{k_z L_z}{2} \end{bmatrix} = -h_z C \exp -\frac{h_z L_z}{2} \tag{11.6.11}$$

Dividing the right by the left yields the characteristic or eigenvalue equation: For even wavefunctions

$$\tan \frac{k_z L_z}{2} = \frac{h_z L_z/2}{k_z L_z/2} \tag{11.6.12a}$$

For odd wavefunctions

$$-\cot \frac{k_z L_z}{2} = \frac{h_z L_z/2}{k_z L_z/2} \tag{11.6.12b}$$

There is only one unknown in these expressions, the allowed energy E above E_{c1}. To make the expressions more convenient, we recall that h_z and k_z were defined by (11.6.9):

$$k_z^2 = \frac{2m^*}{\hbar^2} E \qquad h_z^2 = \frac{2m^*}{\hbar^2} (\Delta E_c - E)$$

and thus we may write

$$h_z^2 = \frac{2m^*}{\hbar^2}\Delta E_c - k_z^2 \tag{11.6.13a}$$

Multiply (11.6.13a) by $(L_z/2)^2$ and identify some dimensionless quantities defined by

$$y^2 = \left(\frac{h_z L_z}{2}\right)^2 = \left(\frac{2m^*}{\hbar^2}\right)\left(\frac{L_z}{2}\right)^2 \Delta E_c - \left(\frac{k_z L_z}{2}\right)^2 = R^2 - x^2 \tag{11.6.13b}$$

where

$$\begin{aligned} x^2 &= (k_z L_z/2)^2 \\ R^2 &= \left(\frac{2m^*}{\hbar^2}\right)\Delta E_c \left(\frac{L_z}{2}\right)^2 \\ y^2 &= (h_z L_z/2)^2 \end{aligned} \tag{11.6.14}$$

Thus our eigenvalue equation becomes

$$\begin{bmatrix} x \tan x \\ \text{or} \\ -x \cot x \end{bmatrix} = y \text{ and } x^2 + y^2 = R^2 \tag{11.6.15a}$$

or

$$\begin{bmatrix} x \tan x \\ \text{or} \\ -x \cot x \end{bmatrix} = y = \sqrt{R^2 - x^2} \tag{11.6.15b}$$

A graphical solution illustrates the character and implications of the solution much better than a dry numerical one, and is shown in Fig. 11.19. The left side of (11.6.15) is the familiar trigonometric functions with the tangent function going to zero at $x = (0, m\pi)$ and infinity at odd multiples of $\pi/2$ and the cotangent function has zero at $\pi/2$ and asymptotes at multiples of π. The second equation has the functional form of a circle of radius R, which is proportional to the square-root of the conduction band discontinuity. The intersection is of course the desired solution for x (i.e., k_z) and the energy E measured from the band edge.*

For the case shown in Fig. 11.19 with the circles having a radius of less than $\pi/2$, it is obvious that there is only one allowed solution of the even wavefunction type. If, however, the energy barrier were large enough such that the "radius" were larger than $\pi/2$, then there would be a second solution or a second bound state with an odd wavefunction. For the odd wavefunctions, the left side starts at -1 for $x = 0$, goes through zero at $x = \pi/2$, and then to infinity at $x = \pi$. If $R < \pi/2$, then there is only one bound state because there is no solution for the odd wavefunction.

*The perceptive student will recognize this analysis as being identical to the slab waveguide problem of Chapter 4. This is not an accident. The normalized variables were chosen to emphasize the similarities.

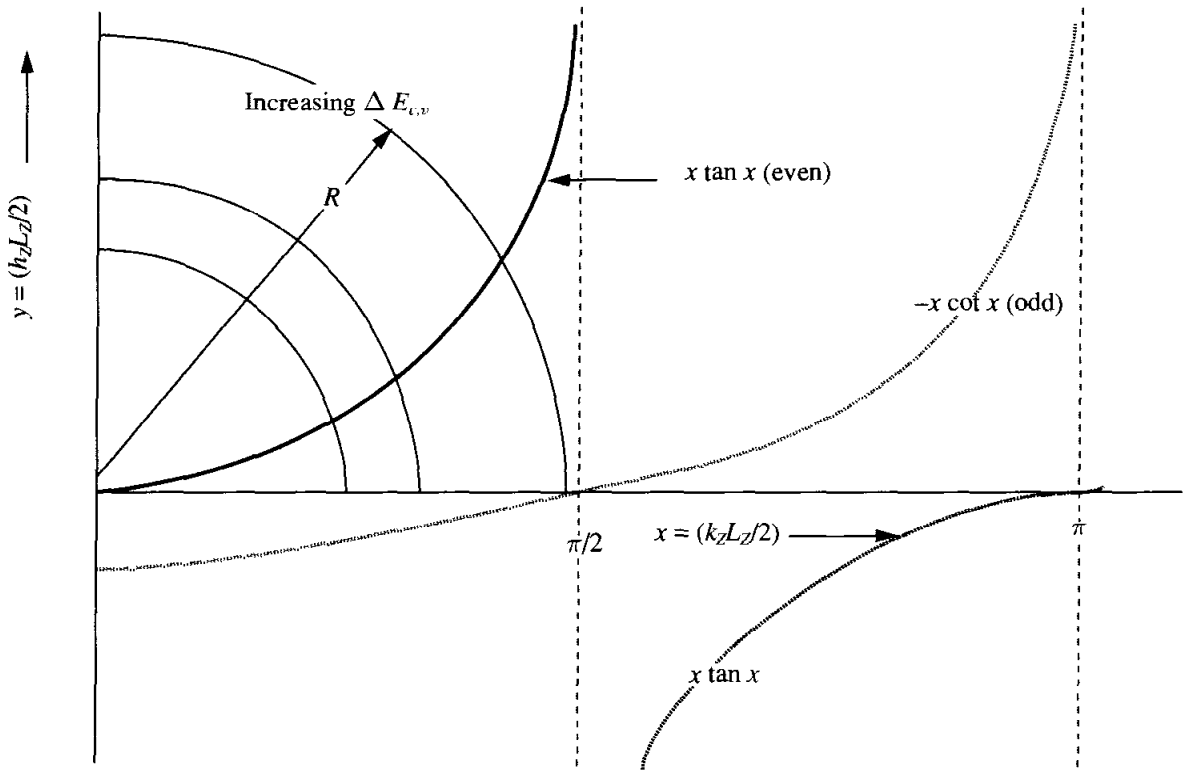


FIGURE 11.19. A graphical solution to the eigenvalue equation.

If $\Delta E_c \to \infty$ as has been assumed in Sec. 11.6.1, then the first intersection with an infinite radius occurs at $x = \pi/2$ and $k_z L_z/2 = \pi/2$ or $k_z = \pi/L_z$ as we had before. Because of the smaller barrier ΔE_c, $k_z < \pi/L_z$. For the dimension in Fig. 11.18 we have

$$R^2 = \left(\frac{2m}{\hbar^2}\right)\Delta E_c \left(\frac{L_z}{2}\right)^2$$

$$= \frac{2 \cdot 0.067 \cdot 9.11 \times 10^{-31}\ \text{kg}}{(1.0545 \times 10^{-34}\ \text{joule sec})^2} \cdot \frac{1.6 \times 10^{-19}\ \text{joule}}{\text{eV}} \cdot (0.16211\ \text{eV}) \left(\frac{10^{-8} m}{2}\right)^2$$

$$= 7.11873\ (\text{rad})^2 = (2.66809\ \text{rad})^2$$

Obviously $x < R$ in order for the right side of (11.6.15) to be real and thus x must be less than $2.6681 < \pi$ rad indicating that only 2 states are confined in the quantum well.

Using a standard root finding technique (on a calculator), we obtain

$x_{1e} = 1.13245\ \text{rad}\ (\pm 10^{-9})$ $\quad y_{1e} = 2.4158$ $\quad h_z = 483.2\ (\mu\text{m})^{-1}$

$x_{2e} = 2.1832\ \text{rad}$ $\quad y_{2e} = 1.5337$ $\quad h_z = 306.8\ (\mu\text{m})^{-1}$

$$(k_{z1} L_z/2)^2 = x_{1e}^2 \quad \text{and} \quad E_{1e} = \frac{\hbar^2}{2m} k_z^2 = \frac{\hbar^2}{2m}\left(\frac{2}{L_z}\right)^2 x_{1e}^2$$

$$\therefore E_{1e} = 4.672 \times 10^{-21}\ \text{joules} \Rightarrow 29.2\ \text{meV} \quad \text{and} \quad E_{2e} = 108.54\ \text{meV}$$

By using the infinite barrier theory (see Problem 11.14), the first allowed state would be at 56.2 meV, and, clearly, the finite barrier made a significant difference, almost a factor of 2 in the energy above the band edge.

To apply this theory for the holes, we use the energy barrier of 0.08729 eV and an effective mass of $0.55\,m_0$ yielding $R_h^2 = (5.60948\text{ rad})^2$. Repeating yields the energy below the valence band edge of the first allowed (heavy) hole state.

$$x_{1hh} = 1.3312\text{ rad} \qquad \text{or} \qquad E_1(hh) = \frac{\hbar^2}{2m^*}\left(\frac{2}{L_z}\right)^2 x_{1hh}^2$$

$$= 7.866 \times 10^{-22}\text{ joules} \Rightarrow 4.916\text{ meV}$$

Thus, the long wavelength limit of the transition between the first allowed electron state and the first hole state is $h\nu/e = (E_g = 1.424\text{ eV}) + (E_{1e} = 2.92 \times 10^{-2}\text{ eV}) + (E_{1hh} = 4.916 \times 10^{-3}\text{ eV}) = 1.4581\text{ eV}$ or $\lambda_0 = 8502\text{Å}$.

The presence of the finite barriers does change the value of the first allowed state and also some of the details of the calculation leading to (11.6.6). For instance, the volume of the cylindrical shell in k space changes because k_z is now $2x_{1e}/L_z$ rather than π/L_z. However, the height of the shaded volume in Fig. 11.14 also changes to that value leaving the density of states in the first sub-band the same as before

$$\rho(E)\,dE = \frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)\left(\frac{\pi}{L_z}\right)dE \quad E > E_1 \tag{11.6.6}$$

The density of states is still independent of energy above E_{1e}, but now there is a limitation on the number of sub-bands within the well. Furthermore, if we attempt to fill a sub-band with too many electrons, the quasi-Fermi level increases and part of the "tail" of the distribution in a sub-band may extend above the barrier as shown in Fig. 11.20. Those electrons, shown as the darkest area, can escape into the 20% AlGaAs, and thus are not confined to the quantum well.

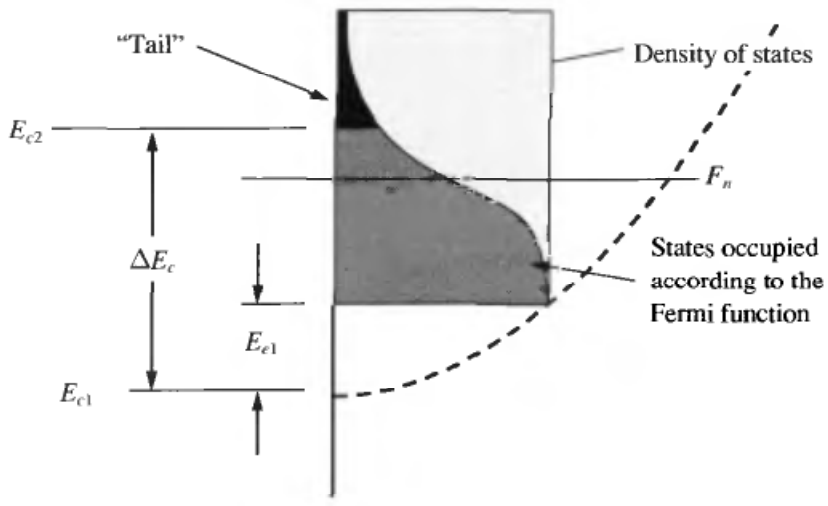


FIGURE 11.20. The distribution of electrons in the first bound state of a quantum well.

The width of the 20% barrier layer of Fig. 11.8 is usually large enough so that the tail of the wavefunction does not penetrate to the 85% interface. For $h_{z1} = 483.2\ (\mu\text{m})^{-1}$ and 500 Å (0.05 μm) of the $Al_{0.2}Ga_{0.8}As$, the wavefunction is only $\exp[-24.2] = 3.2 \times 10^{-11}$ of its value at $z = \pm L_z/2$. That is an adequate approximation to zero.

11.7 VERTICAL CAVITY SURFACE EMITTING LASERS

Most heterostructure quantum well lasers employ an optical cavity parallel to the substrate in the manner shown in Fig. 11.11, where some typical dimensions were added to emphasize the disparity in the dimensions with the thickness of the active region being much less than the width (or stripe) which is much less than the length of the cavity. For a quantum well laser, even the thickness implied in Fig. 11.11 is a gross exaggeration, and a better view of the cross section of the 1 μm thickness of such a laser is shown in Fig. 11.21.

Shown in Fig. 11.21(a) is the variation of the index of refraction and bandgap with the dimension perpendicular to the substrate and illustrates the asynchronous behavior of the bandgap and the index of refraction. Notice that the dimension of the quantum well is only $\sim$ 70 Å and very small compared to the optical confinement layers of $\sim$ 1000 Å and also small compared to the mode size of $\sim \lambda_0/2n$. The aluminum composition increases with distance away from active quantum layer so that the index decreases, thereby providing field guidance. The higher bandgap of the aluminum compositions ensure that most of the injected carriers collect and recombine in the thin GaAs layer, thereby providing enhanced gain there.

However, we have now created a problem: The overlap between the gain region and E^2 is small, which tends to negate the enhanced gain. The quantity that measures this overlap is the confinement factor Γ (not to be confused with the field reflection coefficient).

$$\Gamma = \frac{\int_{QW} E^2(x)\,dx}{\int_{\text{all } x} E^2(x)\,dx} \tag{11.7.1}$$

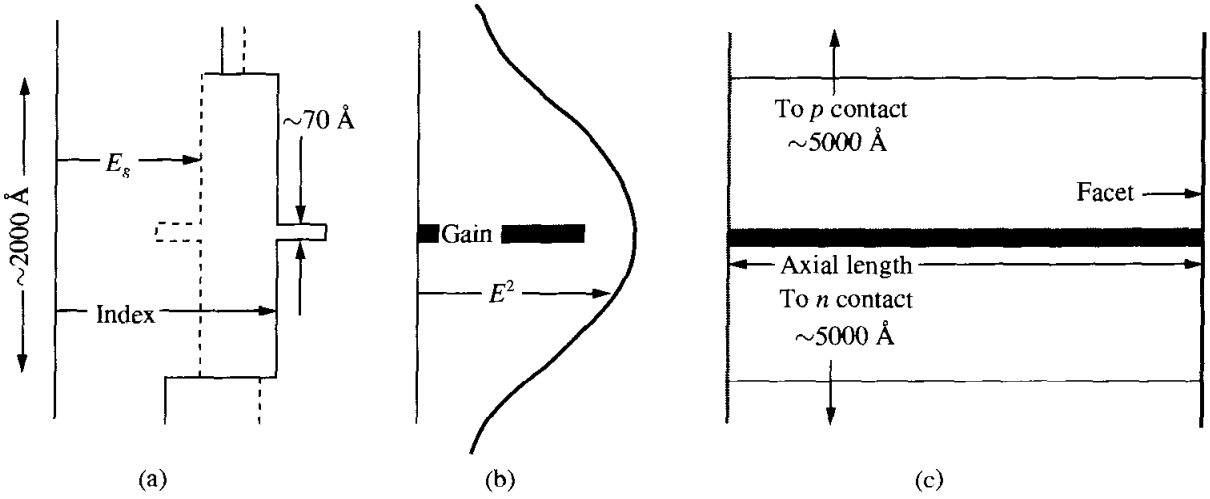


FIGURE 11.21. Typical dimension of a planar quantum well laser.

(~ 0.035 for Fig. 11.21).

Since the quantum well thickness is very small, the modal gain coefficient is considerably smaller than the material gain coefficient.

$$\gamma_{\text{modal}} = \Gamma\gamma_{\text{mat}} \ll \gamma_{\text{mat}} \tag{11.7.2}$$

On the other hand, we can compensate for the small gain coefficient by using longer lengths, and that ploy has been successful.

A major drawback to the conventional planar semiconductor laser is its beam quality: An elliptical rather than a cylindrical mode is generated, and thus conventional circular optics are far from the diffraction limit. While we can phase-lock arrays of planar semiconductor lasers (see Sec. 12.10), but the array is limited to one dimension. If we had a vote, we would opt for a two-dimensional array of laser elements in the manner sketched in Fig. 11.22 and

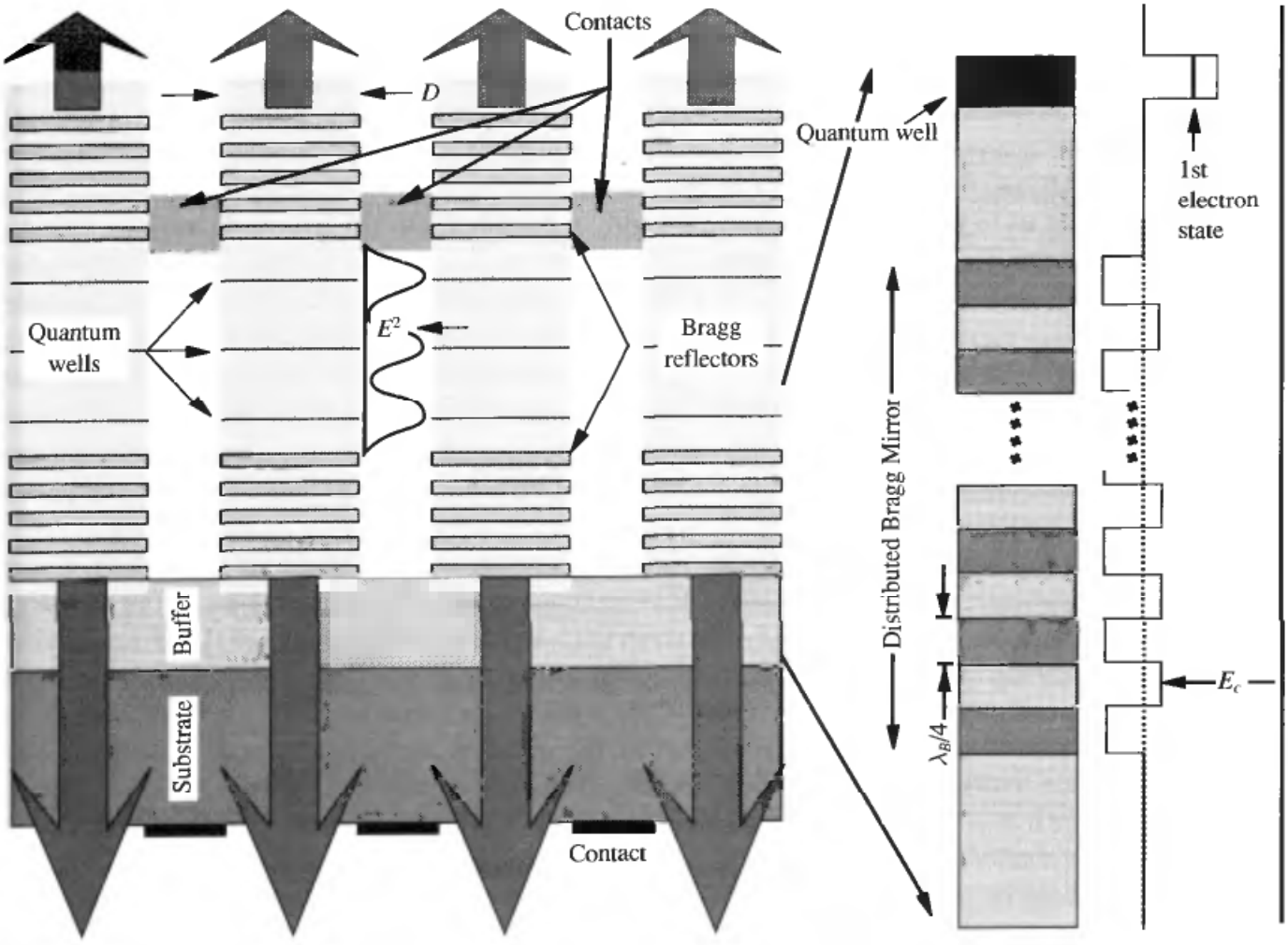


FIGURE 11.22. A possible arrangement of vertical cavity surface emitting lasers (VCSELs). A typical dimension of D might be 5 to 50 μm. The number of wells might be anywhere from 1 (see [33]) to 20 (see [24]).

shown on the front cover of the book. Shown there is a 8×8 array of VCSEL's emitting at 725 nm, with each element being addressable.*

In the ideal world, one starts with a substrate, say GaAs, grows a buffer layer to establish a pristine crystalline surface, and then grows alternating layers of lattice matched materials, each a quarter wavelength thick (in the dielectric) but with differing indices so as to build up the reflection coefficient by successive mismatches, and then places quantum wells at the peak of the standing wave of the electric field. If the photon energy emitted by the quantum well is smaller than bandgap of the substrate, then the output can be taken out of the bottom. We would need to use InGaAs for the quantum well if the substrate were GaAs, and this last option is often chosen. Alternatively, we would take the radiation from the top if a GaAs well were used since the substrate would absorb the radiation emerging through the bottom mirror. Figure 11.22 assumed lasing in a "post" D units in diameter with the active layer placed $\lambda_0/4n$ from the reflecting mirrors, which are presumed to have a field reflection coefficient of $\Gamma = -(\sim 1)$. However, $\Gamma = +(\sim 1)$ would work just as well provided the quantum well is placed at $\lambda_0/2n$ from the mirrors.

These mirrors do not represent a new technology, just a new and innovative application for an old theory from transmission lines. However, since the mirrors are actually thicker than the region with gain, we usually refer to this system as a distributed (Bragg) reflector.† We can use the same growth technology that has been successful in forming the conventional planar quantum well laser and apply it to the vertical cavity laser shown in Fig. 11.22. For the AlGaAs/GaAs system, the differences in indices of different compositions of AlGaAs enable us to grow a lattice matched highly reflecting dielectric mirror directly on a GaAs substrate. Then, we grow various spacer layers to ensure that the quantum well is at the peak of the standing wave. Finally, one grows or deposits a highly reflecting top mirror to couple the laser radiation out perpendicular to the substrate. The top mirror can be composed of the multiple-layer lattice matched semiconductor materials as is the one on the substrate, but the voltage drops across the heterolayers increase the resistance and lower the efficiency of the laser. However, once the lattice matched environment is established up to the top contact, other materials such as the traditional insulators such as SiO_2, TiO_2, Si_3N_4 can be used for the top mirror.

We would hope to apply this technology to the very important 1.5 μm region, where the quaternary III-V compounds lattice matched to InP are commonly used. For instance, $In_xGa_{1-x}As_yP_{1-y}$ can have E_G between 0.75 and 1.34 eV (1.65 μm to 0.925 μm) and is lattice matched to InP ($a = 5.869$ Å) with its bandgap of 1.35 eV provided the fractional composition y is related to x by $y = 2.16(1 - x)$. For such a case, the substrate is transparent to the radiation produced by the quantum wells. Unfortunately, the index of refraction does not change nearly as much with (x, y), as it does for the AlGaAs system. Hence it is difficult to build up a significant reflectivity with a reasonable number of layers if we restrict the materials to only this system of quaternary compounds. There are different methods of overcoming this problem, but they are still in the research stage. Reference

*The author greatly appreciates the generosity of Dr. R. P. Bryan of Photonics Research, Inc. for providing some of the figures for the front cover.

†A general theory for such mirrors will be covered in Sec. 12.5.

[35], for instance, used the $AlAs_{0.52}Sb_{0.48}$/InGaAs(P) layer system on InP to achieve high reflectivity in the 1.7 μm band.

There are some obvious points of contrast and comparison between the two laser geometries.

1. The quantum well material for the gain region can be the same in either case so that the material gain coefficient is identical.
2. The gain length for a planar laser may be anywhere from 100 to 1000 μm, whereas the gain length for a vertical cavity surface emitting laser (VCSEL) is the number (say, $N = 1$ to 3) of quantum wells times each thickness, say, 100 Å, to give a gain length of $N \cdot 100$ Å ~ 300 Å $= 0.03$ μm (Reference [24] used 20 wells of 100 Å length, but the length is still small). In addition, the optical confinement factor is small for the conventional planar laser whereas it is ~ 1 for VCSEL. Ref. [33] estimates a value of 2000 cm^{-1} for the material gain coefficient and thus the maximum modal gain $\times$ gain length is only 0.6% for a VCSEL. We can probably double that number since the gain regions are placed at the antinodes of the fields and thus the stimulated emission rate would be twice that for a traveling wave. In any case, we need highly reflecting mirrors with $R > 98\%$. As a contrast, we should realize that the facet reflectivity is ~ 0.36 for the mirrors of the planar laser.
3. The VCSEL geometry is a "natural" for a two-dimensional (2-D) matrix array that may or may not be phase locked and/or individually addressed. At best, only a one-dimensional (1-D) array can be achieved with a planar system.
4. The VCSEL can produce a cylindrically symmetric beam ([26] and [28]) at its output aperture, which naturally mates to or can be imaged onto a fiber. The beam from a planar laser is elliptical and possibly astigmatic and is badly mismatched to a fiber.
5. Unfortunately, the multiple heterojunctions of the mirrors gives rise to high series resistance. Consequently, current injection into the active region is much more efficient for the planar laser in comparison to the VCSEL. By grading the composition of the layers in the mirrors, the authors in Ref. [34] have shown that a drastic reduction of the series resistance is possible with little degradation in the optical reflectivity. In any case, VCSELs appear to be limited to low power ($\sim$ mW) and parallel beam applications whereas the planar laser is more appropriate for the higher power applications.
6. The number of high Q resonant modes in a VCSEL may be easily counted—say, 1 to 5—since the length between the mirrors is only half to a few wavelengths whereas there are thousands in a planar one. The total number of modes is so small that the presence or absence of a mode can even affect the spontaneous emission (Refs. [33] to [39]).
7. The free spectral range of the microcavity is so large ($\sim$ 1000Å) that the VCSEL tends to be a single mode (wavelength) source whereas multimode operation can easily exist for a planar laser.

The VCSELs are a relatively new innovation in the laser field, and it is not clear whether any of the limitations noted are fundamental or whether any of the advantages can be maintained over a long period. It is surely an area calling for more innovative thinking and research.

11.8 MODULATION OF SEMICONDUCTOR LASERS

The fact that the diode lasers are so conveniently pumped by simple current injection (mA and at a few volts drop) makes them the clear-cut candidate for communications. A typical light-out current-in ($L - I$) curve is shown in Fig. 11.23 with threshold currents as low as 10 to 50 mA and a nearly linear light output above threshold.

It takes little imagination to recognize that if the current is varied by ΔI around an operating level I_0, then the light output will also vary in a more or less synchronous fashion. If the variation is slow enough, then the two will be "in phase." However, the quality of many communication systems depends on the bandwidth, and thus our problem is to ascertain the frequency response of the diode transmitter. We do this by examining the interplay between stimulated emission and the injected carriers.

Let us now formulate the rate equations for a semiconductor laser. Because of the small physical size of these lasers, we again treat the system as a circuit element whereby we ignore the spatial variations of carrier density and photons with z. We also treat the cross sectional overlap between the electromagnetic mode and the inverted population as a constant that only needs to be computed once.

$$\Gamma = \frac{\int_{-d/2}^{+d/2} E^2(x)\,dx}{\int_{-\infty}^{+\infty} E^2(x)\,dx} \tag{11.8.1}$$

In other words, Γ is that fraction of the photon flux of the cavity mode that overlaps the inverted population (assumed to be in the width d).

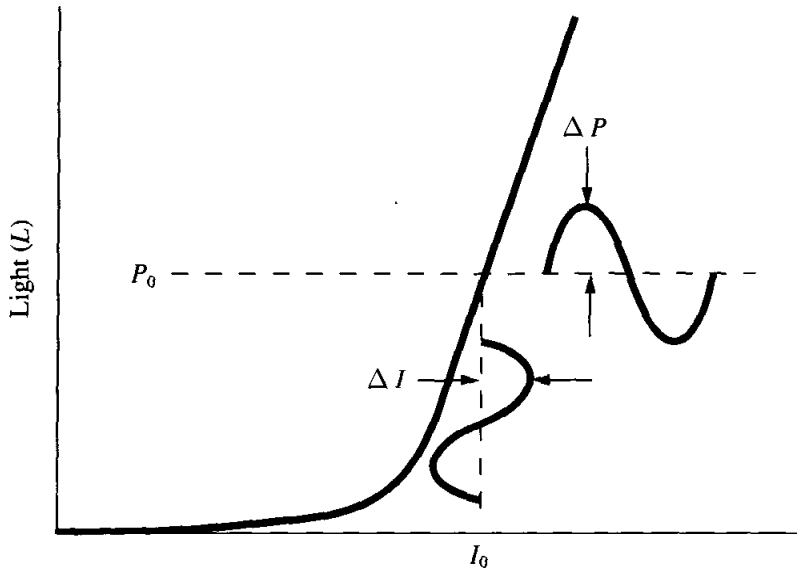


FIGURE 11.23. A modulated laser.

It should be clear from the previous sections that the density of the inverted population N (electrons in the conduction band and holes in the valence band) is equal to the injected current divided by the electronic charge and the distance over which the recombination takes place. Thus the production term for the inversion becomes

$$\frac{dN}{dt}\,(\text{production}) = \frac{J}{ed} \tag{11.8.2a}$$

The carriers are lost spontaneously by both radiative and nonradiative processes. Let us assign a rate for the total recombination in terms of a carrier lifetime $(1/\tau_s)$. Inasmuch as it is a spontaneous decay, the rate of decrease of the inversion is proportional to the inversion multiplying that rate.

$$\frac{dN}{dt}\,(\text{spontaneous recombination}) = -\frac{N}{\tau_s} \tag{11.8.2b}$$

Stimulated emission depletes the inversion in direct proportion to the intensity of the stimulating wave. However, it is never very important until a laser is near or above threshold, and then it is all important. Thus the stimulated recombination rate of carriers should account for the fact that the inversion density must be at least high enough to make the cavity *transparent* (i.e., the gain just about equal to the losses) before stimulated loss of carriers occurs. Furthermore, the net rate of stimulated emission will tend to drive the inversion back to its transparent or threshold condition.

$$\frac{dN}{dt}\,(\text{stimulated emission}) = -A(N - N_{\text{tr}})P \tag{11.8.2c}$$

where N_{tr} is the value of the inversion that makes the material transparent and the factor A is an abbreviation for a host of other quantities. For instance, the photons in the cavity are spread out over a cross section that is much larger than the inverted width d, but only the field in that region does the stimulation. Consequently the optical confinement factor Γ appears as one factor in A. The others are (1) the gain coefficient per unit of inversion, (2) the group velocity of the wave composing the cavity mode, and (3) the volume of the cavity. (Surely the need for an abbreviation should be appreciated.)

Thus the total rate of change of the inversion is the sum of (11.8.2a)→(11.8.2c):

$$\frac{dN}{dt} = \frac{J}{ed} - \frac{N}{\tau_s} - A(N - N_{\text{tr}})P \tag{11.8.3}$$

This equation is merely a system of bookkeeping of the supply and use of the carriers and how they interact with the photons.

The photons in the cavity mode have a separate equation of motion that is coupled to (11.8.3). The stimulated emission term (11.7.2c) is the same except for a sign change.

$$\frac{dP}{dt}\,(\text{stimulated emission}) = A(N - N_{\text{tr}})P \tag{11.8.4a}$$

Only a fraction of the recombination included in (11.8.2b) results in a photon, and only a fraction of that emission enters the cavity mode whose photon number is represented by

P.* We hide our ignorance of the precise amount of spontaneous decay that yields a photon in this mode by assigning a fraction β of (11.7.2b) to the photon equation.

$$\frac{dP}{dt}(\text{spont}) = \beta \frac{N}{\tau_s} \tag{11.8.4b}$$

Finally the photons are lost from the cavity at a rate dictated by the unavoidable resistive losses remaining inside the cavity as well as by coupling to the external world. As usual, the photon lifetime is used here to describe this change in P.

$$\frac{dP}{dt}(\text{coupling, internal losses}) = -\frac{P}{\tau_p} \tag{11.8.4c}$$

Thus the rate equation for the photons becomes

$$\frac{dP}{dt} = +A(N - N_{\text{tr}})P + \beta \frac{N}{\tau_s} - \frac{P}{\tau_p} \tag{11.8.5}$$

These two simple equations contain a lot of information, and it behooves us to extract as much as possible from them.

11.8.1 Static Characteristics

Let us assume a static (DC) excitation so that all time derivatives are zero. We assume $J = J_0$, $N = N_0 > N_{\text{tr}}$ and thus $P = P_0$ (to be determined) so that (11.8.3) can be solved for the inversion density.

$$A[N_0 - N_{\text{tr}}]P_0 + \beta \frac{N_0}{\tau_s} = \frac{P_0}{\tau_p} \tag{11.8.6a}$$

or

$$A[N_0 - N_{\text{tr}}] = \frac{1}{\tau_p} - \beta \frac{N_0}{P_0 \tau_s} \tag{11.8.6b}$$

where the subscript 0 refers to the DC or steady state values of all parameters. The two forms of (11.8.6) state something very simple but very important. Equation (11.8.6a) states that the sum of the stimulated and spontaneous rates must be equal to the photon loss rate (internal plus external), as is true for *any* laser. Equation (11.8.6b), which is a minor rewrite of (11.8.6a), points out that spontaneous emission becomes less and less important as the optical power becomes bigger. It also points out that the net gain rate $A(N_0 - N_{\text{tr}})$ is clamped more or less at threshold and is equal to the photon loss rate (if the last term involving spontaneous emission divided by P_0 is neglected).

Now let us substitute (11.8.6a) into (11.8.3) and solve for the photon density.

$$0 = \frac{J_0}{ed} - \frac{N_0}{\tau_s} - \frac{P_0}{\tau_p} + \beta \frac{N_0}{\tau_s} \tag{11.8.7a}$$

or

*Spontaneous emission goes into any of 4π directions with just about equal probability, but stimulated emission goes into the same direction, at the same frequency, and same polarization as the stimulating wave.

$$P_0 = \frac{J_0 \tau_p}{ed} - (1 - \beta)\frac{N_0 \tau_p}{\tau_s} \tag{11.8.7b}$$

It is convenient to rewrite this equation in a manner that keeps tabs on the spontaneous power.

$$P_0 = \frac{\tau_p}{ed} \cdot \left[J_0 - \frac{N_0 ed}{\tau_s} \right] + \beta \frac{N_0 \tau_p}{\tau_s} \tag{11.8.7c}$$

As mentioned earlier, N_0 is clamped more or less at threshold for lasing, and thus (11.8.7c) suggests that the diode output can be written in the following format:

$$P_0 = K(J - J_{\text{th}}) + P_{\text{spont.}} \tag{11.8.7d}$$

where the threshold current density J_{th} is $N_{\text{tr}} ed/\tau_s$. Such an expression is consistent with the sharp "break" in the light (L) output versus current (I) curves found in most lasers such as that indicated in Fig. 11.23.

There is another way of explaining this sharp threshold behavior. Below threshold, spontaneous emission goes into any of the electromagnetic modes coupling to the active region: $[(8\pi \nu^2 n^3/c^3)\Delta\nu] \cdot$ volume in number. This is a huge number for even a small semiconductor: pick GaAs of volume 1 mm^3 cube, $n = 3.6$, $\lambda_0 = 8400$ Å, and $\Delta\lambda = 50$ Å. The number of modes staggers the imagination; that is, $1.17 \times 10^{+10}$! Most of those modes represent waves propagating in the wrong direction. We, living in the external world, "see" only those photons emitted in our direction and in the bandwidth of our receiving system, which is a very small fraction of the total (say, 1 part in 10^3). When the laser is above threshold, the excess pumping goes into one or at most a few, perhaps 10 to 20, modes, which represent waves propagating toward us. Thus it is no wonder that we see a dramatic increase in intensity when stimulated emission takes place. (Incidentally, this same line of reasoning applies to any laser.)

11.8.2 Frequency Response of Diode Lasers

The above is sufficient when the injected current is a slowly varying function of time, but under high-frequency excitation, such as at a GHz rate, we have to be concerned about the ability of stimulated emission to keep up with the rate of carrier injection.

Let us assume that the diode is biased above threshold by DC current $J_0 > J_{\text{th}}$ and an AC current $\Delta J(t)$ is added:

$$J = J_0 + \Delta J(t) \tag{11.8.8a}$$

Thus the carrier density N and photon density P should have a DC and a time-varying component.

$$N = N_0 + \Delta N(t) \tag{11.8.8b}$$

$$P = P_0 + \Delta P(t) \tag{11.8.8c}$$

where it is presumed that the time-varying components are small compared to the DC values. Now we substitute these equations into (11.8.3) and (11.8.5) to obtain

$$\frac{d\Delta N(t)}{dt} = \frac{J_0}{ed} + \frac{\Delta J(t)}{ed} - \frac{N_0 + \Delta N(t)}{\tau_s} - A[N_0 + \Delta N(t) - N_{tr}] \cdot [P_0 + \Delta P(t)] \tag{11.8.9a}$$

and

$$\frac{d\Delta P}{dt} = A[N_0 + \Delta N(t) - N_{tr}] \cdot [P_0 + \Delta P(t)] + \beta \frac{[N_0 + \Delta N(t)]}{\tau_s} - \frac{P_0 + \Delta P(t)}{\tau_p} \tag{11.8.9b}$$

We neglect products of time varying quantities (i.e., $\Delta N \cdot \Delta P$) as being of second order and equate, separately, the DC variables and time-varying components. The equations involving the DC terms reproduce precisely the material covered in the previous section. The equation for the AC variables becomes

$$\frac{d\Delta N}{dt} = \frac{\Delta J}{ed} - \left(\frac{1}{\tau_s} + AP_0\right) \cdot \Delta N - A[N_0 - N_{tr}] \cdot \Delta P \tag{11.8.10a}$$

$$\frac{d\Delta P}{dt} = A[N_0 - N_{tr}]\Delta P + AP_0\Delta N - \frac{\Delta P}{\tau_p} + \beta \frac{\Delta N}{\tau_s} \tag{11.8.10b}$$

These can be simplified somewhat by the use of (11.8.6b) in which we neglect the spontaneous term as being small when the laser is biased well above threshold so that (11.8.10a) and (11.8.10b) become

$$\frac{d\Delta N}{dt} = \frac{\Delta J(t)}{ed} - \left(\frac{1}{\tau_s} + AP_0\right) \cdot \Delta N - \frac{\Delta P}{\tau_p} \tag{11.8.11a}$$

$$\frac{d\Delta P}{dt} = \left[AP_0 + \frac{\beta}{\tau_s}\right] \Delta N \tag{11.8.11b}$$

where the first and third terms of (11.8.10b) cancel on use of (11.8.6b). We can eliminate the coupling by differentiating one and substituting the other.

$$\frac{d^2\Delta N}{dt^2} + \left[\frac{1}{\tau_s} + AP_0\right] \frac{d\Delta N}{dt} + \frac{1}{\tau_p}\left[AP_0 + \frac{\beta}{\tau_s}\right] \Delta N = \frac{1}{ed} \cdot \frac{d\Delta J}{dt} \tag{11.8.12}$$

The details of the derivation of a similar equation for ΔP are left as a problem. The answer is

$$\frac{d^2\Delta P}{dt^2} + \left[\frac{1}{\tau_s} + AP_0\right] \frac{d\Delta P}{dt} + \frac{1}{\tau_p}\left[AP_0 + \frac{\beta}{\tau_s}\right] \Delta P = \frac{\Delta J}{ed}\left[AP_0 + \frac{\beta}{\tau_s}\right] \tag{11.8.13}$$

It is most informative if we form the "transfer" characteristics by letting $\Delta J(t) = \Delta J_m \exp(j\omega t)$ in (11.8.13) and solving for $\Delta P(t) = \Delta P_m \exp(j\omega t)$. The light-current transfer function becomes

$$\frac{\Delta P_m}{\Delta J_m} = \frac{(1/ed)A\left(P_0 + (\beta/\tau_s)\right)}{\left[(1/\tau_p)\left(AP_0 + (\beta/\tau_s)\right) - \omega^2\right] + j\omega \cdot \left[(1/\tau_s) + AP_0\right]} \tag{11.8.14}$$

This modulation response has the functional form of a second-order lowpass filter. The resonance in the transfer characteristic occurs when

$$\omega^2 = \frac{1}{\tau_p} \cdot \left[AP_0 + \frac{\beta}{\tau_s} \right] \approx \frac{AP_0}{\tau_p} \tag{11.8.15}$$

Equation (11.8.15) implies that the resonant frequency increases as the photon lifetime decreases, which is to be expected, and also increases as CW power increases. At higher frequencies, the response drops off as ω^{-2} or 40 dB per decade of frequency.

Fig. 11.24, taken from a publication by Lau and Yariv [4], illustrates a qualitative agreement with this theory. The resonance does increase with DC current (i.e., laser power) but seems to fall off faster than 40 dB/decade suggested by the analogy with a second-order circuit filter, but this is most likely caused by parasitic circuit effects. Although the theory suggests that a high average power leads to a high modulation capability, there is a practical upper limit to this approach. At too high a power, the photon flux at the output facet of the diode literally destroys the surface—it self-destructs. Other trends that are in accordance with the theory are clearly demonstrated in that paper. However, the important point is that multi-GHz modulation is possible. Let us end this chapter with the same thought as before: Semiconductor lasers are clearly the optical device of choice for low power control, interrogation, and communication. Although the semiconductor laser suffers in comparison to the much larger gas and solid-state laser in terms of power/energy and beam quality, many applications do not need those extremes. Finally, the semiconductor laser naturally mates to the rest of modern electronics, and that convenience makes up for a lot of its minor

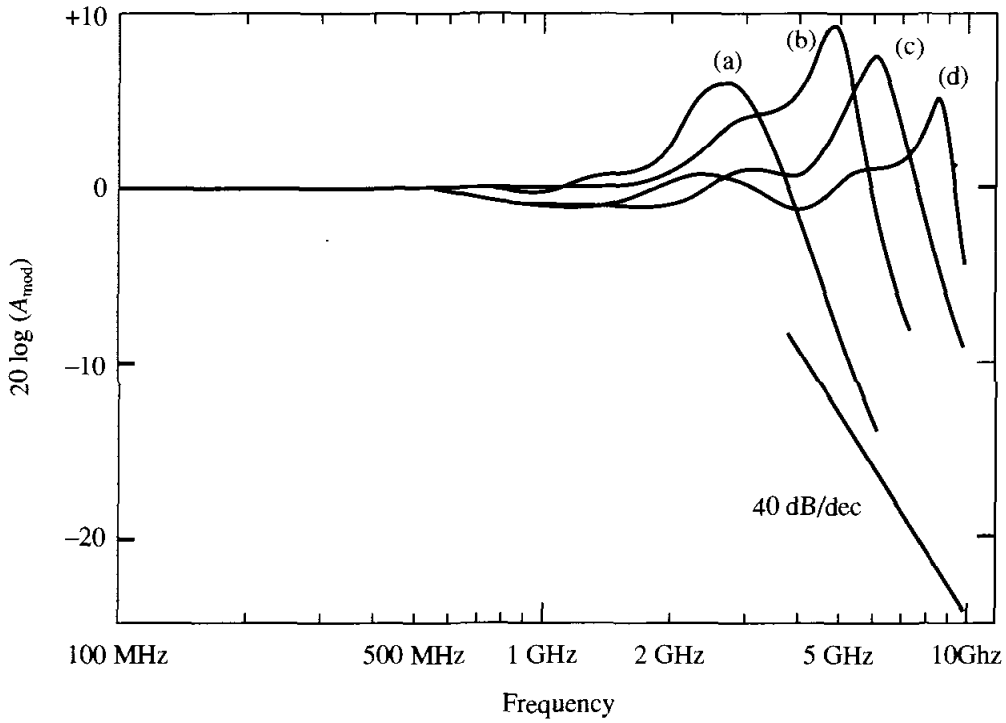


FIGURE 11.24. Modulation characteristics of a short-cavity (120 μm), buried-hetero-structure laser as a function of bias levels: (a) 1 mW, (b) 2 mW, (c) 2.7 mW, and (d) 5 mW. (Data from Lau and Yariv [4].)

shortcomings. Some of the detailed electromagnetic issues associated with semiconductor lasers are addressed in Chapter 12.

PROBLEMS

11.1. Consider an intrinsic GaAs semiconductor with $m_e^* = 0.067\, m_0$, $m_{hh}^* = 0.55\, m_0$, $m_{lh}^* = 0.067\, m_0$, and $E_g = 1.43$ eV. Optical pumping creates 5×10^{18} cm^{-3} electrons in the conduction band, leaving an equal number of holes in the valence band. Assume $T = 0$ K.

(a) What is the position of the quasi-Fermi level, F_n, relative to the conduction band? (Ans.: 0.1596 eV.)

(b) What is the position of F_p relative to E_v? (Ans.: 0.0189 eV.)

(c) What are the densities of the light and heavy holes? (Ans.: $n_{hh} = 4.8 \times 10^{18}$ cm^{-3}; $n_{lh} = 2 \times 10^{17}$ cm^{-3}.)

(d) What is the "speed" of the electrons in the highest filled state? (Ans.: 9.14×10^7 cm/ sec.)

11.2. Electron-hole pairs are created in an intrinsic GaAs semiconductor (at 0 K) by the absorption of a focused argon-ion laser at 5145 Å. Assume a volumetric pumping rate of 10^3 W/cm^3 and that every absorbed photon creates one electron-hole pair. Assume that the electron-hole pairs are generated by the ion laser, recombine according to (11.4.23) with $\beta = 2 \times 10^{-10}$ cm^3/ sec and come into an equilibrium such that $\partial/\partial t = 0$. (Since $T = 0$, all of the carriers are created by the optical pump.)

(a) What is the steady state density of the electrons (or holes)? (Ans.: 3.6×10^{15}/cm^{+3}.)

(b) What is the spacing $F_n - E_c$ (in eV)? (Ans.: 1.28×10^{-3} eV.)

11.3. Show that (11.4.12c) is the same as (11.4.12a).

11.4. Show that the requirement of $f_n(E_2) > f_v(E_1)$ reduces to $F_n - F_p > h\nu$.

11.5. An intrinsic GaAs semiconductor is irradiated by a wave with $h\nu - E_g = 0.05$ eV. Assume k conservation and 0 K and ignore light holes ($E_g = 1.43$ eV).

(a) Identify the energy levels in the conduction and valence band that can participate in absorption or gain (i.e., find $E_2 - E_c$ and $E_v - E_1$). (Ans.: $\Delta E_c = 0.044$ eV; $\Delta E_v = 0.0054$ eV.)

(b) What is the minimum number of electron-hole pairs necessary to achieve optical gain at the wavelength? (Ans.: $N > 7.4 \times 10^{17}$ cm^{-3}.)

11.6. Show the steps in the derivation of (11.7.11) from (11.7.9a) and (11.7.9b).

11.7. Read the article entitled "Ultra-High Speed Semiconductor Lasers" by K. Y. Lau and A. Yariv, *IEEE J. Quant. Elect. QE-21*, 121-137, 1985, and answer the following questions.

(a) Which figure illustrates the fact that the resonant frequency varies inversely with photon lifetime? (Ans.: Fig. 5.)

(b) Use the data of this figure and some reasonable approximations to estimate the functional dependence of ω on τ_p and compare with theory.

(c) The authors estimate that intrinsic differential gain increases by cooling from $+22°$ to $-50°$C. What is that ratio? (Ans.: $\sim$ 1.8; see p. 123.)

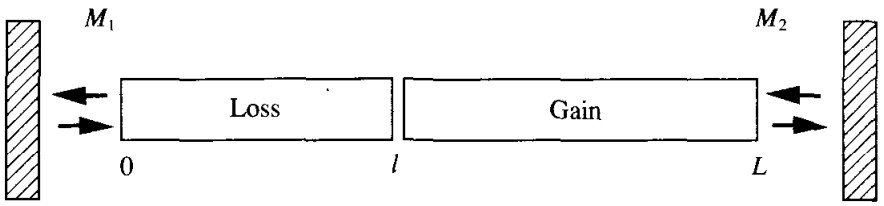
11.8. Consider the following problem sketched on the diagram below, which was inspired by the article by D. Marcuse and F. Nash, "Computer Model of an Injection Laser with Asymmetrical Gain Distribution," *IEEE J. Quant. Electr. QE-18*, 30-43, 1982.

(a) If α is the power loss coefficient over the length 0 to l with a gain coefficient of γ from l to L, what is the ratio of the power emerging from M_2 compared to that from M_1? (Express your answer in terms of mirror reflectivities $R_{1,2}$ alone.)

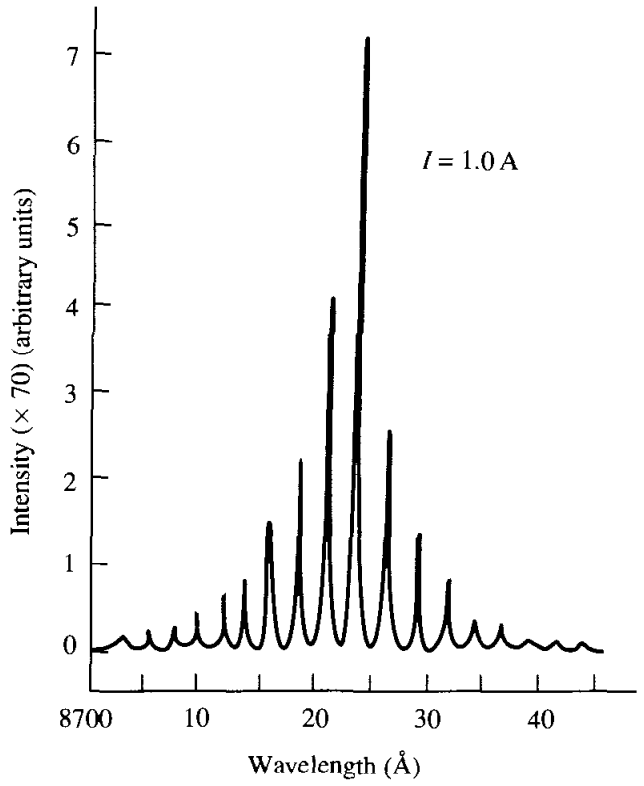
(b) Does this law depend on a specific saturation law or assumption of spatial uniformity?

(c) What do the authors mean by a "soft" transition?

(d) At the end of Appendix I, the authors suggest a means for including spontaneous emission in the formulation of the equation for the laser power. Carry out this suggestion for the situation shown on the diagram below.



11.9. The following is the mode structure obtained from a diode laser 380 μm long. (From T. Paoli, *IEEE J. Quant. Electr. QE-9*, 267, 1973.)



(a) Find the effective index of refraction defined by

$$n(\lambda) - \lambda \partial n/\partial \lambda = n_{\text{eff}}$$

(b) What is the value of the dispersion term, $\partial n/\partial \lambda$ if $n = 3.6$?

11.10. Assume that the radiation field from the laser sketched in Fig. 11.10 is given by

$$\mathbf{E} = E_0 \exp[-(y/w_y)^2] \exp[-(x/w_x)^2]\mathbf{a}_y$$

where x is perpendicular to the junction, y is parallel, $w_x = 1\,\mu$m, and $w_y = 10\,\mu$m.

(a) If the total power radiated by this laser were 100 mW, what is the value of E_0 (in V/cm)?

(b) Assume that the index of refraction of the semiconductor were 3.6 and that the standard transmission line formulas for the reflection or transmission coefficients were applicable (they are, almost). Find the field in the device at the output facet.

(c) Find the angles $\theta_\perp$ and $\theta_\parallel$ for this laser.

11.11. Read and report on the paper by K. Y. Liou, et al., *Appl. Phys. Lett. 50*, 380, 1987.

(a) The authors used a scanning Fabry-Perot interferometer to measure the mode hopping. If we assume the laser mode is a δ function in frequency, what is the finesse of the instrument used?

(b) Explain the difference between Figs. 2a and 2b. Why did the sharp peaks degrade?

(c) What is the frequency change (in Hz) when the diode current is changed from 55 to 57 mA?

11.12. The following questions refer to "Fine Structure of Frequency Chirping and FM Sideband Generation of a Single-Longitudinal-Mode Semiconductor Laser under 10 GHz Direct Intensity Modulation," by C. Lin, et al., *Appl. Phys. Lett. 46*, 12, 1985.

(a) What are the major points covered in this paper?

(b) Why is the FM spectra "intrinsic" to a directly modulated semiconductor laser?

(c) Show that the carrier should be depressed to zero at an FM index of 2.4 (cf. paragraph 2, p. 13). Show, by a sketch, what the spectrum would be for an index of 5.

(d) If the carrier density varied as $N_0 + \Delta N \cos \omega t$, derive a formula for the variation of the refractive index as a function of time.

11.13. The dimensions in a semiconductor may become so small that diffusion of the carriers away from the point of birth can become a problem. The dynamics of the carrier generation and loss become

$$\frac{dN}{dt} = G - \beta N^2 - \frac{N}{\tau_D}$$

where τ_D = diffusion lifetime = 2.6 ns

β = electron-hole recombination rate = 2×10^{-10} cm^3/sec

(a) What must be the carrier concentration for the recombination rate to exceed the loss by diffusion? (Ans.: $N > 1.9 \times 10^{18}$ cm^{-3}.)

(b) If a carrier density of 3×10^{18} cm^{-3} is to be maintained by optical pumping by an argon laser at 5145 Å, and each absorbed photon creates 1 electron-hole pair, what must be the absorbed power (per unit of volume) from the pump? (Ans.: $p = 1.14 \times 10^9$W/cm^3.)

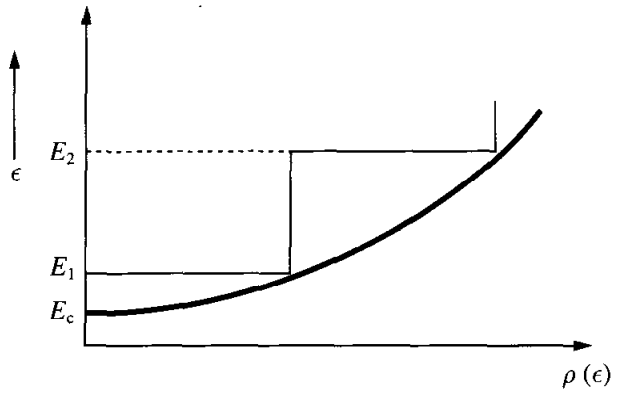
11.14. The density of states in the conduction band of a semiconductor with all dimensions large compared to a deBroglie wavelength is given by

$$\rho(\epsilon) = \frac{1}{2\pi^2} \cdot \left[\frac{2m^*}{\hbar^2}\right]^{3/2} (\epsilon - E_c)^{1/2} \qquad \epsilon > E_c$$

If one dimension is small, say, $L_z = 100$ Å, the density of states in the first allowed band becomes

$$\rho(\epsilon) = \frac{1}{2\pi^2} \cdot \left[\frac{2m^*}{\hbar^2}\right] \cdot \frac{\pi}{L_z} \qquad \epsilon > E_1 > E_c$$

where E_1 results from the quantization of the z component of the momentum, $k_z \cdot L_z = \pi$. This yields the stair-step density of states shown on the diagram below.



(a) If $L_z = 100$ Å, find the energy spacing $(E_1 - E_c)$ and $(E_2 - E_c)$ assuming $m_e^* = 0.067m_0$. (Ans.: $E_1 - E_c = 56.3$ meV.)

(b) Suppose electron densities of 5×10^{17} cm^{-3} were maintained in the conduction bands of both a "large" and a "small" slab of semiconductor. Find the energy interval over which the electrons are distributed (assume $T = 0$ K for the two cases). (Ans.: $\Delta E = 34.3$ meV and 17.89 meV.)

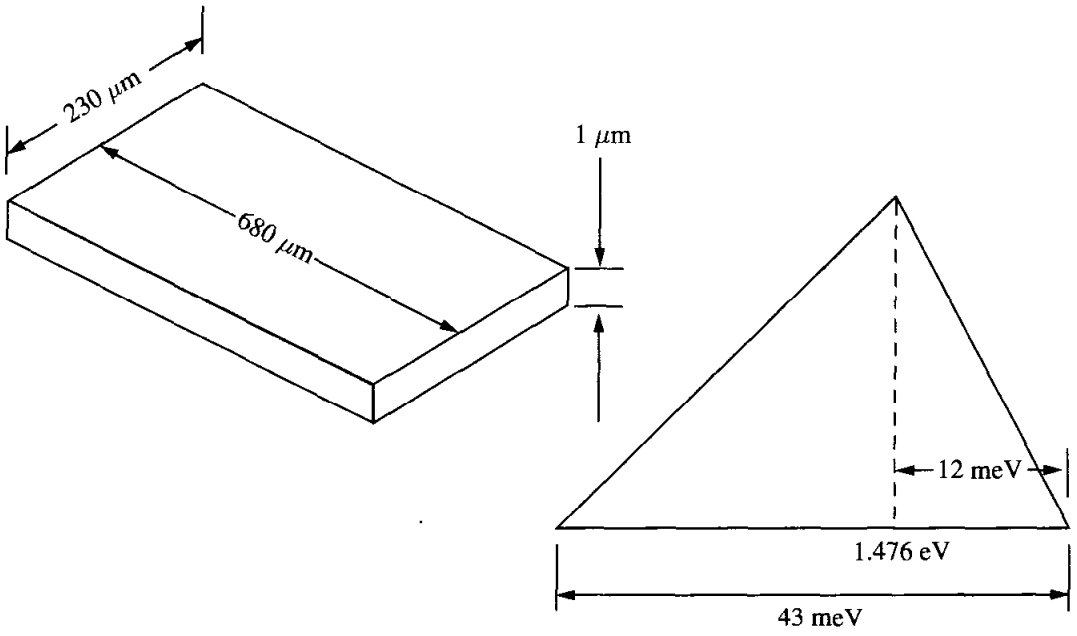
(c) The allowed energies in the valence band are given by the same relation as used in (a). Evaluate the spacing $E_v - E_1(lh)$ or $E_1(hh)$ for $m_{lh}^* = m_e^*$ and $m_{hh}^* = 0.55m_0$. (Ans.: $hh = 6.86$ meV and $lh = 56.3$ meV.)

(d) Use the results of (a) and (c) to find the long wavelength limit of the recombination radiation between the first allowed state in the conduction band and the first heavy-hole state, assuming $E_c - E_v = 1.43$ eV. (Ans.: 1.49 eV.)

(e) Assume a recombination lifetime of 1 ns, approximate the frequency distribution of the transition by the reduced density of filled states found in (d), and

find the gain coefficient of both a "large" and a "small" semiconductor with $n = p = 5 \times 10^{17}$ cm^{-3} and $T = 0$ K. (Ignore light holes.)

11.15. The spontaneous emission from a GaAs semiconductor laser can be approximated by the graph shown below. The length of the wafer is 680 μm, the index of refraction is 3.6, the facet reflectivity is 0.3, the residual absorption coefficient in the crystal is 10 cm^{-1}, and the recombination lifetime is 1 ns.



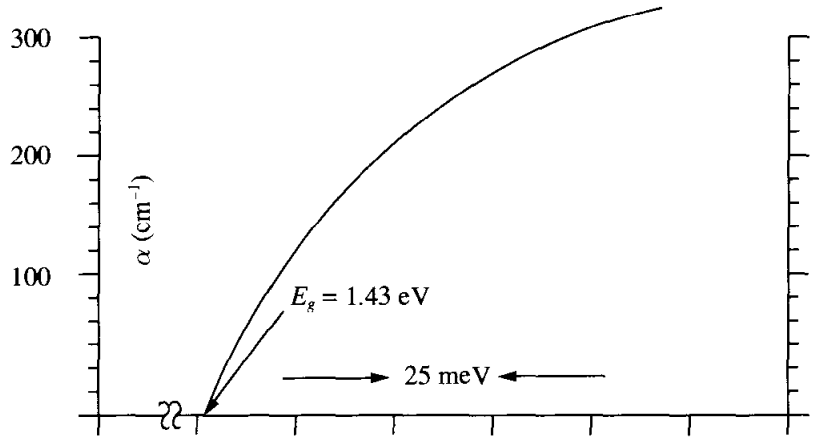
(a) What is the wavelength of peak gain? (Ans.: 0.84 μm.)

(b) What is the FWHM of the gain coefficient in Hz and cm^{-1}? (Ans.: 5.1×10^{12}Hz and 169 cm^{-1})

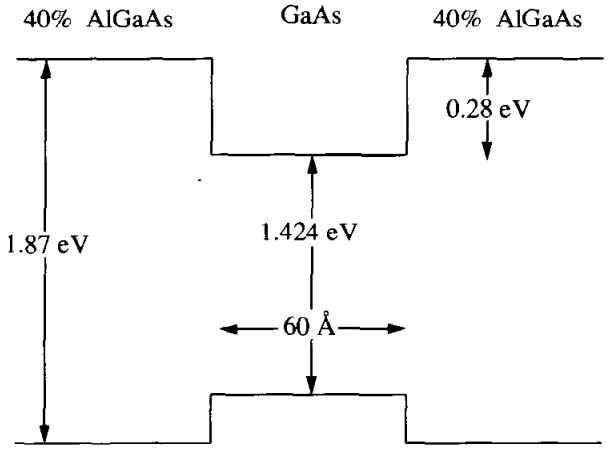
(c) What must be the inverted carrier density to bring this wafer to threshold? (Ans.: 6.5×10^{15} cm^{-3})

(d) This carrier density must be sustained by some sort of pumping—carrier injection, photo pumping, or E-beam excitation. Estimate the minimum pumping power to maintain an inversion of 10^{16} cm^{-3} throughout the wafer. (Ans.: 369 mW.)

11.16. The graph below is the absorption coefficient of a semiconductor at $T = 0$ K. If that sample is photopumped such that $F_n - E_c = 0.050$ eV and $E_v - F_p = 2$ meV, find the peak gain coefficient and the photon energy at which it occurs.



11.17. The diagram below indicates a 60 Å quantum well GaAs layer embedded between two $Al_{0.4}Ga_{0.6}As$ confining layers with the band gaps of the bulk materials given.



(a) Compute the energy (in eV) and wavelength (in cm^{-1} and Å units) corresponding to the transition between the first allowed state in the conduction band and the heavy hold state in the valence band, assuming that the barrier is infinite (even though this is *not* correct): $m_e = 0.067m_0$ and $m_{hh} = 0.55m_0$

(b) What would be the energy of the second allowed electron state, still assuming an infinite barrier, and would that state be confined in the quantum well?

(c) The actual energy states are closer to the band edges. Give a physical explanation why this is true *without* solving the particle-in-a-box problem. (Hint: Sketch the wave function and indicate the relationship of the momentum vectors in the two cases.)

11.18. Redo the calculations in Sec. 11.6.2 for the case of $L_z = 50$ Å, and evaluate the following quantities. For these calculations use $m_e^* = 0.067m_0$ and $m_{hh}^* = 0.55m_0$.

(a) How many electron states are confined in the conduction band well?

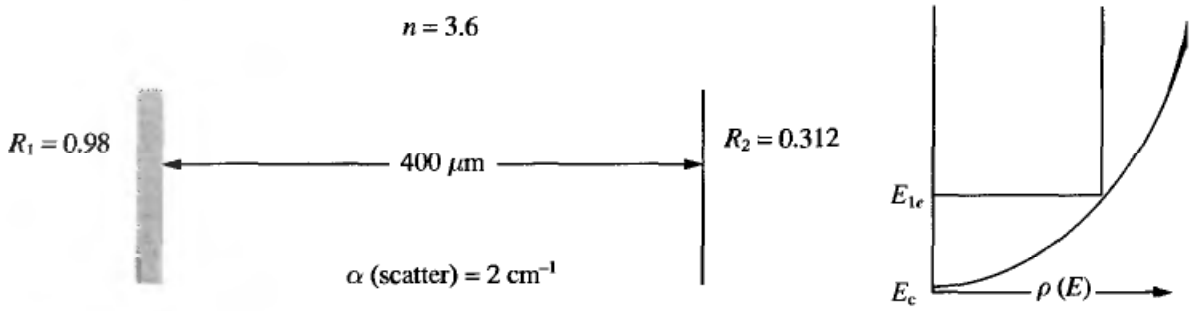
(b) What is the energy of the first confined electron state in the conduction band well?

(c) What is the energy of the first confined hole state in the valence band well?

(d) What is the wavelength of the transition from the bottom of the first state in the conduction band to the first state in the valence band?

(e) How many electron states (per unit of volume) are available to be filled between E_{1e} and E_{c2}? (Assume $T = 0$ for this calculation, but use your values found above.)

11.19. Consider the semiconductor quantum well laser shown in the diagram below along with the density of states diagram for the conduction band. Oscillation takes place in the wavelength region around 8500 Å.



(a) What is the photon lifetime for this cavity?

(b) What is the separation (in Å) between the longitudinal modes?

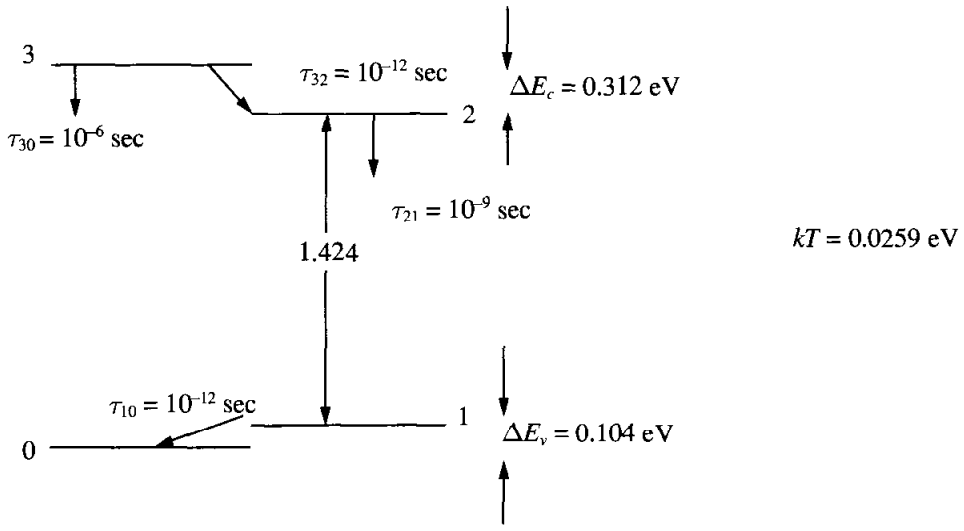
(c) What is the threshold gain coefficient for a mode in this cavity?

(d) Indicate on the density of states diagram where the quasi-Fermi level F_n must be in order to obtain gain.

(e) Assume $T = 0$ K and an electron density of 10^{18} cm^{-3}, $m_e^* = 0.067m_0$, $L_z = 100$ Å, and $x_{1e} = 1.13$. What is the separation $F_n - E_1$ in meV?

(f) Assume that the equal electron and hole densities of (e) were created by photopumping at 5145 Å and are lost by recombination (at a rate βnp, $\beta = 2 \times 10^{-10}$ cm^3/ sec) and by diffusion (n/τ_D, $\tau_D = 10$ ns). What must be the absorbed power per unit of volume to maintain this electron-hole population?

11.20. Consider the following four-level system. Any possible resemblance to a quantum well laser is intentional, but semiconductor theory is at most incidental. The system can be pumped via the $0 \rightarrow 3$ route. State 3 can relax back to 0 at a rate of 10^6 sec^{-1} or to state 2 at a rate of 10^{12} sec^{-1} and can receive population back from 2 at a rate of 5.87×10^6 sec^{-1}. State 2 decays to 1 at a rate of 10^9 sec^{-1} and 1 relaxes to 0 at a rate of 10^{12} sec^{-1}. Thermal processes keep the populations in (3, 2) and in (1, 0) related by the Boltzmann factor involving the appropriate energies. All states have a degeneracy equal to 2. In the absence of pumping, the absorption coefficient on the $2 \rightarrow 1$ transition is 20 cm^{-1}, and the density of active atoms is 10^{20} cm^{-3}.



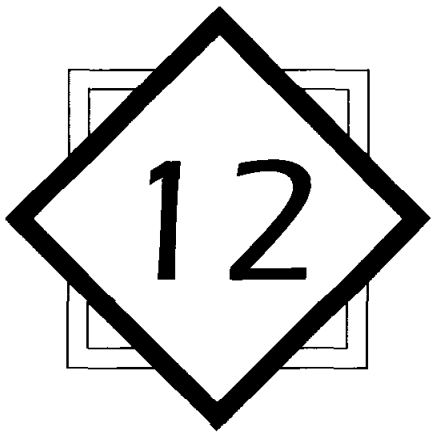
(a) In the absence of pumping, what is the population in state 1?
(b) What is the absorption cross section on the 2 → 1 transition?
(c) What is the ratio N_3/N_2?
(d) What must be the density of atom in state 2 to reach optical transparency?
(e) How much pump power must be expended on the 0 → 3 route to maintain a population in state 2 of $2 \times 10^{+18}$ cm^{-3}?

REFERENCES AND SUGGESTED READINGS

1. It is interesting to note that three separate research groups obtained injection lasers almost simultaneously:
   (a) M. I. Nathan, W. P. Dumke, G. Burns, F. H. Dills, and G. Lasher, "Stimulated Emission of Radiation from GaAs *p-n* Junctions," *Appl. Phys. Lett. 1*, 62, 1962.
   (b) T. M. Quish, R.J Keyes, W. E. Krag, B. Lax, A. L. McWhorter, R. H. Redike, and H. J. Zeiger, "Semiconductor Maser of GaAs," *Appl. Phys. Lett. 1*, 91, 1962.
   (c) N. Holonyak, Jr. and S. F. Bevacqua, "Coherent (Visible) Light Emission from $Ga(As_{1-x}P_x)$," *Appl. Phys. Lett. 1*, 82, 1962.
2. See the Special Issue on Semiconductor Lasers in *IEEE J. Quant. Electron. QE-21*, No. 6, 1985. This contains 38 separate papers on various topics such as single frequency, high-speed operation, optical amplifiers high power, quantum well lasers, novel structures spectra, modeling and noise, and electron-hole recombination.
3. See also another Special Issue on Semiconductor Lasers, *IEEE J. Quant. Electron. QE-23*, No. 6, 1987. The editorial, introduction and historical papers are especially recommended for easy reading. The first paper in the AlGaAs and Visible Lasers: "Optimization and Characterization of Index-guided Visible AlGaAs/GaAs Graded Barrier Quantum Well Laser Diodes," L. J. Maust, M. E. Givens, C. A. Zmudzinski, M. A. Emanuel, and J. J. Coleman, was the basis for the discussion of Fig. 11.17.
4. K. Y. Lau and A. Yariv, "Ultra-High Speed Semiconductor Lasers," *IEEE J. Quant. Electron. QE-21*, No. 2, 1985. This was the basis of Section 11.8.
5. See *IEEE J. Quant. Electron. QE-22*, No. 9, 1986. Special Issue on Semiconductor Quantum Wells and Superlattices: Physics and Applications.

6. G. H.B. Thompson, *Physics of Semiconductor Laser Devices* (New York: John Wiley & Sons, 1980).
7. H. K. Kressel and J. K. Butler, *Semiconductor Lasers and Heterojunction LED* (New York: Academic Press, 1977).
8. S. M. Sze, *Semiconductor Devices, Physics and Technology* (New York: John Wiley & Sons, 1985). See also, S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (New York: John Wiley & Sons, 1981), Chap. 12.
9. K. A. Jones, *Introduction to Optical Electronics* (New York: Harper & Row, 1987).
10. R. G. Hunsperger, *Integrated Optics, Theory and Techniques*, 2nd ed. (New York: Springer-Verlag, 1984), Chaps. 10–14.
11. J. I. Pankove, *Optical Processes in Semiconductors* (Englewood Cliffs, N.J.: Prentice Hall, 1971.)
12. D. Botez and G. J. Herskowitz, "Components for Communication Systems: A Review," *Proc. of IEEE 68*, 689-731, June 1980.
13. N. Holonyak, Jr., R. Kolbas, R. D. Dupris, and P. D. Dapkus, "Quantum Well Heterostructure Lasers," *IEEE J. Quant. Electron QE-16*, 170-180, 1980.
14. B. G. Streetman, *Solid State Electronic Devices*, 2nd ed. (Englewood Cliffs, N.J.: Prentice Hall, 1980), Chap. 10.
15. W. Shockley, *Electrons and Holes in Semiconductors* (Princeton, N.J.: D. Van Nostrand, 1950).
16. M. G. Bernard and G. Duraffourg, "Laser Conditions in Semiconductors," *Phys. Status Solidi 1*, 699, 1961.
17. W. P. Dumke, "Interband Transitions and Maser Action," *Phys. Rev. 127*, 1559, 1962.
18. F. Stern, "Semiconductor Lasers: Theory," Vol. 1, Article B4, p. 425a; and H. Kressel, "Semiconductor Lasers: Devices," in *The Laser Handbook*, Vol. 1, Article B5, p. 441, Eds. F. T. Arecchi and E. O. Schulz-DuBois, (New York: North Holland, 1972).
19. H. C. Casey, Jr. and F. Stern, "Concentration Dependent Absorption and Spontaneous Emission in Heavily Doped GaAs," *J. Appl. Phys. 47*, 631, 1976.
20. G. Lasher and F. Stern, "Spontaneous and Stimulated Recombination Radiation in Semiconductors," *Phys. Rev. 133*, A 553, 1964.
21. Zh. I. Alferov, V. M. Andreev, V. I. Korol'koy, E. L. Portnic, and D. N. Tret'yakov, "Injection Properties of Ω-$Al_xGa_{1-x}AsP$/GaAs Heterojunctions," *Friz. Tekh, Poluprov, 2*, 1016, 1968. See also *Sov. Phys. Semicond. 2*, 843, 1969.
22. (a) H. C. Casey, Jr. and M. B. Panish, *Heterostructure Lasers* (New York: Academic Press, 1978).
    (b) H. C. Casey, Jr. and M. B. Panish, "Composition dependence of the $GA_{1-x}Al_xAs$ direct and indirect energy gaps," *J. Appl. Phys., 40*, 4910, 1969.
    (c) H. C. Casey, Jr., D. D. Sell, and M. B. Panish, "Refractive index of $Al_xGa_{1-x}As$ between 1.2 and 1.8 eV," *Appl. Phys. Lett., 24*, 63–65, 1974.
    (d) D. D. Sell, H. C. Casey, Jr., and K. W. Wecht, "Concentration dependence of the refractive index for n- and p-type GaAs between 1.2 and 1.8 eV," *J. Appl. Phys., 45*, 2650, 1974.
23. C. M. Wolfe, N. Holonyak, Jr., and G. E. Stillman, *Physical Properties of Semiconductors* (Englewood Cliffs, N.J.: Prentice Hall, 1989).
24. Y. H. Wang, K. Tai, J. D. Wynn, M. Hong, R. J. Fischer, J. P. Mannaerts, and A. Y. Cho, "GaAs/AlGaAs Multiple Quantum Well GRIN-SCH Vertical Cavity Surface Emitting Laser Diodes," *IEEE Photonics Technol. Lett. 2*, 456–458, 1990.

25. R. P. Schneider, Jr., R. P. Bryan, J. A. Lott, and G. R. Olbright, "Visible (657 nm) InGaP/InAlGaP Strained Quantum Well Vertical-Cavity Surface-Emitting Laser," *Appl. Phys. Lett. 60*, 1830–1832, 1992.
26. T. Erdogan, O. King, G. W. Wicks, D. G. Hall, E. H. Anderson, and M. J. Rooks, "Circularly Symmetric Operation of a Concentric-Circle-Grating, Surface-Emitting, AlGaAs/GaAs Quantum-Well Semiconductor Laser," *Appl. Phys. Lett. 60*, 1921–1923, 1992.
27. C. Lei and D. G. Deppe, "Optical Gain Enhancement in Fabry-Perot Microcavity Lasers," *J. Appl. Phys. 71*, 2530-2535, 1992.
28. C. J. Chang-Hasnain, M. Orenstein, A. Von Lehmen, L. T. Florez, J. P. Harbison, and N. G. Stoffel, "Transverse Mode Characteristics of Vertical Cavity Surface-Emitting Lasers," *Appl. Phys. Lett. 57*, 218–220, 1990.
29. B. Tel, Y. H. Lee, K. F. Brown-Groebele, J. L. Jewell, R. E. Leibenguth, M. T. Asom, G. Livescu, L. Luther, and V. D. Mattera, "High-Power CW Vertical-Cavity Top Surface-Emitting GaAs Quantum Well Lasers," *Appl. Phys. Lett. 57*, 1855-1857, 1990.
30. T. J. Rogers, D. G. Deppe, and B. G. Streetman, "Effect of an AlAs/GaAs Mirror on the Spontaneous Emission of an InGaAs-GaAs Quantum Well," *Appl. Phys. Lett. 57*, 1858–1860, 1990.
31. D. G. Deppe and C. Lei, "Spontaneous Emission From a Dipole in a Semiconductor Microcavity," *Appl. Phys. 70*, 3443–3448, 1991.
32. T. Yamauchi, Y. Arakawa, and M. Nishioka, "Enhanced and Inhibited Spontaneous Emission in GaAs/AlGaAs Vertical Microcavity Lasers With Two Kinds of Quantum Wells," *Appl. Phys. Lett. 58*, 2339–2341, 1991.
33. J. L. Jewell, K. F. Huang, K. Tai, Y. H. Lee, R. J. Fischer, S. L. McCall, and A. Y. Cho, "Vertical Cavity Single Quantum Well Laser," *Appl. Phys. Lett. 55*, 424–426, 1989.
34. K. Lai, L. Yang, J. D. Wynn, and A. Cho, "Drastic Reduction of Series Resistance in Doped Semiconductor Distributed Bragg Reflectors for Surface Emitting Lasers," *Appl. Phys. Lett. 56*, 2496–2498, 1990.
35. K. Tai, R. J. Fischer, A. Cho, and K. F. Huang, "High-Reflectivity $AlAs_{0.52}Sb_{0.48}$/GaInAs(P) Distributed Bragg mirror on InP Substrate for 1.3–1.55 μm Wavelengths," *Electron. Lett. 25*, 1159–1160, 1989.
36. J. L. Jewell, J. P. Harrison, A. Scherer, Y. H. Lee, and L. T. Florez, "Vertical Cavity Surface Emitting Laser: Design, Growth, Fabrication, Characterization," *IEEE J. Quantum Electron. QE-27*, 1332–1346, 1991.



Advanced Topics in Laser Electromagnetics

12.1 INTRODUCTION

The electromagnetic aspects of lasers are very important. Indeed we could argue that these are the only ones that can be affected by the design since the atomic physics aspects were fixed at the moment of creation.* The quality of the cavity affects the magnitude of the fields, which, in tum, affects the stimulated emission rate, and thus the electromagnetics play a very pervasive part in lasers.

The earlier chapters on open resonators gave a simple introduction to terminology but barely scratched the surface of electromagnetic problems in various lasers. For instance, we can list some very practical issues and questions about various lasers that are basically electromagnetic issues.

1. Semiconductor lasers use an "optical waveguide" to connect two waveguide discontinuities for a cavity. The transverse geometry is definitely not circular, with heterostructures affecting the field locations as well as confining the electron-hole pairs. There is usually a well defined dielectric waveguide in the direction perpendicular to the junction, but no "obvious" reasons why the optical field should be confined

*That is not really true. For instance, the physics of a semiconductor was determined at $t = 0$ (creation), but it took a bit of time to arrive at the heterostructure and quantum well semiconductor laser.

in the plane of the junction other than that provided by the gain. How strong is that confinement?

2. Not all gas, dye, and solid-state lasers use "stable" resonators, and surely not semiconductors. What are the field configurations in such unstable cavities, and what is the expansion law for a non-Hermite Gaussian beam? We cannot use a uniform plane approximation (since it has an infinite cross sectional area).
3. Some lasers (mostly semiconductor ones) use a "distributed feedback" (DFB) scheme with hundreds of partial reflections rather than the simple two mirror scheme. Such DFB lasers have very desirable characteristics but are still difficult to produce. They promise to be the dominant type in telecommunications applications. Our round-trip gain and phase shift specification takes on a deeper meaning for such a geometry. Coupled mode theory will be needed for this analysis.
4. Laser arrays are very desirable. A multiplicative factor of (N) in power can be realized from the N coupled lasers with the theoretical possibility of electronically scanning the array through the N allowed coherent superpositions of the individual modes. The latter has not been accomplished (for good and real reasons), but it is still an intriguing possibility.
5. We need a formal method of addressing electromagnetic problems, one which is more specific than merely stating that the field should obey Maxwell's equations. The method should be based on them, but it must be easily applied to some obvious problems, such as unstable cavities, hole coupled mirrors, tapered mirrors, or propagation of an arbitrary beam.

Thus, this chapter addresses the more practical and therefore harder electromagnetic problems of lasers. Since the semiconductor laser seems predestined to become dominant, problems associated with it are addressed first.

12.2 SEMICONDUCTOR CAVITIES

The cavities used for semiconductor lasers are characterized by transverse dimensions comparable to the wavelength being amplified, and hence a waveguide approach is most often used in the analysis. The vertical geometry of a typical stripe laser is shown in Fig. 12.1. The injected current is usually confined to a "stripe" whose width is much larger than the depth over which the recombination of the carriers takes place.

As mentioned in Chapter 11, we can substitute Al for Ga in the GaAs lattice, which raises the band gap and simultaneously lowers the index of refraction but keeps the crystal structure more or less identical. For various choices of x in $Al_xGa_{1-x}As$ and various layer thicknesses, we can have from 0 to 4 heterojunctions for the simultaneous confinement of the injected carriers and the electromagnetic wave. Most of the essential features of the wave guidance can be obtained by analyzing the simpler case of a three region waveguide, named 1, 3, and 5 for historical reasons, shown in Fig. 12.1. We search for fields that propagate in the $\pm z$ direction as $\exp \mp(\gamma z)$ and obey the wave equation. There are two general classes

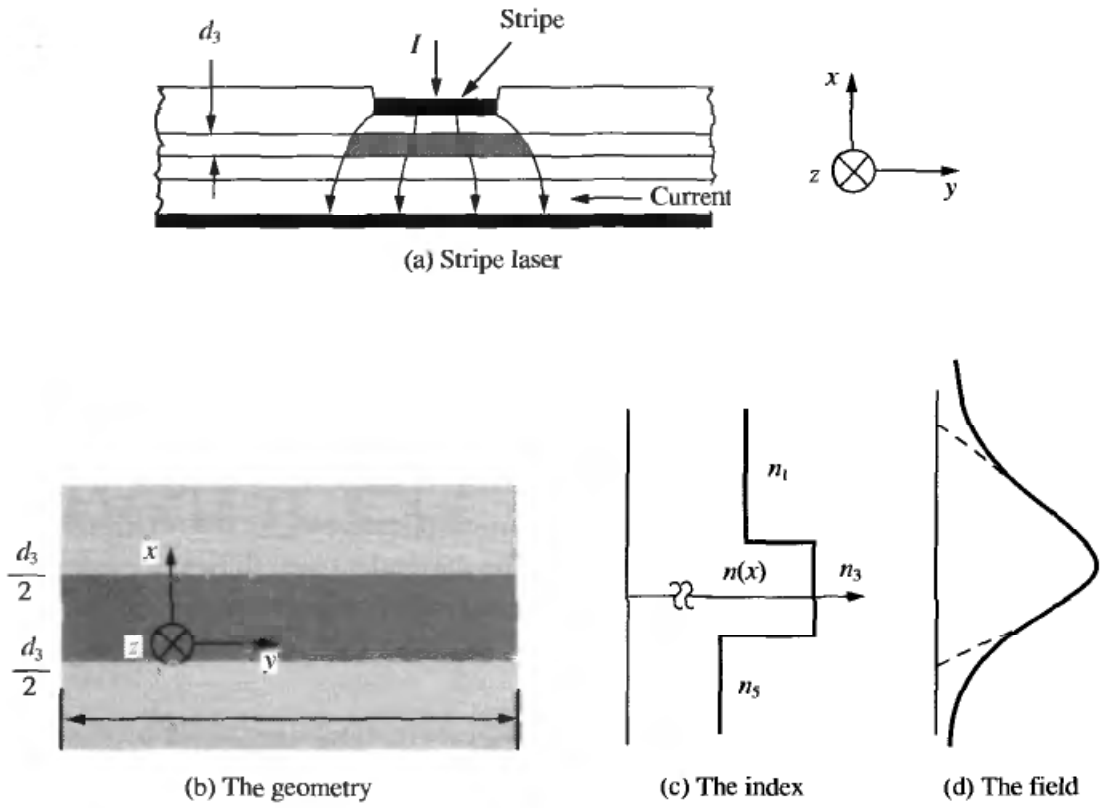


FIGURE 12.1. (a) The geometry of a stripe laser. (b) The asymmetric slab waveguide representative of many semiconductor lasers. (c) A sketch of the mode in the slab.

of such fields, TE or TM, which are of primary interest:

$$\nabla_t^2 E_z + \left[\gamma^2 + \frac{\omega^2}{c^2} n^2(x, y)\right] E_z = 0 \qquad (H_z = 0 \text{ or TM}) \qquad (12.2.1a)$$

$$\nabla_t^2 H_z + \left[\gamma^2 + \frac{\omega^2}{c^2} n^2(x, y)\right] H_z = 0 \qquad (E_z = 0 \text{ or TE}) \qquad (12.2.1b)$$

where

$$\nabla_t^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

The index of refraction can be continuous functions of x and y as implied by (12.2.1) or discontinuous as shown in Fig. 12.1.

Both types of modes obey the same type of scalar wave equation, and thus some general conclusions can be obtained by examining the characteristics of the solution for the various regions. We, or course, are primarily interested in the case $\gamma = j\beta$, that is, wave motion in the z direction with size of the mode being bounded in the transverse directions. If we neglect the variation of the fields in the y direction for a moment, then (12.2.1) can

be expressed as

$$\frac{\partial^2 \Psi}{\partial x^2} + \left[\frac{\omega^2}{c^2} n^2(x, y) - \beta^2\right] \Psi = 0 \tag{12.2.2}$$

This equation states a very important point about either mode if the desired goal of obtaining a *guided* field distribution of such lasers is remembered. We prefer a high peak transverse electric field in the immediate vicinity of the recombination region d_3 and then dropping to zero for $|x| \to \infty$. This characteristic is achieved if $\beta^2 < (\omega n_3/c)^2$ in the central region, which results in a standing wave or trigonometric type of solution there, and $\beta^2 > (\omega n_{1,5}/c)^2$ for the exterior regions, which leads to the exponential tails on the modes. Thus the phase constant β divided by $\omega/c = k_0$ lies between two extremes for these desired guided (i.e., nonradiating) modes.

$$\boxed{n_{1,5} < \frac{\beta}{k_0} < n_3} \qquad \text{for guided modes} \tag{12.2.3}$$

Values of β/k_0 less than n_1 or n_5 (or both) lead to waves propagating or radiating in the x direction and thus are not modes guided along z and suitable for the laser.

Although the presence or absence of a z component of an electric or magnetic field characterizes the mode as TM or TE, respectively, most prefer to work with the transverse components E_x or E_y (or H_y, H_x), which are more easily visualized in terms of intensity. The relationships between the transverse fields and the z component were given earlier (4.3.1) and are repeated below using γ instead of $j\beta$ for the propagation constant and $k = \omega n/c$.

$$\mathbf{E}_t = -\frac{1}{\gamma^2 + k^2}[\gamma \nabla_t E_z - j\omega\mu\, \mathbf{a}_z \times \nabla_t H_z] \tag{12.2.4a}$$

$$\mathbf{H}_t = \frac{1}{\gamma^2 + k^2}[-j\omega\epsilon_0 n^2 \mathbf{a}_z \times \nabla_t E_z - \gamma \nabla_t H_z] \tag{12.2.4b}$$

The following analysis closely parallels that of a symmetric slab of Chapter 4, but now we face up to the fact that seldom is $n_1 = n_5$, which makes the arithmetic somewhat tedious. However, the importance of such lasers dictates that the effort be expended to extract all the information possible for the asymmetric slab typical of the lasers.

12.2.1 TE Modes ($E_z = 0$)

For fields uniform in the y direction in Fig. 12.1 ($\partial/\partial y = 0$), the component E_y satisfies

$$\frac{\partial^2 E_y}{\partial x^2} + \left(\gamma^2 + \frac{\omega^2}{c^2} n_{1,3,5}^2\right) E_y = 0 \tag{12.2.5}$$

in each of the three regions. We anticipate a wave-guided z along with $\gamma = j\beta$ and its amplitude vanishing at $|x| \to \infty$. Thus

$$E_y^{(1)} = A_1 \exp\left[-h_1\left(x - \frac{d_3}{2}\right)\right] \tag{12.2.6a}$$

$$E_y^{(3)} = A_3 \cos h_3 x + B_3 \sin h_3 x \tag{12.2.6b}$$

$$E_y^{(5)} = A_5 \exp\left[+h_5\left(x + \frac{d_3}{2}\right)\right] \tag{12.2.6c}$$

where

$$h_1^2 = \beta^2 - (\omega n_1/c)^2 \tag{12.2.7a}$$

$$h_3^2 = (\omega n_3/c)^2 - \beta^2 \tag{12.2.7b}$$

$$h_5^2 = \beta^2 - (\omega n_5/c)^2 \tag{12.2.7c}$$

If β/k_0 satisfies the inequality of (12.2.3), then the right-hand sides of (12.2.7a) to (12.2.7c) are positive real.

Equations (12.2.6) can be rewritten to emphasize the continuity of the fields at the boundaries $x = \pm d_3/2$.

$$E_y^{(1)} = A\left\{\cos\left(\frac{h_3 d_3}{2} - \phi\right)\exp\left[-h_1\left(x - \frac{d_3}{2}\right)\right]\right\} \tag{12.2.8a}$$

$$E_y^{(3)} = A[\cos(h_3 x - \phi)] \tag{12.2.8b}$$

$$E_y^{(5)} = A\left\{\cos\left(\frac{h_3 d_3}{2} + \phi\right)\exp\left[+h_5\left(x + \frac{d_3}{2}\right)\right]\right\} \tag{12.2.8c}$$

This form also emphasizes the fact that the field need not be "centered" at $x = 0$ either as a symmetric or an asymmetric function. A sketch of the E_y field as a function of x is shown in Fig. 12.2.

Although the form of (12.2.8) guarantees the continuity of E_y at $x = \pm d_3/2$, one must also match the tangential components of $H_z^{(1,3,5)}$ along these same planes. Combining

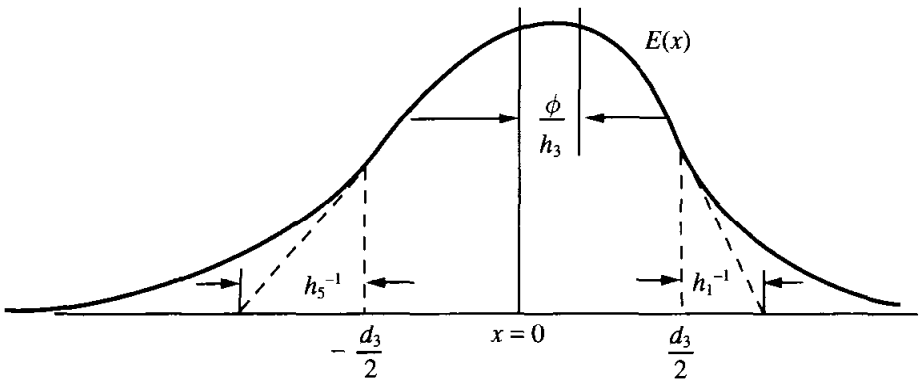


FIGURE 12.2. A sketch of the dominant TE_0 mode.

(12.2.4b) and (12.2.1b) yields

$$H_z = -\frac{1}{j\omega\mu_0}\frac{\partial E_y}{\partial x} \tag{12.2.9}$$

Thus

$$H_z^{(1)} = \frac{h_1}{j\omega\mu_0}\left\{A\cos\left(\frac{h_3 d_3}{2} - \phi\right)\exp\left[-h_1\left(x - \frac{d_3}{2}\right)\right]\right\} \tag{12.2.10a}$$

$$H_z^{(3)} = \frac{h_3}{j\omega\mu_0}[A\sin(h_3 x - \phi)] \tag{12.2.10b}$$

$$H_z^{(5)} = -\frac{h_5}{j\omega\mu_0}\left\{A\cos\left(\frac{h_3 d_3}{2} + \phi\right)\exp\left[+h_5\left(x + \frac{d_3}{2}\right)\right]\right\} \tag{12.2.10c}$$

The continuity of H_z at the boundaries ($x = \pm d_3/2$) leads to

$$\left(\frac{h_3 d_3}{2} - \phi\right) = \tan^{-1}\left(\frac{h_1}{h_3}\right) = A \tag{12.2.11a}$$

$$\left(\frac{h_3 d_3}{2} + \phi\right) = \tan^{-1}\left(\frac{h_5}{h_3}\right) = B \tag{12.2.11b}$$

Thus the propagation constant γ, which is hidden in the h parameters (12.2.7), is determined implicitly by the sum of (12.2.11a) and (12.2.11b) or

$$\tan(h_3 d) = \frac{\tan A + \tan B}{1 - \tan A \tan B} = \frac{h_3(h_1 + h_5)}{h_3^2 - h_1 h_5} \tag{12.2.12}$$

The difference between (12.2.11b) and (12.2.11a) yields an expression for ϕ:

$$\tan 2\phi = \frac{h_3(h_5 - h_1)}{h_3^2 + h_1 h_5} \tag{12.2.13}$$

12.2.2 TM Modes ($H_z = 0$)

The equation for H_x of the TM modes is identical to (12.2.5), and thus its solution has the same format as (12.2.6a) through (12.2.6c) with the same definitions of $h_{1,2,3}$ as given by (12.2.7a) through (12.2.7c). However, different multiplication factors appear in the secular equation (12.2.12) and for ϕ in (12.2.13) because of the different indices of refraction. From (12.2.4)

$$E_z = \frac{1}{j\omega\epsilon_0 n^2}\frac{\partial H_y}{\partial x} \tag{12.2.14}$$

and matching H_y and E_z at $x = \pm d_3/2$, we obtain

$$\tan(h_3 d_3) = \frac{\dfrac{h_3}{n_3^2}\left[\dfrac{h_1}{n_1^2} + \dfrac{h_5}{n_5^2}\right]}{\left(\dfrac{h_3}{n_3^2}\right)^2 - \left(\dfrac{h_1}{n_1^2}\right)\left(\dfrac{h_5}{n_5^2}\right)} \tag{12.2.15}$$

and

$$\tan 2\phi = \frac{\dfrac{h_3}{n_3^2}\left[\dfrac{h_5}{n_5^2} - \dfrac{h_1}{n_1^2}\right]}{\left(\dfrac{h_3}{n_3^2}\right)^2 + \left(\dfrac{h_1}{n_1^2}\right)\left(\dfrac{h_5}{n_5^2}\right)} \tag{12.2.16}$$

12.2.3 Polarization of TE and TM Modes

Although there is a difference in the propagation constant and ϕ between the two configurations, that is a numerical problem rather than a physical phenomenon that is easily detected with a minimum of sophisticated instruments. Both modes are superpositions of plane waves internally reflected at the boundaries at $x = \pm d_3/2$. The distinguishing, easily detected, physical feature is the difference in polarization of the fields as shown in Fig. 12.3. The optical electric field is in the y direction for the TE modes and lies in the xz plane for the TM orientation. Most semiconductor lasers oscillate in the TE modes because the reflectivity of a cleaved facet for that orientation is higher than for the TM case. To compute this reflectivity is a rather formidable task because the radiation field is composed of a distribution of plane waves to match the internal mode at the semiconductor facet. That is a task reserved for more advanced texts. Many use the standard Fresnel formula in conjunction with Fig. 12.3 to estimate the reflectivity. If the angle θ were 0°, then the standard transmission line formula yields:

$$R = \left[\frac{n_3 - 1}{n_3 + 1}\right]^2 \sim 0.32 \text{ for } n_3 = 3.6 \text{ (TE and TM modes)}$$

However, Fig. 12.3 suggests that there is the possibility of Brewster's angle occurring for the TM orientation but not for the TE case. At Brewster's angle ($\theta \cong 16°$ in Fig. 12.3), the reflectivity would be zero if the waves were uniform plane waves. Because the fields are *not* uniform plane waves, zero reflectivity does not occur. But this line of reasoning does

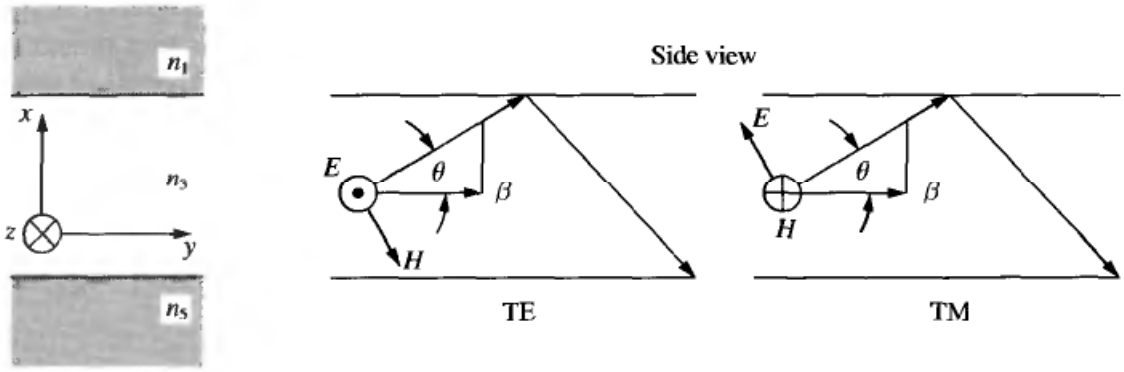


FIGURE 12.3. A comparison between TE and TM slab waveguide modes showing the difference in polarization. The dashes indicate the orientation of a cleaved surface.

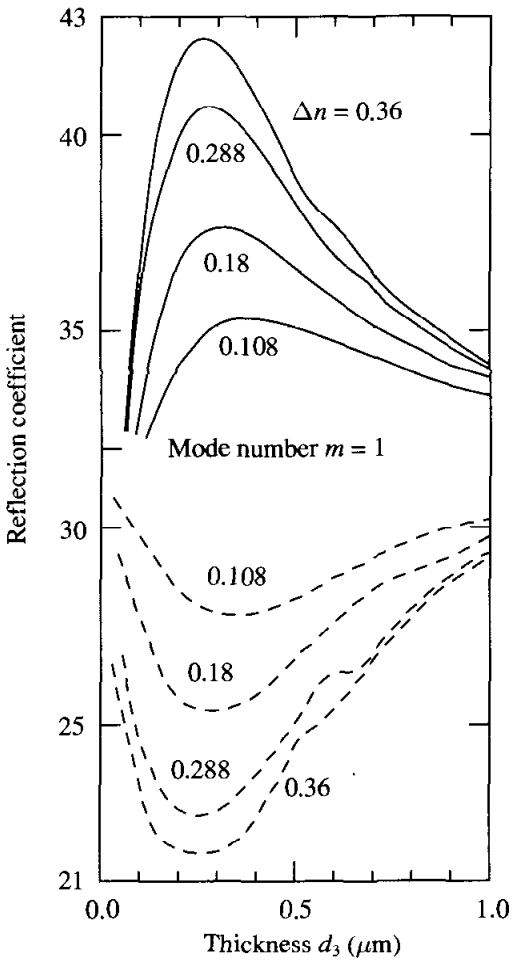


FIGURE 12.4. The facet reflection coefficient for the TE (solid) and TM (dashed) modes. (From Ikegami [19].)

indicate that the TM reflectivity and thus the feedback are always less than those of the TE modes. This is borne out by detailed calculations and is shown in Fig. 12.4.

12.3 GAIN GUIDING: AN EXAMPLE

The previous section is an example of the use of a spatial variation of the index of refraction to create a waveguide to confine the wave that is propagating in the z direction. The index can have discontinuous jumps as was considered previously (12.1) or be a continuous variable, as was considered in Chapter 4 for fibers. For most heterostructure lasers, there is always an index variation in the direction perpendicular to the plane of the junction. Spatial variation of the gain along the plane of the junction can also provide guidance of a mode, and this is the problem to be addressed here. The practical situation is sketched in Fig. 12.5 for a heterostructure laser.

To make the following arithmetic tolerable, we assume the fractional concentrations of aluminum of the confining p and n layers are chosen to generate a smooth variation of dielectric constant in the x direction as sketched in Fig. 12.5(b). Alternatively, we can interpret the parabolic variation as a simple continuous approximation to the discontinuous jumps considered in the previous section, and thus L_x is an adjustable fitting parameter.

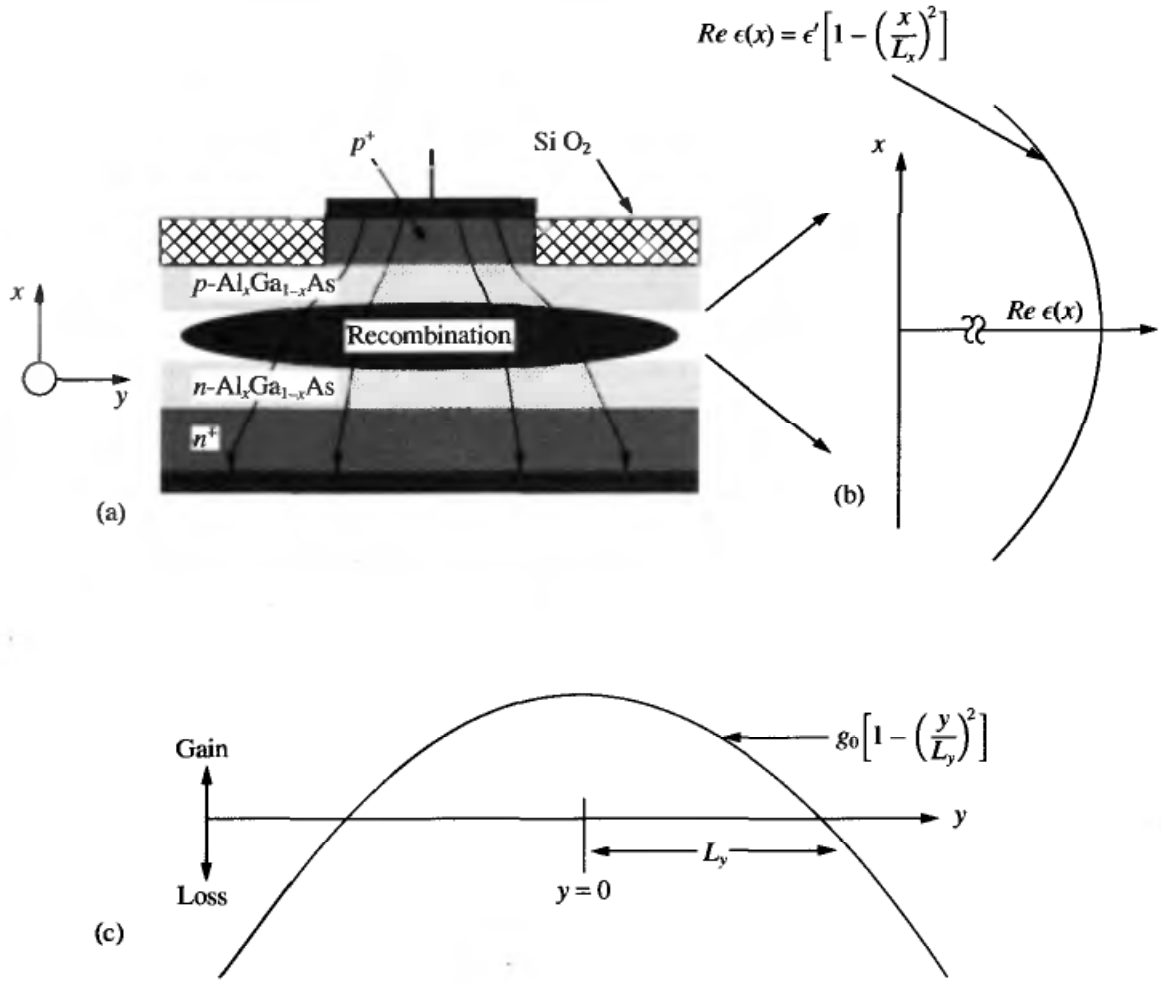


FIGURE 12.5. The assumed variation of the dielectric constant and gain in a heterostructure laser. The dielectric constant is assumed to be uniform with y. The cross-hatch region is to indicate an etched region back filled with SiO_2 to obtain a real index guided laser (see *Sec. 12.4*).

Obviously, the extent of the region along x must be much less than L_x to avoid the absurdity of $\epsilon(x)$ becoming negative.

The amplification is provided by the recombination of the carriers in a very small region near the p-n junction (or in the quantum well) as discussed in Chapter 11. Thus the material gain coefficient does depend upon x, rather drastically so for a quantum well laser. It is large in the very small quantum well and zero elsewhere. However, the variation of the real index along x (or the real part of the dielectric constant) is much more important in the determination of the field variation than is the gain. Thus we assume an effective gain coefficient independent of x, but which can still depend upon the coordinate y, parallel to the junction.

$$g(y) = g_0(1 - y^2/L_y^2) \tag{12.3.1}$$

where we acknowledge the fact that the gain coefficient can be negative owing to the lack of current density at $y > L_y$. It is important to have the variation of the gain with y because

gain provides the only encouragement for the mode to be confined in that direction. We should also realize that both g_0 and L_y are functions of the injected current.

Thus the electromagnetic problem consists of finding solutions to the wave equation with a complex inhomogeneous dielectric constant of the form:

$$\epsilon(x, y) = \epsilon_0 \left[\epsilon' \left(1 - \frac{x^2}{L_x^2} \right) + j\epsilon'' \left(1 - \frac{y^2}{L_y^2} \right) \right] \tag{12.3.2}$$

Hence the wave equation becomes

$$\nabla^2 \begin{Bmatrix} E \\ H \end{Bmatrix} + \frac{\omega^2}{c^2} \left[\epsilon' \left(1 - \frac{x^2}{L_x^2} \right) + j\epsilon'' \left(1 - \frac{y^2}{L_y^2} \right) \right] \begin{Bmatrix} E \\ H \end{Bmatrix} = 0 \tag{12.3.3a}$$

If the medium were uniform and of infinite extent, and if we were dealing with uniform plane waves propagation as $\exp[(g_0/2 - j\beta)z]$, then we would relate the field gain coefficient $(g_0/2)$ and phase constant β to the dielectric parameters (ϵ', ϵ'') according to

$$\beta^2 = \frac{\omega^2}{c^2} \epsilon' = \frac{\omega^2}{c^2} n_r^2 \tag{12.3.4a}$$

$$g_0 \beta = \frac{\omega^2}{c^2} \epsilon'' \tag{12.3.4b}$$

Hence it is appropriate to rewrite (12.3.3a) with those abbreviations:

$$\nabla^2 \begin{Bmatrix} E \\ H \end{Bmatrix} + \left[\beta^2 \left(1 - \frac{x^2}{L_x^2} \right) + j g_0 \beta \left(1 - \frac{y^2}{L_y^2} \right) \right] \begin{Bmatrix} E \\ H \end{Bmatrix} = 0 \tag{12.3.3b}$$

Now we follow a procedure similar to that used in Chapter 3: assume that the fields vary as

$$(E, H) = \psi(x, y, z) e \exp[(g_0/2 - j\beta)z] \tag{12.3.5}$$

and assume

$$\frac{\partial^2 \psi}{\partial z^2} \ll 2j\beta \frac{\partial \psi}{\partial z}$$

as was argued in Chapter 3. The wave function ψ obeys the following equation with the additional assumption

$$|g_0| \ll |\beta|$$

which is always true except for pathological cases.

$$\nabla_t^2 \psi - 2j\beta \frac{\partial \psi}{\partial z} - \frac{\beta^2 x^2}{L_x^2} \psi - j g_0 \beta \frac{y^2}{L_y^2} \psi = 0 \tag{12.3.6}$$

We should not fall into the trap of thinking that (amplified) uniform plane waves have been assumed with a growth of $\exp(g_0 z/2)$ and a phase change of $\exp(-j\beta z)$. Not so. Those terms are factors in the expression for the field but the wave function ψ depends on z, and thus the total modal growth and phase change remain to be determined.

Now the procedure becomes identical (rather than merely similar) to Chapter 3. Assume that ψ can be expressed as

$$\psi(x, y, z) = \left\{\exp\left[-j\left(P_x(z) + \frac{\beta x^2}{2q_x(z)}\right)\right]\right\}\left\{\exp\left[-j\left(P_y(z) + \frac{\beta y^2}{2q_y(z)}\right)\right]\right\} \tag{12.3.7}$$

where this form emphasizes that the wave function is expressed as a separated product of two functions $\psi_x(x, z) \cdot \psi_y(y, z)$. We need not go through the standard steps of separating the variables, because the unknown functions, $P_x(z)$, $P_y(z)$, $q_x(z)$, and $q_y(z)$, are already identified and appear in the exponent. Equation (12.3.7) is substituted into (12.3.6) and terms involving x^2, x^0, y^2, and y^0 are grouped together.

$$\begin{aligned}\left\{\beta^2\left[\frac{q_x' - 1}{q_x^2} - \frac{1}{L_x^2}\right]x^2 - \beta\left[\frac{j}{q_x} + 2P_x'\right]x^0\right.\\ \left. + \beta^2\left[\frac{q_y' - 1}{q_y^2} - j\frac{g_0/\beta}{L_y^2}\right]y^2 - \beta\left[\frac{j}{q_y} + 2P_y'\right]y^0\right\} = 0\end{aligned} \tag{12.3.8}$$

All factors of the powers of x or y must be separately equal to zero, which thus yields ordinary differential equations for the complex beam parameters $q_{x,y}(z)$ and $P_{x,y}(z)$. However, we are interested in a guided mode, not freely expanding waves as were considered in Chapter 3. Hence, we also require that $q_x'(z) = q_y'(z) = 0$; that is, the complex beam parameter q should be independent of z. From the coefficient of x^2 in (12.3.8) (with $q_x' = 0$), we obtain the following:

$$q_x^2 = -L_x^2 \quad \text{or} \quad \frac{1}{q_x} = -j\frac{1}{L_x} \tag{12.3.9a}$$

From the coefficient of x^0

$$P_x'(z) = -j\frac{1}{2q_x} \qquad -jP_x(z) = +j\frac{z}{2L_x} \tag{12.3.9b}$$

From the coefficient of y^2 (with $q_y' = 0$):

$$\frac{1}{q_y^2} = -j\frac{g_0}{\beta L_y^2} \text{ or } \frac{1}{q_y} = \left(\frac{g_0}{2\beta}\right)^{1/2}\frac{1}{L_y}(1 - j1) \tag{12.3.9c}$$

and from the coefficient of y^0

$$P_y'(z) = \frac{-j}{2q_y} \qquad -jP_y(z) = -\left(\frac{g_0}{2\beta}\right)^{1/2}\frac{z}{2L_y} + j\left(\frac{g_0}{2\beta}\right)^{1/2}\frac{z}{2L_y} \tag{12.3.9d}$$

where the signs of all square roots are chosen to ensure that the field vanishes at $\pm\infty$. Now the physical interpretation of the complex beam parameter is the same as that assigned in Chapter 3; that is,

$$\frac{1}{q} \equiv \frac{1}{R} - j\frac{\lambda}{\pi w^2}$$

The significant changes are that we have demanded that the radius of curvature of the phase front R, and the spot size, w, *be independent of* z, and both parameters be allowed to be different in the x and y direction. With this in mind, we can reassemble the field:

$$\begin{aligned}\frac{E(xyz)}{E_0} &= \left\{\exp\left[-\left(\frac{\beta x^2}{2L_x}\right)\right]\right\} \\ &\times \left\{\exp\left[-\frac{\beta y^2}{2L_y}\left(\frac{g_0}{2\beta}\right)^{1/2}\right]\right\} \times \left\{\exp\left[-j\frac{\beta y^2}{2L_y}\left(\frac{g_0}{2\beta}\right)^{1/2}\right]\right\} \\ &\times \left\{\exp -j\left[\beta - \frac{1}{2L_x} - \frac{1}{2L_y}\left(\frac{g_0}{2\beta}\right)^{1/2}\right]z\right\} \\ &\times \left\{\exp\left[\frac{g_0}{2} - \left(\frac{g_0}{2\beta}\right)^{1/2}\frac{1}{2L_y}\right]z\right\} \end{aligned} \tag{12.3.10}$$

There are various parts of this equation that are easily discernible:

1. The first line of (12.3.10) obviously describes a Gaussian beam variation in the x direction with a spot size w_x^2 given by

$$w_x^2 = \frac{2L_x}{\beta} \tag{12.3.11a}$$

with the wave front being planar (i.e., $R_x = \infty$) along x.

2. The first factor of the second line of (12.3.10) describes a Gaussian beam with a spot size given by

$$\frac{y^2}{w_y^2} \equiv \frac{\beta y^2}{2L_y}\left(\frac{g_0}{2\beta}\right)^{1/2} \quad \text{or} \quad w_y^2 = \frac{2\sqrt{2}L_y}{(g_0\beta)^{1/2}} \tag{12.3.11b}$$

The quadratic variation of the phase with y, the second factor of line 2 of (12.3.10), indicates that the wave front is curved with the radius of curvature given by

$$\exp\left(-j\frac{\beta y^2}{2R}\right) = \exp\left[-j\frac{\beta y^2}{2L_y}\left(\frac{g_0}{2\beta}\right)^{1/2}\right]$$

or

$$R = L_y\left(\frac{2\beta}{g_0}\right)^{1/2} \tag{12.3.11c}$$

3. The third line merely indicates that the wave propagates with a modal phase constant given by

$$\beta_{\text{modal}} = \beta - \frac{1}{2L_x} - \left(\frac{g_0}{2\beta}\right)^{1/2}\frac{1}{2L_y} \tag{12.3.11d}$$

with β given by (12.3.4a). The factors "correcting" β are the result of the penetration of the field into regions of lower dielectric constant (for x variation) and into regions of lower gain—even loss—for the y variation.

4. The net *power* gain coefficient for this *mode* is also reduced from the peak value of g_0 because of the penetration of the field into the loss region.

$$g_{\text{modal}} = g_0 - \left(\frac{g_0}{2\beta}\right)^{1/2} \frac{1}{L_y} \tag{12.3.11e}$$

Example

Let us choose some typical numerical values so as to appreciate the physical significance of the previous development.

Let $\lambda_0 = 8400$ Å, $n = 3.6$; $g_0 = 100\ \text{cm}^{-1}$; $L_y = 5 \times 10^{-4}$ cm; and $L_x = 10^{-3}$ cm

$$\therefore \qquad \beta = \frac{2\pi n}{\lambda_0} = 2.69 \times 10^5\ \text{rad/cm}$$

Hence $w_x = 0.862\ \mu$m by (12.3.11a). The real index of refraction decreases from 3.6 to $3.6(1 - 0.0074) = 3.573$ in this "spot size" distance. This calculation points out that it does not take much of a change in the index to guide the wave. The field is described by a Gaussian with a "spot size" in the y direction, $w_y = 5.22\ \mu$m, from (12.3.11b).

The gain coefficient defined by (12.3.1) changes quite significantly over this spot size, decreasing from a peak value of 100 cm^{-1} at $y = 0$ to become *negative* (i.e., absorptive) at $y = L_y$. Because of the field penetrating into the absorptive regions, the effective (or modal) gain coefficient is given by (12.3.11e) to be $g_{\text{modal}} = 72.75\ \text{cm}^{-1}$. The situation is sketched in Fig. 12.6(a) where the field is plotted directly below the assumed spatial variation of the gain coefficient. The positive value of g_{modal} reflects the fact that the peak field lies near $y = 0$ where the gain is largest. The tails of the field penetrate into the lossy region and thus account for the reduction of the modal gain coefficient.

The radius of curvature of the phase front is given by (12.3.11c) to be $R = 366.9\ \mu$m. The fact that the radius is not infinity indicates that the wavefront is *curved* in the y direction (but not in the x direction). In other words, the beam is *astigmatic*. This is sketched in Fig. 12.6. To evaluate the far-field radiation pattern from such lasers, we must account for both the amplitude and phase variation over the output facet.

The astigmatism can be observed to us in the outside world by the radiation pattern. The Gaussian beam in the x direction has the $z = 0$ plane located at the output facet, and therefore the standard laws from Chapter 3 for the expansion into free space are applicable. However, the wave front is *curved* in the y direction as it arrives at the output facet, and thus some additional work is involved in computing the free space radiation. We need to compute the y variation of the complex beam parameter on the *air* side of the output facet mirror by using the *ABCD* matrix for a dielectric-air interface for that purpose. The spot size w_y will be the same in the air as in the laser; however, the radius of curvature will be different. This makes the beam appear to originate from a plane behind the output mirror, which in turn increases the beam spread in the y direction. The details are left for a problem (naturally).

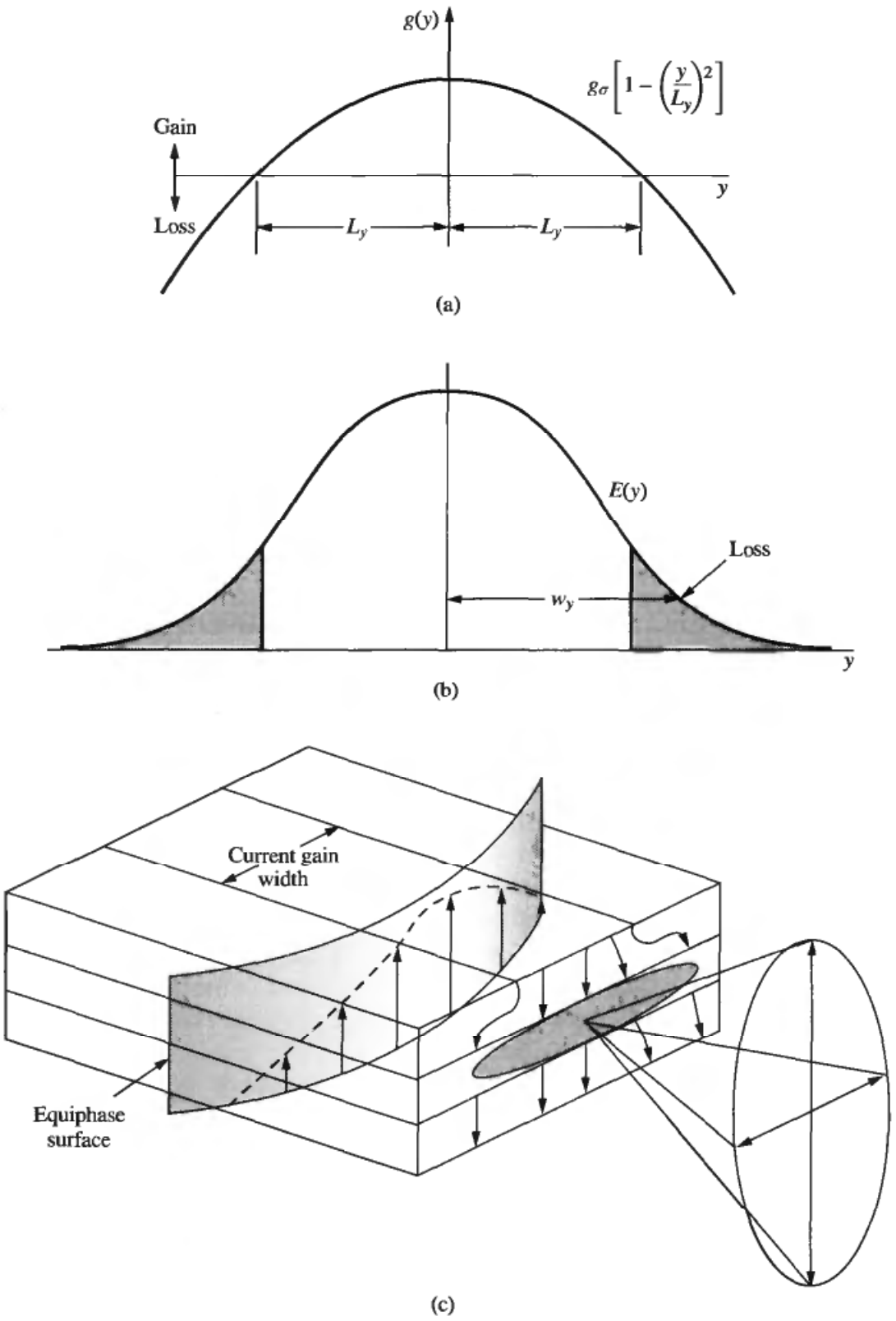


FIGURE 12.6. (a) The variation of the gain along the plane of the junction. (b) The resulting electric field of the mode. (c) The equiphase surface for a gain-guided laser.

12.4 OPTICAL CONFINEMENT AND EFFECTIVE INDEX

The example of the last section illustrated a rather benign case of a mode (i.e., a field configuration) in which some fraction of the field was in the gain medium, other parts in the absorbing region, and still others sampled a different dielectric constant. With all those "differences," the mode maintained its *relative* shape while propagating and being amplified along z.

As was demonstrated by the example in the previous section, the modal gain is always less than the peak gain and is less than the material gain. This is an important issue for all lasers but is particularly significant for the semiconductor case where the active region may be very small compared with the spatial extent of the field. For instance, suppose the recombination region of Fig. 12.1 were the 100 Å (in the x direction) quantum well considered in Sec. 11.6. That 100 Å distance is almost insignificant compared to the spatial extent of the field in the x direction [cf. (12.3.11b)].* Yet it is that very small region that provides the modal gain.

Since stimulated emission depends upon the intensity, the modal gain is related to the gain provided by the material by the optical confinement factor that expresses what fraction of the optical power of the mode overlaps the inverted material.

$$\Gamma = \frac{\iint_{\text{gain}} E^2(x, y)\, dx\, dy}{\iint_{-\infty}^{+\infty} E^2(x, y)\, dx\, dy} \tag{12.4.1}$$

and thus the gain coefficient is related to the material value by

$$g = \Gamma\gamma \tag{12.4.2}$$

where γ is the material gain coefficient from Chapter 11. For the larger gas, solid state, and dye lasers, the overlap is virtually complete, and $\Gamma \sim 1$. For the quantum well and heterostructure lasers, the confinement factor is much less than 1, and small changes in n_3, n_5 of Fig. 12.1 can significantly change the overlap.

The effective index (or effective dielectric constant) is defined (to give the correct answer!) by

$$\beta = k_0 n_{\text{eff}} \qquad \text{or} \qquad \beta^2 = k_0^2 \epsilon_{\text{eff}} \tag{12.4.3}$$

This definition is not terribly satisfying since we need the final answer before identifying the product. We could use the perturbation formula (4.5.4) to derive a variationally correct formula for it, but that is of limited use in itself. However, the concept leads to an "effective index method" of approximating intractable problems where the confinement in one direction—say, perpendicular to the junction—is much greater than along the plane of the junction. For instance, suppose the region outside the contact in Fig. 12.5 were etched down from the top to a distance comparable to w_x from the p-n junction and then back filled with a low dielectric constant material such as SiO_2. In such a case, the variation of $\epsilon'(x)$ is different depending upon the coordinate y. The effective index method would address this geometry by analyzing the slab geometry at least three times:

*Indeed, it was ignored in the computation of the field in Sec. 12.3.

1. We would solve the slab waveguide problem as in Sec. 12.2 for β for the region directly under the contact by using the assumption that $\partial(\ )/\partial y = 0$. It could be a 3 or 5 slab case (symmetric or otherwise). We would then find an effective index for this region under the contact. Let us call that value n_c.
2. We would then go to the region beyond the contact where there is a different slab waveguide problem to be addressed and solved with the same assumption of $\partial(\ )/\partial y = 0$. From the solution, we could then define another different effective index called n_a. A moment's thought will convince us that $n_a < n_c$.
3. Now the above two steps have defined a new symmetric "three slab" waveguide problem with plane of the slabs running in the x direction with the central index being equal to n_c and that of the outside slabs equal to n_a, which is precisely the problem addressed in Sec. 4.4. This yields the real index confinement of the mode in the y direction.

Clearly, this is not a calculation that can be put on one or two sheets of paper, but it is a very simple concept.

12.5 DISTRIBUTED FEEDBACK AND BRAGG REFLECTORS

12.5.1 Introduction

In most of our work, we have made the tacit assumption of simple mirrors for feedback elements. While this is prevalent, there are good reasons for distributing the feedback, although we will have to wait to appreciate them. As a start, let us convince ourselves that it is an interesting problem.

Consider the system shown in Fig. 12.7, in which a strong field is propagating to the right in a medium composed of periodic small reflections. To the zeroth-order approximation, this strong signal propagates more or less as if the medium were uniform and ignores the small reflections. At each of the planes $a, b, c, d, \ldots$, the wave propagating in the positive z direction generates a "small" wave propagating back to the left with the total "reflected" field being composed of all the individual reflections. The problem with this zeroth-order picture is that the total field propagating in the negative direction does not remain small if the phase delay between the individual small wave packets were chosen correctly. For instance, the wave generated at the plane b lags the corresponding field generated at plane a by $k\Lambda$ radians. When the b wave propagates back to position a, it experiences another $k\Lambda$ phase delay, yielding a total phase delay of b with respect to a (at the plane a) of $k(2\Lambda)$. Thus the total field propagating to the left in Fig. 12.7 is

$$E^-(z=a) = \underbrace{\Delta\Gamma E_0}_{a} + \underbrace{\Delta\Gamma E_0 e^{-jk2\Lambda}}_{b} + \underbrace{\Delta\Gamma E_0 e^{-jk4\Lambda}}_{c} + \underbrace{\Delta\Gamma E_0 e^{-jk6\Lambda}}_{d} + \cdots$$

If $2k\Lambda$ were an integral number of 2π radians, then the individual small wavelets add and, with N planes of reflections, E^- is $N\Delta\Gamma E_0$. For N large, this reverse wave could be

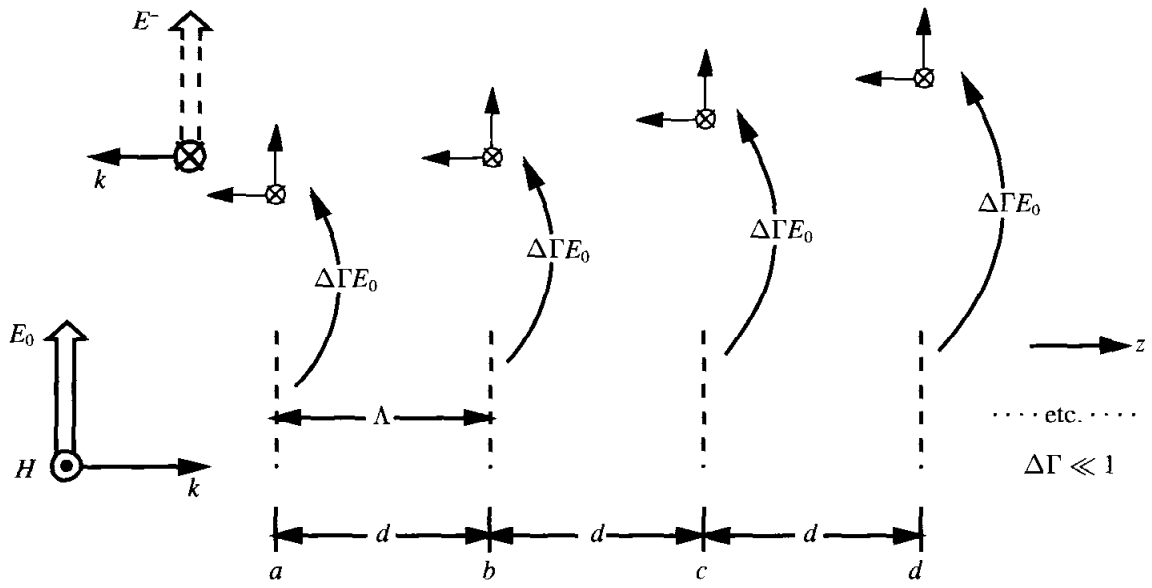


FIGURE 12.7. "Small" reflections from a periodic discontinuity.

comparable to the incident one. Obviously, we are "manufacturing" energy in a completely passive system, a neat (but impossible) trick, a consequence of our simplifying zeroth-order approximation. Thus we must resort to a more precise analysis.

However, this simple picture does provide us with a motivation for proceeding. Note that this feedback is frequency selective, since only those frequencies satisfying the condition

$$k(2\Lambda) = \frac{\omega}{c} n(2\Lambda) = m2\pi$$

or (12.5.1)

$$\Lambda = m\frac{\lambda}{2} \text{ with } \lambda = \frac{\lambda_0}{n} \qquad (n = \text{index of refraction})$$

generate components which add in-phase, and thus result in a wave of significant amplitude. The condition expressed by (12.5.1) is the same as found for Bragg scattering of electrons from the periodic arrangement of atoms in a crystal. The integer m corresponds to the various orders; in what follows, we will pick the simplest case of $m = 1$.

Such a feedback system can be realized in a semiconductor by etching a periodic structure along the axis of the diode and then regrowing a different material on top of the array. Fig. 12.8 shows an idealized version of that published by Nakamura et al. [21]. It is a fact of life that the index of refraction goes down as the aluminum fraction goes up, and thus the guiding properties and the effective width of the guide are both periodic functions of z. Now the phase constant of the guided modes is found from the matching of the fields at these transverse boundaries (cf Sec. 12.3), and thus the effective dielectric constant is also a periodic function of z as shown in Fig. 12.8(c). We should also expect the gain coefficient to have a similar variation as shown in Fig. 12.8(d).

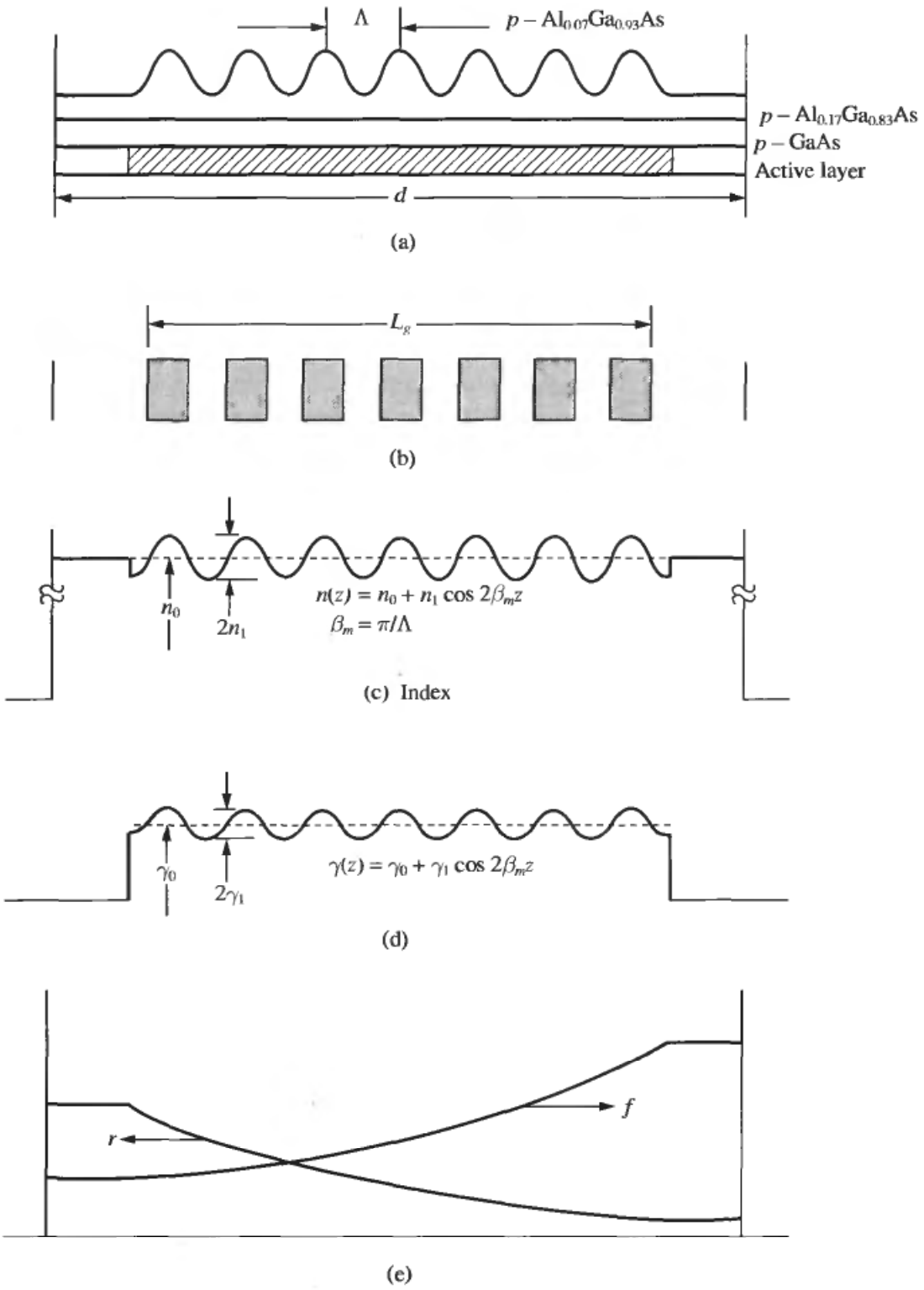


FIGURE 12.8. (a) Geometry for a distributed feedback semiconductor laser (after Nakamura [21]); (b), (c), and (d) are sketches of the material density, index, and inversion with z; and (e) is a first-order guess as to the variation of the *f*orward and *r*everse waves inside the cavity.

12.5.2 Coupled Mode Analysis

Let us focus our attention on the wave equation describing the variation of the field with respect to z and hide our ignorance of the details of the field matching in the transverse plane by using the periodic representation of n and γ shown in Fig. 12.8. We also travel the same logic path relating the gain coefficient and the index to the complex dielectric constant as was used in the previous section (12.3.4) to obtain

$$\frac{d^2E}{dz^2} + \left\{ \frac{\omega^2}{c^2}[n_0 + n_1 \cos(2\beta_m z)]^2 + j\frac{\omega}{c} n_0[\gamma_0 + \gamma_1 \cos(2\beta_m z + \theta)] \right\} E = 0 \tag{12.5.2}$$

where n_1 represents the spatial Fourier component of the "ripples" in the index around the effective value n_0 and γ_1 represent the corresponding ripples in the gain. The angle θ is included to allow for the possibility that the spatial variation in gain is shifted in space with respect to that of the index. However, to include that possibility in this first analysis complicates matters far in excess of its practical importance, so we will set $\theta = 0$ with a warning that the generality of our analysis is somewhat compromised. With this warning in mind, let us abbreviate terms and make some reasonable approximations.

$$\begin{aligned}
\beta_0 &= \frac{\omega}{c} n_0\text{: the phase constant of unperturbed mode} \\
g_0 &= \gamma_0/2\text{: the } \textit{field} \text{ gain coefficient} \\
g_1 &= \gamma_1/2\text{: the "ripple" in the } \textit{field} \text{ gain coefficient} \\
\kappa &= (\omega/c) \cdot (n_1/2) + j(g_1/2)\text{: the } \textit{coupling} \text{ coefficient} \\
2\beta_m &= 2\pi/\Lambda\text{: } \Lambda \text{ is mechanical periodicity, } \lambda_B/2n_0
\end{aligned} \tag{12.5.3}$$

We have also assumed a first-order grading [$m = 1$ in (12.5.1)] but that is easily corrected by redefining $\Lambda = m\Lambda_0$ at the end. The coupling coefficient κ will play a central role in the theory to follow. If g_1 is ignorable, then κ is real, and the system is described as an index coupled DFB; if g_1 dominates, then the system is named a gain coupled DFB. In general, κ is complex, which makes our mathematical life rather miserable, but the laser has no problem with complex arithmetic. For this first analysis, we will assume $g_1 = 0$ and an index coupled DFB and thus κ is real.

If we make the very reasonable assumptions that $n_1 \ll n_0$ and $\gamma_0 \ll \beta_0$ (per usual), (12.5.2) becomes

$$\frac{d^2E}{dz^2} + [\beta_0^2 + j2g_0\beta_0 + 4\beta_0\kappa \cos 2\beta_m z]E = 0 \tag{12.5.4}$$

which of course degenerates to the usual wave equation when the coupling coefficient κ equals zero. When $\kappa \neq 0$, we run into a problem not encountered in simple wave theory.

Usually, we can assume a solution of the form of a superposition of waves traveling in the $\pm z$ direction and then shorten our work by analyzing the propagation of each one separately. With $\kappa \neq 0$, such a ploy will not work, and we must acknowledge the presence of

both waves existing simultaneously in space. Based upon our earlier zeroth-order analysis, we anticipate interesting solutions in the near vicinity of the Bragg wavelength, so we assume

$$\mathbf{E}(\mathbf{z}) = [f(z)e^{-j\beta_m z} + r(z)e^{+j\beta_m z}]\mathbf{a}_t \tag{12.5.5a}$$

with the magnetic field being found from $-j\omega\mu_0\mathbf{H} = \nabla \times \mathbf{E}$ (as usual) to be

$$\frac{\omega\mu_0}{\beta_m}\mathbf{H}_t = [f(z)e^{-j\beta_m z} - r(z)e^{+j\beta_m z}]\mathbf{a}_z \times \mathbf{a}_t \tag{12.5.5b}$$

where terms on the order of $(f', r')/\beta_m$ have been neglected. We have not made an approximation in the formulation of (12.5.5a) and (12.5.5b). Rather we are attempting to fold in as much of our physical picture of Fig. 12.7 as possible without violating any of the mathematics. The functions $f(z)$ (the forward wave propagating in the $+z$ direction) and $r(z)$ (the reverse wave propagating in the $-z$ direction) remain to be determined by demanding a self-consistent solution. We substitute this form into (12.5.4) and collect terms with common exponential multipliers.

$$\begin{bmatrix} \dfrac{\partial^2 E}{\partial z^2} \\ + \\ [\quad]E \end{bmatrix} = \begin{bmatrix} -\beta_m^2 f - j2\beta_m f' + f'' \\ + \\ (\beta_0^2 + j2g_0\beta_0)f \\ + \\ 2\beta_0\kappa r \end{bmatrix} e^{-j\beta_m z} + \begin{bmatrix} -\beta_m^2 r + j2\beta_m r' + r'' \\ + \\ (\beta_0^2 + j2g_0\beta_0)r \\ + \\ 2\beta_0\kappa f \end{bmatrix} e^{+j\beta_m z}$$
$$+ [2\beta_0\kappa]fe^{-j3\beta_m z} + [2\beta_0\kappa]re^{+j3\beta_m z} \tag{12.5.6}$$

So far we are exact. Now we start the approximation process:

1. We neglect the second derivative terms since they are additive to the other terms f, f' which are multiplied by large values β_m^2 and β_m, respectively.
2. We set each multiplier of $\exp[\pm j\beta_m z]$ equal to zero separately because the other one oscillates quickly through all complex phases of 1 in a length $z = 2\Lambda$. (A formal approach uses the following logic: multiply (12.5.6) by the complex conjugate of one of the factors and integrate over a length 2Λ. The integrals of the asynchronous terms are zero if (f, r) vary slowly with z. This is called the rotating wave approximation (RWA) and will be encountered many times in electromagnetic and quantum problems.)
3. We ignore the terms varying as $\exp[\pm j3\beta_m z]$ (or they disappear by virtue of the RWA).

These approximations lead to

$$\left\{[(\beta_0^2 - \beta_m^2) + 2jg_0\beta_0]f - 2j\beta_m f' + 2\beta_0\kappa r\right\} e^{-j\beta_m z} = 0 \tag{12.5.7a}$$

$$\left\{[(\beta_0^2 - \beta_m^2) + 2jg_0\beta_0]r + 2j\beta_m r' + 2\beta_0\kappa f\right\} e^{+j\beta_m z} = 0 \qquad (12.5.7b)$$

A further approximation is in order: We need not distinguish between β_0 and β_m unless the difference occurs. Hence, let $\beta_0 = \beta_m + \delta$ with δ being the phase shift per unit of length in *excess* of the mechanical value $\beta_m = \pi/\Lambda = 2\pi n_0/\lambda_B$ built into the structure.

$$\beta_0^2 - \beta_m^2 = (\beta_0 + \beta_m)(\beta_0 - \beta_m) \cong 2\beta_m\delta$$

Equation (12.5.7) reduces to a more manageable form:

$$f' - (g_0 - j\delta)f = -j\kappa r \qquad (12.5.8a)$$

$$r' + (g_0 - j\delta)r = +j\kappa f \qquad (12.5.8b)$$

Let us stop for a moment to ensure that these equations make sense in the limit of $\kappa \to 0$. If $\kappa = 0$, then a solution to (12.5.8a) is $f(z) = f(0)e^{(g_0 - j\delta)z}$. The reinsertion of the phase factor $\exp[-j\beta_m z]$ and the definition of $\delta = (\beta_0 - \beta_m)$ leads to the following expression for the field, $E(z) = [E_0 e^{-j\beta_0 z}]e^{g_0 z}$, which is what would be expected for an amplifying medium.

We can combine (12.5.8a) and (12.5.8b) to obtain

$$\begin{bmatrix} f'' \\ r'' \end{bmatrix} - p^2 \begin{bmatrix} f \\ r \end{bmatrix} = 0 \qquad (12.5.9a)$$

where

$$p^2 = \kappa^2 + (g_0 - j\delta)^2 \qquad (12.5.9b)$$

Let us consider the following problem associated with a distributed feedback structure, which is illustrated in Fig. 12.9. We presume that a wave $\mathbf{E}^+$ is incident on the boundary at $z = -L/2$ and excites the forward f wave propagating in the periodic medium. It, in turn, generates a reverse wave r, which becomes the reflected wave $\mathbf{E}^-$ at $z = -L/2$. A fraction of the forward wave may make it to the right boundary at $z = L/2$ and become the transmitted wave $\mathbf{E}_t$. The reverse wave is, of course, zero at $z = +L/2$ since, for the problem at hand, there is nothing in the uniform medium beyond that point to generate a reverse wave.

With this sort of first-order guess at a solution it is appropriate to solve (12.5.9a) in a format that encompasses as much of this intuition as possible and still be mathematically

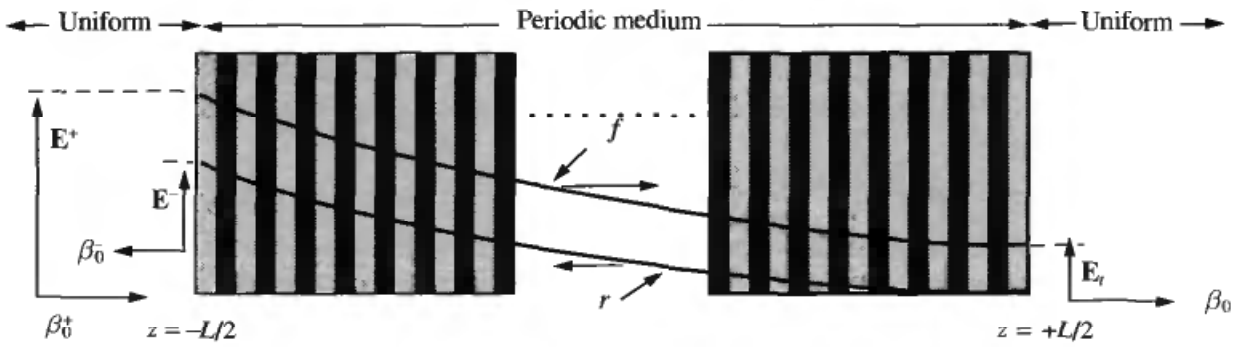


FIGURE 12.9. Geometry of a DFB medium (either a Bragg reflector or a distributed feedback amplifier) being excited by a source located at the left.

correct. A solution to (12.5.9b) for $r(z)$ can be written as

$$r(z) = A \cosh p(z - L/2) + B \sinh p(z - L/2) \tag{12.5.10a}$$

Now apply the obvious boundary condition that $r(z = +L/2) = 0$ to find that $A = 0$ and thus

$$r(z) = B \sinh p(z - L/2) \tag{12.5.10b}$$

Use (12.5.8b) to find $f(z)$

$$\begin{aligned} f(z) &= \frac{1}{j\kappa}[r' + (g_0 - j\delta)r] \\ &= \frac{B}{j\kappa}[p \cosh p(z - L/2) + (g_0 - j\delta) \sinh p(z - L/2)] \end{aligned} \tag{12.5.11}$$

Evaluate the coefficient B by relating $f(z = -L/2)$ to the incident field driving the structure and express all phases relative to that of the incident wave at $z = -L/2$

$$E_{\text{inc}}(z = -L/2) = E^+ = f(z = -L/2)e^{+j\beta_m L/2} \tag{12.5.12}$$

yielding

$$B = \frac{j\kappa E^+ e^{-j\beta_m L/2}}{p \cosh(pL) - (g_0 - j\delta) \sinh(pL)} \tag{12.5.13}$$

Reassemble the results and express them in an easily interpreted form. The reflected field (in the uniform medium to the left) is equal to

$$E^-(z = -L/2) = r(z = -L/2)e^{-j\beta_m L/2} = \frac{-j\kappa \sinh(pL)e^{-j\beta_m L}}{p \cosh(pL) - (g_0 - j\delta) \sinh(pL)} E^+ \tag{12.5.14}$$

Hence, the reflection coefficient at the plane $z = -L/2$ is

$$\boxed{\Gamma \triangleq S_{11} = S_{22} = \frac{E^-(z = -L/2)}{E^+(z = -L/2)} = \frac{-j\kappa L \cdot \sinh(pL) \cdot e^{-j\beta_m L}}{pL \cdot \cosh(pL) - (g_0 L - j\delta L) \cdot \sinh(pL)}} \tag{12.5.15}$$

where a symmetric structure has been assumed so that it "looks" the same from the right as it does from the left. Solve for the transmitted field

$$E_t(z = +L/2) = f(z = L/2) \exp[-j\beta_m L/2] = \frac{B}{j\kappa L} pL \cdot \cosh(0) \cdot \exp[-j\beta_m L/2]$$

Hence, the field transmission coefficient S_{12} is given by

$$\boxed{t \triangleq \frac{E_t(z = +L/2)}{E^+(z = -L/2)} = S_{12} = S_{21} = \frac{pL \cdot e^{-j\beta_m L}}{pL \cdot \cosh(pL) - (g_0 L - j\delta L) \cdot \sinh(pL)}} \tag{12.5.16}$$

Having the elements of the scattering matrix enables us to examine a variety of cases of practical importance with minimal effort. First, we will consider two passive cases for which total power is conserved and thus

$$|S_{11}|^2 + |S_{12}|^2 = 1 \qquad (\text{for } g_0 = g_1 = 0) \tag{12.5.17}$$

with the proof being saved for a problem.

12.5.3 Distributed Bragg Reflector

If $g_0 = 0$ (and thus g_1 is also), then κ is real, and we have the design equations for a dielectric mirror with (12.5.9b) assuming a simple form

$$(pL)^2 = (\kappa L)^2 - (\delta L)^2$$

There is a band of frequencies (or wavelengths) centered around the Bragg value ($\delta = 0$) where $(pL)^2$ is positive and thus (pL) is real. For reasonable values of (κL) and for $(\delta L) = 0$, the hyperbolic functions take on their limiting values of $\sinh(x) \sim \cosh(x) \sim e^x/2$, and the power transmission and reflection coefficients (at band center) are given by

$$R = |S_{11}|^2 = \left.\frac{(\kappa L)^2 \sinh^2 pL}{(pL)^2 \cosh^2 pL + (\delta L)^2 \sinh^2 pL}\right|_{(\delta L)=0} = \tanh^2 \kappa L$$

$$\approx [1 - 2e^{-2\kappa L}]^2 \tag{12.5.18a}$$

and

$$T = |S_{12}|^2 = \left.\frac{(pL)^2}{(pL)^2 \cosh^2 pL + (\delta L)^2 \sinh^2 pL}\right|_{(\delta L)=0} = \frac{1}{\cosh^2 \kappa L}$$

$$\approx 4e^{-2\kappa L} \tag{12.5.18b}$$

If κL is a reasonable value – say 3 – then the power reflection coefficient is high (99.01%) with the remainder being transmitted.

If $\delta \neq 0$, then the value of pL decreases toward zero, and in the limit of $\delta L \to \kappa L$, (12.5.18a) becomes

$$R(\delta L \to \kappa L) = \frac{(\kappa L)^2}{1 + (\kappa L)^2} \tag{12.5.18c}$$

and

$$T(\delta L \to \kappa L) = \frac{1}{1 + (\kappa L)^2} \tag{12.5.18d}$$

If κL is 3, then the power reflectivity is still respectable (90%), but R and T are starting to change rapidly with δL.

The region

$$-\kappa L < [\delta L = (\beta_0 - \beta_m)L] < \kappa L \tag{12.5.18e}$$

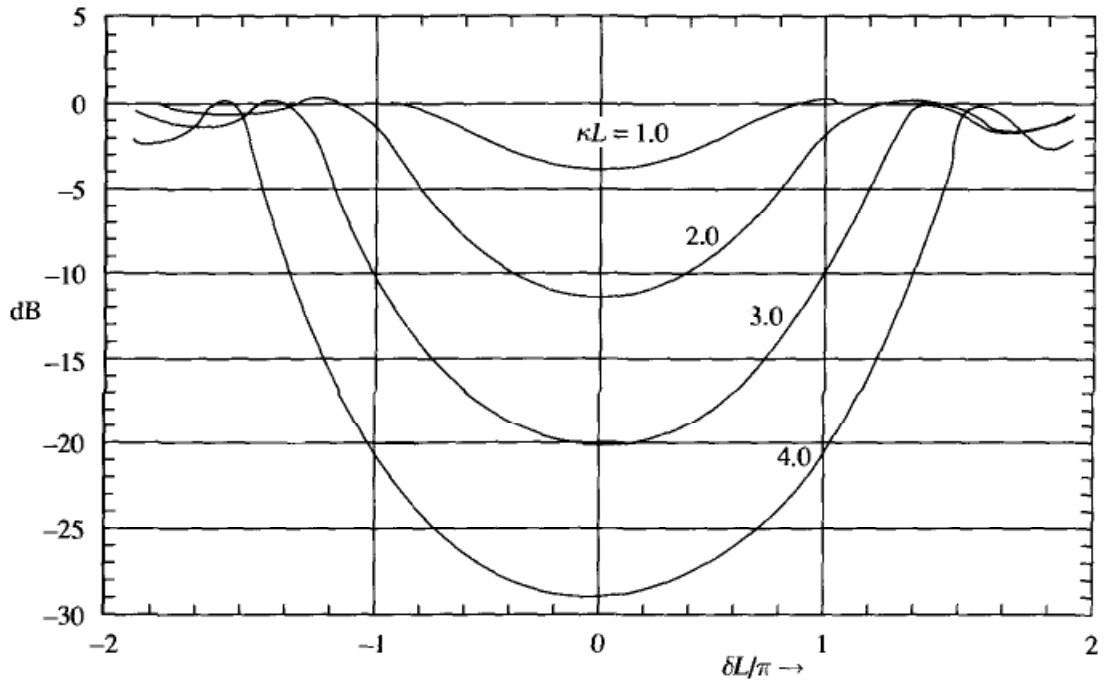


FIGURE 12.10. The relative power transmission (in dB) of a Bragg reflector as a function of the detuning parameter $\delta L/\pi$ for various values of the coupling coefficient ($\kappa L = 1.0$, 2.0, 3.0, and 4.0). The Bragg wavelength corresponds to $\delta L = 0$.

is called the "stop band," a term borrowed from lumped-element filter theory. The power transmission (in dB) is plotted in Fig. 12.10 for various values of κL. It should also be apparent that, in the stop band, the amplitudes of the fields decrease exponentially in the manner sketched in Fig. 12.9.

12.5.4 A Quarter-Wave Bandpass Filter

Once the elements of the scattering matrix are known, (12.5.15) and (12.5.16), it is a simple matter to cascade two (or more) Bragg reflectors in the manner shown in Fig. 12.11.

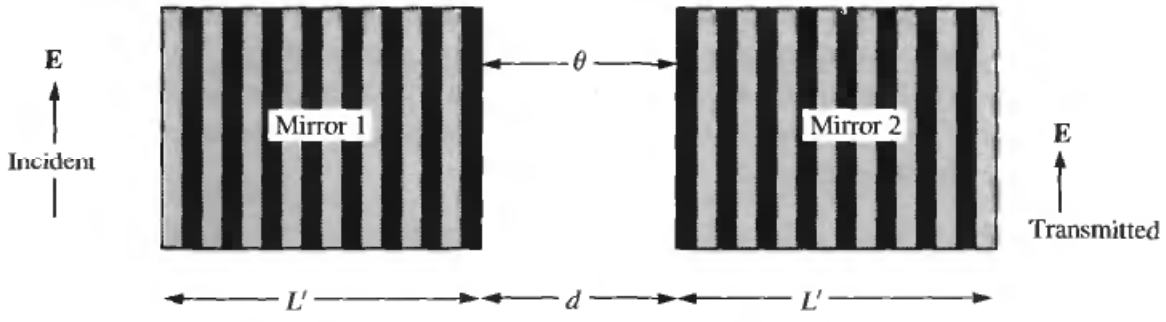


FIGURE 12.11. Two Bragg mirrors separated by a distance d leading to a phase shift of $\theta = \beta_0 d$ between the two mirrors. To compare this system with that shown in Fig. 12.9, the distance L' is $L/2$. The distance d is usually much smaller than L'.

We can use the scattering matrix formalism of Appendix I (see problem) to find that the composite field transmission coefficient is given by

$$(S_{12})_{\text{NET}} = \frac{[^1S_{12}][^2S_{12}]e^{-j\theta}}{1 - [^1S_{11}][^2S_{11}]e^{-j2\theta}} \tag{12.5.19}$$

where the prior superscripts refer to the respective mirrors. This should look very familiar, especially if we substitute Γ for S_{11} and t for S_{12}, for then (12.5.19) becomes identical to (6.3.4) for a simple Fabry–Perot cavity. There is a bit of difference here that was absent in Chapter 6, however—the reflection coefficient is complex whereas previously we had treated Γ as a real number. To illustrate the significance of that difference, let us evaluate S_{11} from (12.5.15) for the case of $\delta = 0$ and hence p is real $= \kappa$.

$$\Gamma(\delta = 0) = S_{11}(\delta = 0) = \frac{-j\kappa \sinh(pL')e^{-j\beta_m L'}}{\kappa \cosh(pL')}$$

Now $L' = N'\Lambda$, N' is number of periods of the mechanical ripples in each mirror, and $\beta_m = \pi/\Lambda$ [see (12.5.3)]. Thus, $\exp[-j\beta_m L'] = \exp[-jN\pi] = \pm 1$, and hence Γ is a pure imaginary number. If we assume the two mirrors of Fig. 12.11 to be identical, then the product $[^1S_{11}] \cdot [^2S_{11}]$ is a negative number and "resonance" [i.e., the denominator of (12.5.19) becoming small] will occur when $2\theta =$ odd multiple of π.

$$2\theta = (2q + 1)\pi \qquad \text{(for resonance)}$$

or

$$\theta = (2q + 1)\pi/2 \tag{12.5.20a}$$

For frequencies or wavelengths "close" to the Bragg value, $\theta = 2\pi n_0 d/\lambda_B$, and a resonance will occur when the spacing d is an odd multiple of $\lambda_B/4n_0$.

$$d = (2q + 1)\lambda_B/4n_0 \tag{12.5.20b}$$

It is left for a problem to show that the net power transmission coefficient for this system is 1 at $\delta = 0$ for any value of κ, if d is given by (12.5.20). Figure 12.12 shows the power transmission coefficient for various values of κL as a function of $\delta L/\pi$ and should be compared to Fig. 12.10. (L' of Fig. 12.11 was chosen to be $L/2$, and hence this represents the same structure as before but separated in the middle by the $\pi/2$ phase shifting section.)

Thus, a combination of two Bragg mirrors separated by a "quarter-wave" separation yields an optical filter that passes a narrow band of wavelengths and reflects others in the near vicinity. This promises to be a most useful component in optical communication systems employing wavelength division multiplexing.

Example

It is worthwhile to translate some of those curves into a format easier to interpret. Consider the transmission through the filter with $\kappa L = 3.0$ in Fig. 12.12. The "3 dB" reflection bandwidth corresponds to a change in $[(\delta L)^+ - (\delta L)^-]/\pi$ of $2 \times 1.7 = 3.4$, and the width of the narrow transmission resonance at the center corresponds to a FWHM of $\delta L/\pi \sim 0.2$ with the 10 dB bandpass corresponding to a change in $\delta L/\pi$ of 0.37×2. We can convert these changes to a

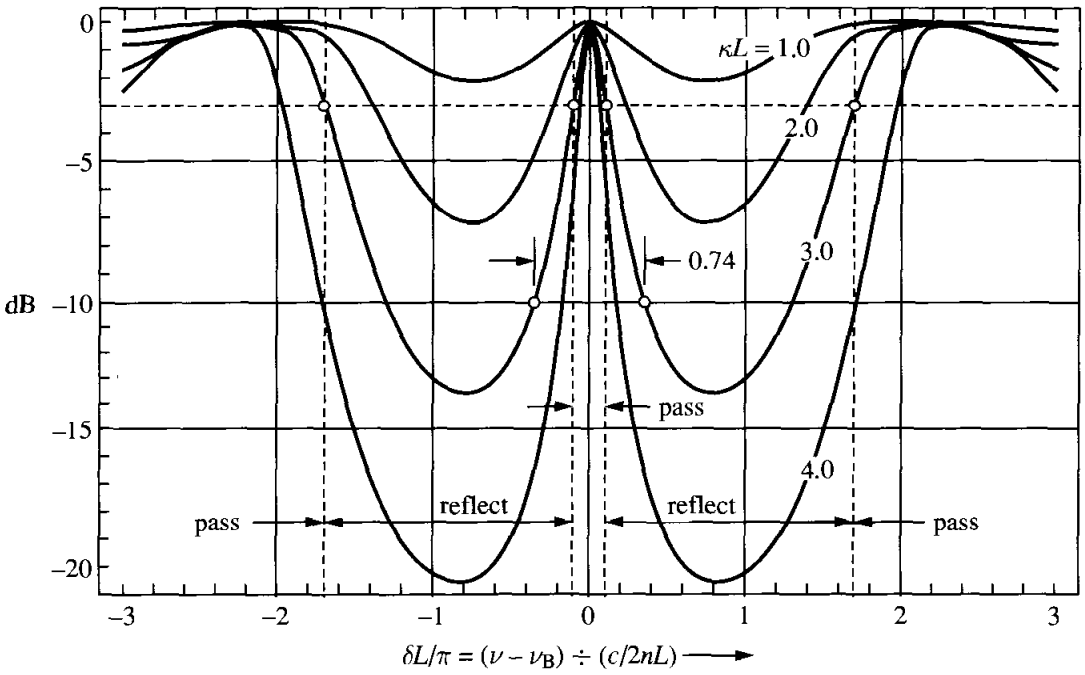


FIGURE 12.12. The relative power transmission (in dB) as a function of the detuning parameter $\delta L/\pi$ for various values of the coupling coefficient ($\kappa L = 1.0, 2.0, 3.0$, and 4.0) with a $\pi/2$ phase-shifting section in the center of the structure. The length L is the same as in Fig. 12.10, and thus $L' = L/2$.

wavelength interval. From the first specification, the filter reflects (> 50%) over the interval

$$\frac{(\beta^+ - \beta_m)L}{\pi} - \frac{(\beta^- - \beta_m)L}{\pi} = \frac{2\pi n_0 L}{\pi}\left[\frac{1}{\lambda^+} - \frac{1}{\lambda^-}\right] = 3.4$$

Now $2n_0L = 2n_0N\Lambda$, where $\Lambda = \lambda_B/2n_0$ and N is the number of layer *pairs* in the two mirrors (say, $N = 100$). If we assume a Bragg wavelength of $\lambda_B = 1.55$ μm and that $(\lambda_+) \cdot (\lambda_-) = \lambda_B^2$, then

$$\Delta\lambda_{\text{ref}} = (\lambda_+ - \lambda_-)_{\text{ref}} = 3.4\lambda_B/N = 527 \text{ Å} \qquad \text{(3dB reflection band)}$$

The narrow passband corresponds to a wavelength difference of 1/17 of the above value or

$$(\lambda_+ - \lambda_-)_{\text{pass}} = 0.2\,\lambda_B/N = 31 \text{ Å} \qquad \text{(3 dB passband)}$$

and the 10 dB bandpass is only 3.7 times as large. This wavelength selectivity is caused by a filter (say, in $Ga_xIn_{1-x}P_yAs_{1-y}$ with $n_0 = 3.37$) of length $L = 10^2 \cdot (\lambda_B/2 = 0.775\ \mu\text{m}/3.37) = 0.023$ mm. For $\kappa L = 3$

$$\kappa L = 3 = \frac{\omega}{c}\frac{n_1}{2}L = \frac{2\pi}{\lambda_B} \cdot \frac{n_1}{2} \cdot N\frac{\lambda_B}{2} = \frac{\pi}{2}N\left(\frac{n_1}{n_0}\right) \text{ or } \frac{n_1}{n_0} = 1.91 \times 10^{-2}$$

Thus the "ripple" need not be very large. Indeed, the reproducibility of the ripple and the uniformity of the spacing are much more critical than the amplitude.

12.5.5 Distributed Feedback Lasers (Active Mirrors)

We are now in the position to consider cases where the gain (g_0L) is positive and thus have the possibility of the transmitted (or reflected) field in Fig. 12.9 or 12.11 being larger than the incident one for various frequencies/wavelengths. If (g_0L) is chosen properly, then the ratios E_t/E_{inc} or $E_{\text{ref}}/E_{\text{inc}}$, will go to infinity at a specific value of the detuning δ, implying an oscillator in which there is an output wave without an incident driving one. Thus an outline of the mathematics required to compute the *threshold* is quite simple.*

We merely set the denominator of the net transmission coefficient $(S_{12})_{\text{net}}$ [i.e., (12.5.19)] equal to zero: If $\theta = 0$, then the threshold for the simple structure of Fig. 12.9 is computed, and if $\theta = \pi/2$, then the threshold for two active mirrors is found.

Unfortunately, the facts that $p^2 = \kappa^2 + (g_0 - j\delta)^2$ is complex and that the hyperbolic functions involve complex arguments make an analytic approach extremely tedious. At best, we can hope to obtain a "simpler" set of transcendental equations with the analytic effort, but, ultimately, we must resort to a numerical solution for cases of practical importance. However, the trends for the two cases and the *approach* to oscillation can be appreciated by examining Figs. 12.13(a) and 12.13(b).

In Fig. 12.13(a), the (linear) power transmission coefficient $|S_{12}|^2$ is plotted as a function of δL for various values of g_0L for the single DFB structure shown in Fig. 12.9, whereas Fig. 12.13(b) assumes that same structure, but split in the middle and separated by a $\lambda_B/4n_0$ or a $(\pi/2)$ phase shifting section. (The transmission is symmetric about $\delta = 0$.) The values of $\delta L/\pi$ of primary interest are those that maximize $|S_{12}|^2$. As we can discern from the trend, one of the maximums becomes larger than the others and, with a sufficient value of g_0L, will go to infinity (i.e., the structure will break into oscillation).

The transmission gain of the simple DFB structure plotted in Fig. 12.13(a) illustrates the discouraging fact that oscillation will *not* take place at the Bragg wavelength ($\delta = 0$) but at either (or both) at $\delta L/\pi \approx \pm 1.2$ (remember: the response is symmetric about $\delta = 0$). Indeed, the value of $|S_{12}|^2 \ll 1$ at $\delta = 0$ (i.e., at the Bragg or design wavelength) even though the system has gain. If the facts that the wavelength is not as designed and we do not obtain a single wavelength (frequency) of oscillation are not dismaying enough, then note the local variation of those peaks as a function of $\delta L/\pi$: They are comparatively broad, which implies a less than desired frequency selectivity.

The transmission gain of the $\pi/2$ shifted structure plotted in Fig. 12.13(b) "cures" the problems listed above and adds some additional benefits. Note that there is a relatively sharp resonance at $\delta = 0$ whose maximum is larger than those at $\delta L/\pi = \pm 1.8$. Thus, if the gain is made large enough, the mode at $\delta = 0$ will break into oscillation. Furthermore, the enhancement in $|S_{12}|^2$ at $\delta = 0$ is much larger in Fig. 12.13(b) for a given gain than it is at $\delta L/\pi = \pm 1.2$ in Fig. 12.13(a). Thus, the threshold for oscillation is much less for the $\lambda_B/4n_0$ shifted structure.

Knowing that the frequency of oscillation will occur at $\delta = 0$ simplifies the analytic task enormously since p is real and thus we merely require that $|S_{11}| = 1$ for the denominator

*Since saturation effects have been ignored, we cannot compute the amplitude of oscillation; at most, we can only compute the threshold gain to reach that condition.

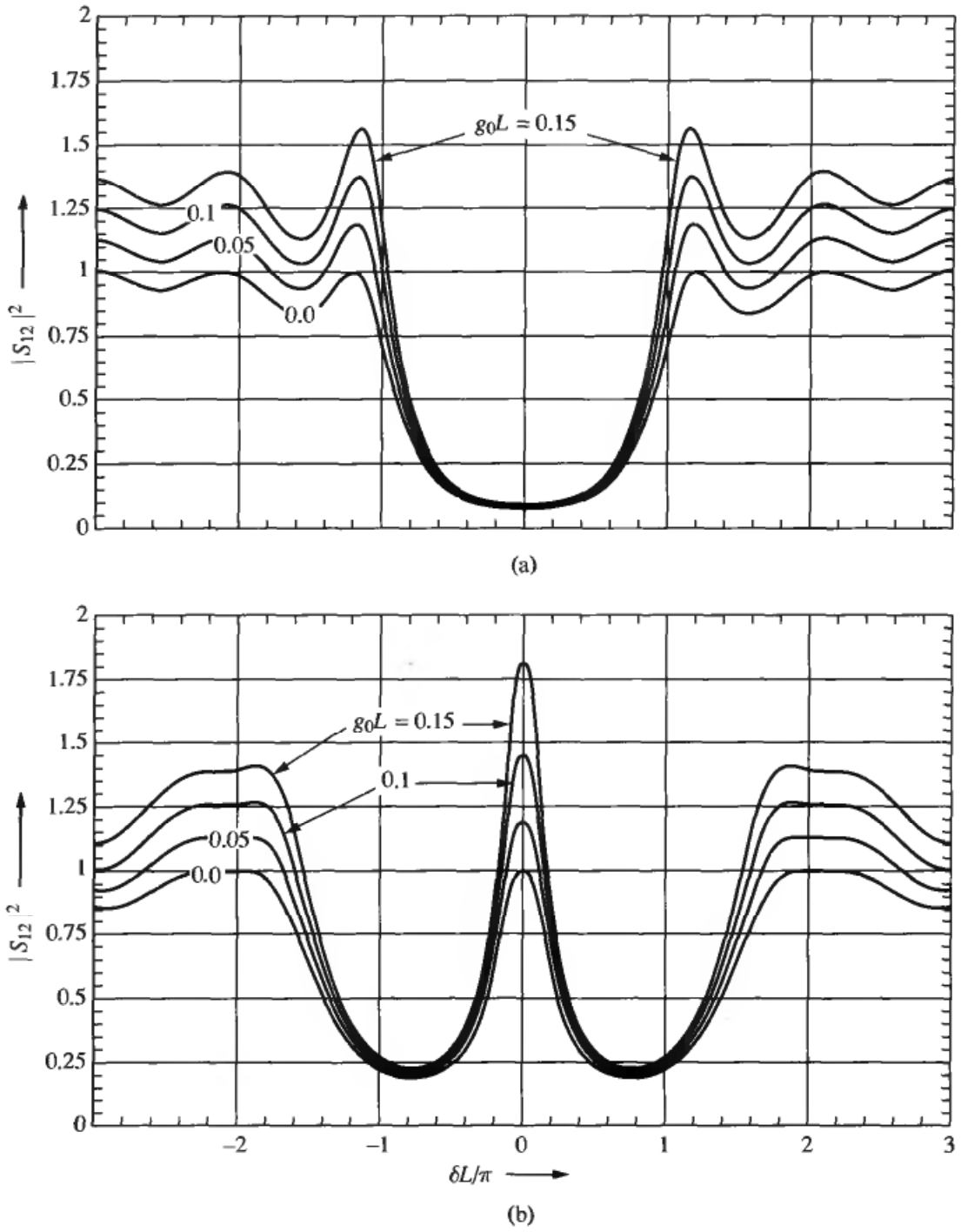


FIGURE 12.13. The power transmission coefficient for increasing values of the integrated gain $g_0L = 0.0, 0.05, 0.10$, and 0.15. The ordinate is $\delta L/\pi = (\nu - \nu_B) \div c/2nL$. (a) A simple DFB structure was assumed; (b) a $\pi/2$ (or $\lambda_B/4n_0$) phase shift was incorporated in the middle, but $\kappa L = 2.0$ in both cases.

in (12.5.19) to go to zero.

$$\frac{(\kappa L') \cdot \sinh(pL')}{(pL')\cosh(pL') - (g_0 L') \cdot \sinh(pL')} = \pm 1$$

with $p^2 = \kappa^2 + g_0^2$. After squaring, using the definition of p^2, and applying the identity $\cosh^2(\ ) - \sinh^2(\ ) = 1$, we obtain

$$2(g_0 L') \cdot (\kappa L') \cdot \left\{ \frac{\sinh \sqrt{(\kappa L')^2 + (g_0 L')^2}}{\sqrt{(\kappa L')^2 + (g_0 L')^2}} \right\}^2 = 1 \qquad (12.5.21)$$

This is solvable on almost all handheld calculators with the results being shown in Fig. 12.14. Notice that as $\kappa L'$ becomes larger the line integrated gain $g_0 L'$ (per mirror) for threshold of oscillation becomes smaller as we would expect. The threshold gain for the simple DFB will be higher than that shown in Fig. 12.14 because oscillation occurs at $\delta \neq 0$, and thus the feedback from the ripples is not quite as effective. This can be verified by comparing the maximums in Figs. 12.13(a) and 12.13(b).

There are many other considerations of DFB lasers that are not addressed in the above analysis. For instance,

1. We have assumed κL is real (index coupled), yet there are advantages to having a "gain coupled" system in which κL is imaginary. For such a system, oscillation will occur at the Bragg frequency without the necessity of the $\pi/2$ phase shifting section. However, it is very difficult, if not downright impossible to obtain a purely gain

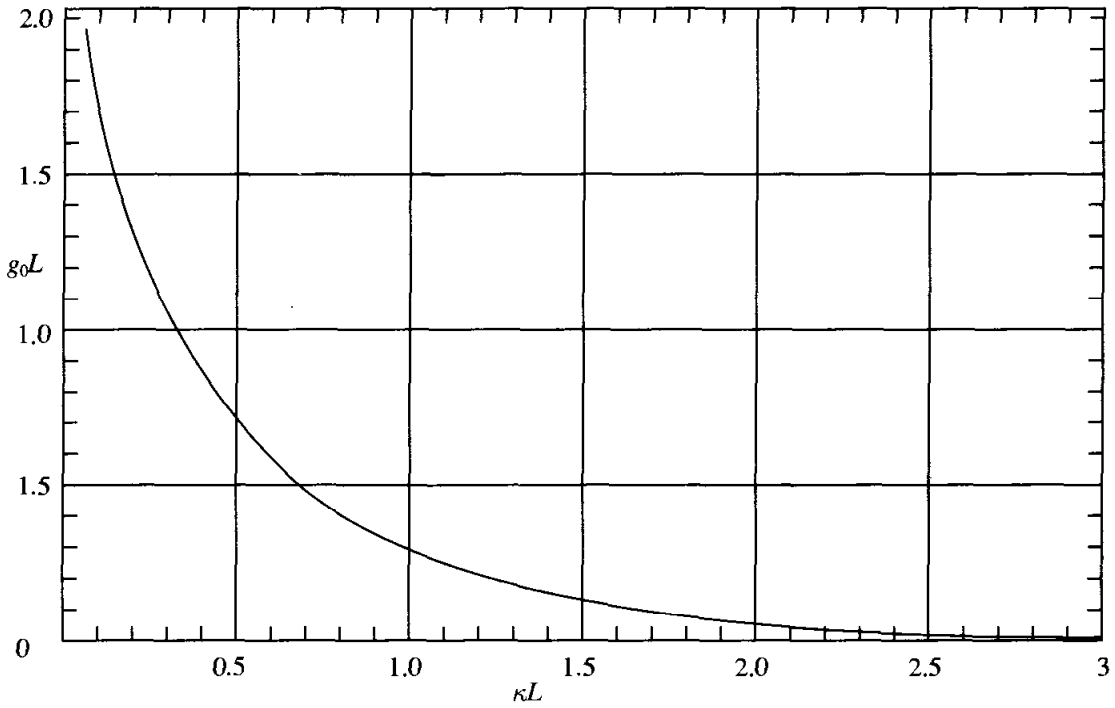


FIGURE 12.14. The line integrated threshold gain for a "quarter-wave" phase shifted DFB.

coupled system. The practical case always has a mixture of both, but predominantly index coupled.

2. We have assumed no reflections in the uniform regions to the right and left of the DFB structures. This may be difficult to accomplish in a practical sense, especially for the most common application, that of semiconductor lasers. The chip is cleaved somewhere to extract the optical beam and reflections from that interface should be taken into account. The theory presented above can be easily adapted to include reflections with a penalty of additional arithmetic. In any case, we would expect that the threshold would be lower if such reflections were included and would also expect some complicated interrelationship between the Fabry–Perot modes and the DFB modes.
3. The "quarter-wave" DFB has clear-cut advantages over the simple structure, but it does demand extra precision as compared to the simple bulk structure. In one sense, the $\lambda_B/4n_0$ spacer is just a "missing $\Lambda/2$ or half-of-a-cycle" in the middle of the periodic structure. This means that the active mirror 1 must be created independently of mirror 2 and separated precisely by $\lambda_B/4n_0$. This requirement places rather significant demands on the writing (or exposure) system creating the periodic structure.
4. Figure 12.14 might suggest that larger values of the coupling coefficients (κL) might be desirable and in the sense of minimizing pumping for threshold that conclusion is valid. However, above threshold, saturation effects must be included with the gain of the forward wave being affected by it and the local reverse wave. The standing waves can burn "spatial holes" in the gain profile and allow other modes to oscillate. In any case, the complexity of the problem increases rather dramatically.

12.5.6 Tunable Semiconductor Lasers

One of the ways to increase the channel capacity of any communication network is to use *w*avelength *d*ivision *m*ultiplexing (WDM) in which different wavelengths/frequencies are used for independent channels of communication. This is, of course, the system used at the lower radio, TV, and microwave frequencies where each "channel" is allotted a certain bandwidth to dispense the information. This same ploy is applicable at optical frequencies (where the available bandwidth is enormous) and thus there is the need for *integrated* tunable lasers and filters, just as there is a need for those components at the lower frequencies.*

There are two criterion that determine the oscillation frequency/wavelength of a laser: It oscillates at the wavelength at which the gain exceeds the loss by the largest amount (with both being wavelength dependent), and/or is determined by the requirement for the round-trip phase shift be an integrable multiple of 2π radians so as to maximize the stimulated emission rate. Continuous as well as step tunable semiconductor lasers have been obtained using various physical effects that either changes the phase and thus the condition for resonance, or change the gain-to-loss ratio or both.

*A few of the methods for tuning other broad-band lasers were covered in Sec. 10.6.4. However, those techniques are not amenable to integration. Here we restrict our attention to methods of tuning a semiconductor laser over its gain bandwidth with all of the components capable of being integrated.

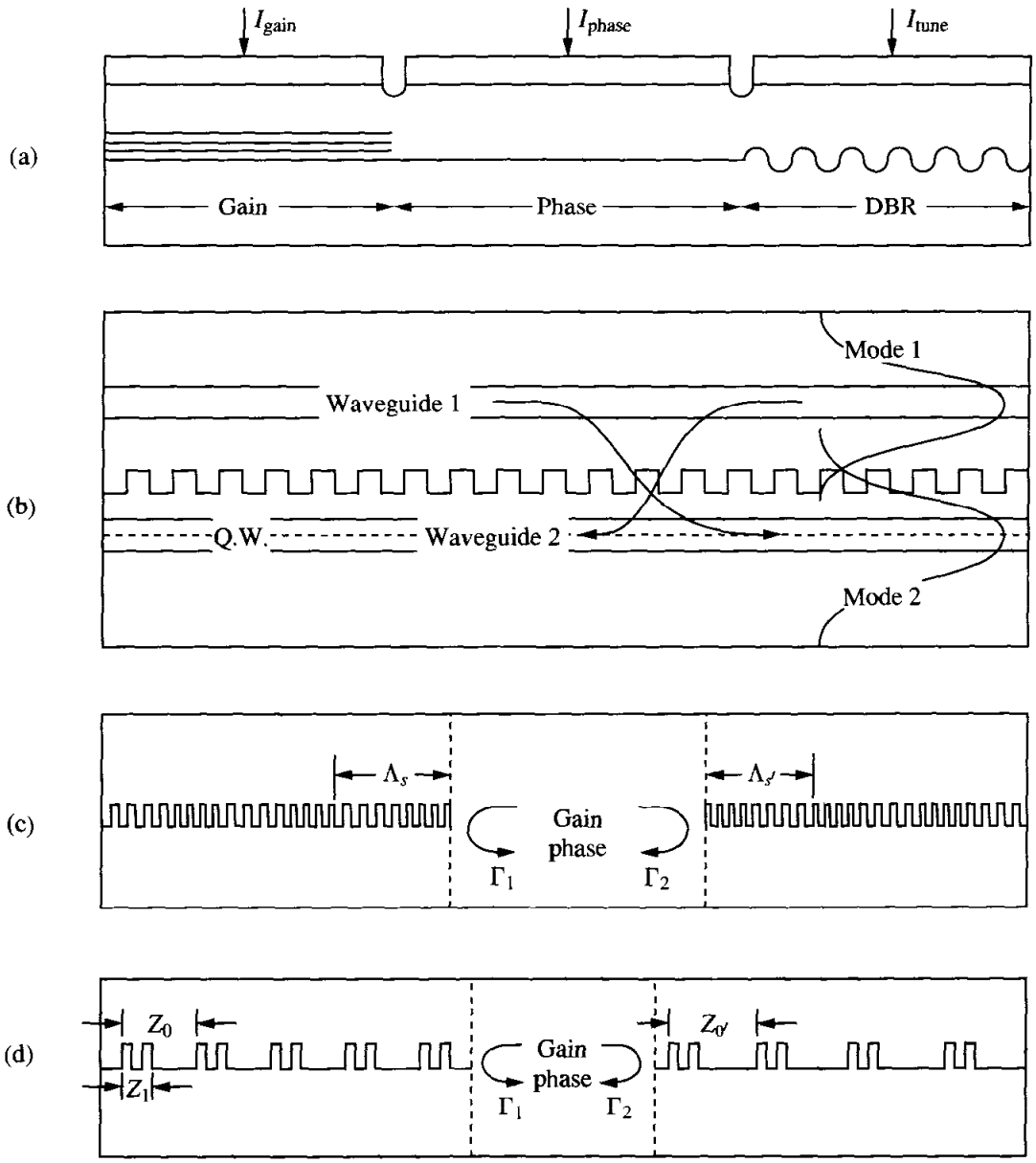


FIGURE 12.15. Schematics for tunable semiconductor lasers. (a) is taken from [37]; (b) is from [42]; (c) is from [47]; and (d) is from [51].

A small sampling of the successful methods is shown in Fig. 12.15, but there are probably other techniques being investigated aimed at obtaining a simpler structure, with fewer processing steps (which implies a greater yield in manufacturing), and with wider tuning.

There are two physical effects utilized in cases shown in Fig. 12.15. The first is caused by the injected free carriers and band filling effects, which contributes a *negative* change to the index of refraction

$$\Delta n = n - n_0 = -\frac{1}{n_0}\frac{\omega_p^2}{\omega^2} = -\frac{1}{n_0}\frac{n_e e^2}{m_e^* \epsilon_0}\frac{\lambda^2}{4\pi^2 c^2} \tag{12.5.22}$$

where n_0 is the local index of the bulk material and n_e is the concentration of the carriers and is obviously proportional to the local current density. Only the electron contribution is indicated (because of their smaller mass), but the holes also contribute in a negative sense using the same formula. (This is sometimes referred to as the "plasma" effect since the same formula applies to free electrons in an ionized gas.) Thus the phase velocity of the wave in the central or phase shifting section of Fig. 12.15a is controlled by the current I_{phase}. In a similar manner, the peak of the Bragg reflection can be adjusted by the current I_{tune}. Since the lifetime of the carriers being injected is on the order of a few nanoseconds, which is much larger than the photon lifetime, this is the time scale for such frequency changes. In addition to these relatively fast changes, there are also slower changes due to the local heating of the waveguides by the current which also changes the index and can be utilized for tunability.

The same two effects play a role in the schemes outlined in Fig. 12.15b to 12.15d, but the electromagnetic issues are more sophisticated and complicated. In Fig. 12.15b, two vertically separated guides are coupled by a grating. The two guides are designed such that the respective phase velocities differ by the wavenumber periodicity of the grating (which need not be given by the Bragg condition)

$$|\beta_2 - \beta_1| = \frac{2\pi}{\Lambda} \tag{12.5.23}$$

which maximizes the interchange of energy between the two guides. Since the phase constants can be expressed as $\beta = 2\pi n_{\text{eff}}/\lambda_0$ for each guide, one obtains a rather simple expression for the wavenumber of the oscillation

$$\frac{1}{\lambda_0} = \frac{1}{|n_{2\text{eff}} - n_{1\text{eff}}|\Lambda} \tag{12.5.24}$$

where n_{eff} is the effective index for each guide. Since both effective indices are close to each other, one obtains a differential effect in which a small change in one or both of the indices results in a large change in the wavenumber.

The "chirped" or superstructure grating shown in Fig. 12.15c achieves a similar differential effect by modulating the grating periodicity (in real space) every Λ_s units. This put sidebands on the high reflectivity band of the DBR's whose separation in wavelength is given by $\Delta\lambda = \lambda_0^2/(2n_{\text{eff}}\Lambda_s)$. A different $\Lambda_{s'}$ is chosen for the other mirror so that the combination appears as two combs with slightly different spacing of the teeth. The minimum loss condition will occur when one set of sidebands (or teeth) from each mirror are aligned. A small change in injected current into the Bragg mirrors results in a different set of sidebands aligning and thus a significant stair-step tuning can be achieved with a small change in index. With current control on both mirrors, continuous tuning can be achieved.

The sampled grating approach of Fig. 12.15d uses a similar approach but with a simpler scheme. Imagine starting with two Bragg mirrors with the same periodicity Λ and then removing part of the corrugations with a spatial period of Z_0 leaving the original grating present for a distance $Z_1 < Z_0$. This puts sidebands spaced at $2\pi/Z_0$ around the original Bragg peak of $2\pi/\Lambda$. Now by choosing different values of the sampling period Z_0 for the front and back mirror, these sidebands can be brought into coincidence by the injected current and thus the laser can be tuned in the same manner as in Fig. 12.15c.

12.6 UNSTABLE RESONATORS

12.6.1 General Considerations

Although the use of stable resonators was instrumental in the development of lasers, they do have a disadvantage of having a very *small* mode volume. Consequently, it is difficult to pack enough excited atoms into this volume to generate high power in the laser oscillator. Furthermore, it is difficult to restrict oscillation to the $TEM_{0,0}$ mode unless we insert apertures to increase the diffraction losses on the higher-order modes. Even this is a nontrivial task. We must adjust the *size* and *position* of the aperture with respect to an *unknown* axis of the cavity.

Thus we naturally look fondly at the edges of the unstable regime, where spot sizes at the various mirrors tend to become very big (see Fig. 5.5 for the hemispherical cavity). The spot size at the spherical mirror goes to infinity as $d \to R$. Unfortunately, we can expect the accuracy of the theory of Gaussian beams to degenerate rapidly as we approach the edges of the stability diagram and be totally inapplicable in the unstable regime.

A valuable alternative to this was provided by Siegman [1]. He reasoned that we could consider the field between mirrors of an unstable resonator to be that of a limited extent *spherical* wave whose size is determined by the aperture of the mirror.

For instance, consider the geometry shown in Fig. 12.16(a), obviously unstable by our previous analysis. If we assume a spherical wave to originate from the point P_1 as in Fig. 12.16(a) (undefined at this state) with a size determined by the aperture a_1, a considerable fraction of this wave misses M_2. However, this incident wave does illuminate the mirror M_2 more or less uniformly. Consequently, it will generate a limited-extent spherical wave whose extent is dictated by a_2 apparently originating from a point P_2, as indicated in Fig. 12.16(b). Obviously, P_2 is the image of the source at P_1. We have a self-consistent picture if the point P_1 is also the image of the source at P_2. The part of the wave that leaks around the mirrors can be considered useful output, and the fraction that makes a complete round trip determines the required gain of the laser medium.

Let us now try to translate these ideas into an analytic description. Consider the geometry shown in Fig. 12.17 and postulate the existence of a set of points P_1 and P_2 that act as virtual sources (or the object) for the waves that illuminate the other mirror. It will be convenient to measure all distances in units of the mirror spacing, d; that is, P_1 is located $r_1 d$ from M_1 and $(r_1 + 1)d$ from M_2. Thus P_1 is the virtual source of radiation impinging on M_2. The reflected wave appears to originate from the object point P_2. The use of standard mirror formulas relating the image and object distances yields

$$\left.\begin{aligned} \frac{1}{(r_1 + 1)d} - \frac{1}{r_2 d} &= \frac{2}{R_2} \\ \text{or} \qquad \frac{1}{r_2} - \frac{1}{r_1 + 1} &= 2(g_2 - 1) \end{aligned}\right\} \tag{12.6.1}$$

where

$$g_2 = 1 - \frac{d}{R_2} \tag{12.6.2}$$

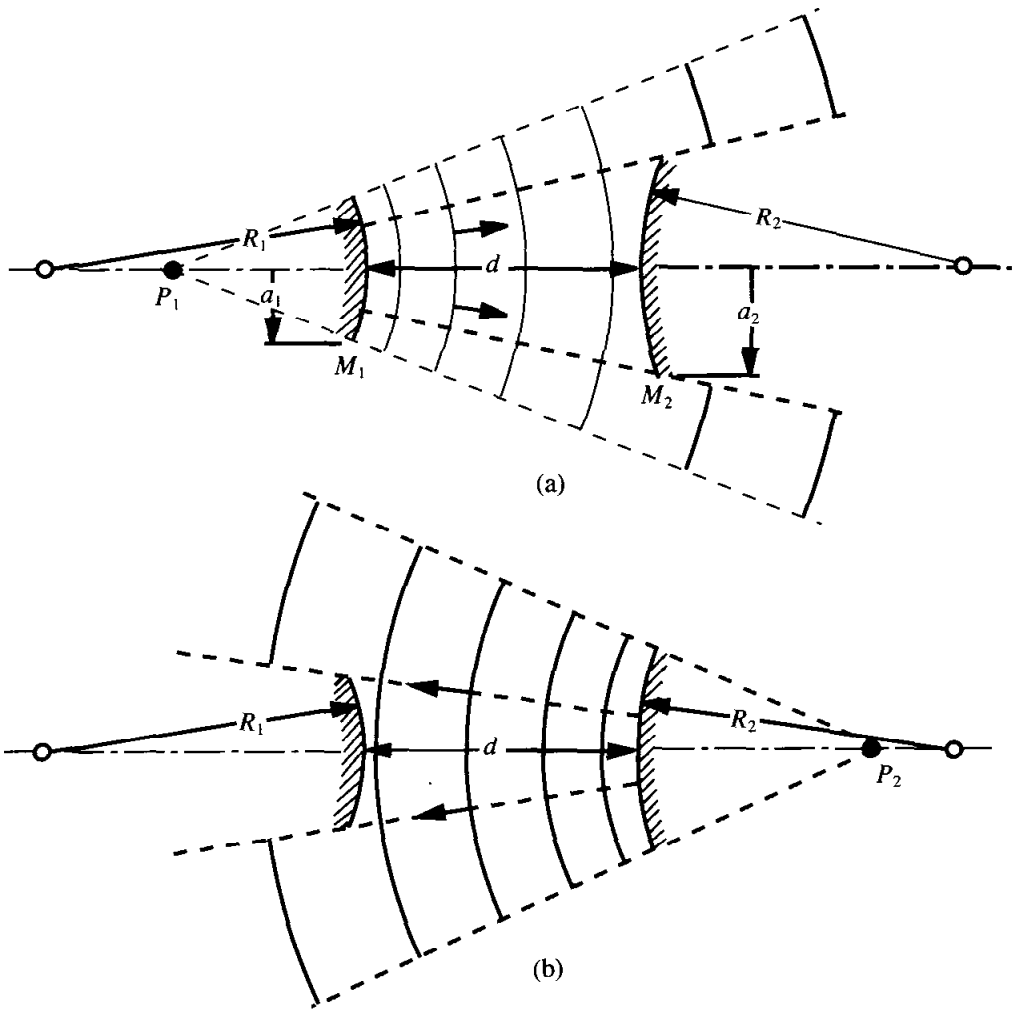


FIGURE 12.16. Wave fronts in an unstable cavity.

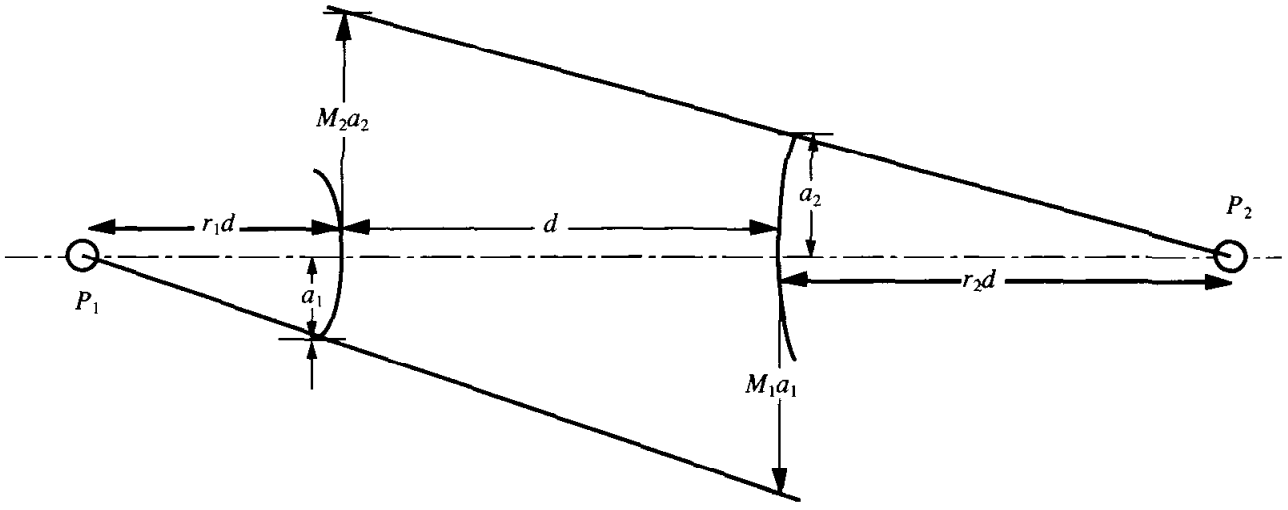


FIGURE 12.17. The unstable cavity illustrating the relative distances $r_{1,2}$ and the magnification of the beam size.

A self-consistent picture is obtained if the new source at P_2 has its image point at the original P_1.

$$\left.\begin{aligned} \frac{1}{(r_2+1)d} - \frac{1}{r_1 d} &= \frac{2}{R_1} \\ \text{or} \qquad \frac{1}{r_1} - \frac{1}{r_2+1} &= 2(g_1 - 1) \end{aligned}\right\} \tag{12.6.3}$$

where

$$g_1 = 1 - \frac{d}{R_1} \tag{12.6.4}$$

We must now solve these equations simultaneously to find these characteristic points $r_1 d$ and $r_2 d$. Note that if the r's are positive numbers, these points lie *outside* the cavity; a negative number would indicate that a source point lies on the reflecting side of the mirror. Perseverance with the arithmetic leads to

$$r_1 = \frac{[1-(g_1 g_2)^{-1}]^{1/2} - 1 + g_1^{-1}}{2 - g_1^{-1} - g_2^{-1}} \tag{12.6.5a}$$

$$r_2 = \frac{[1-(g_1 g_2)^{-1}]^{1/2} - 1 + g_2^{-1}}{2 - g_1^{-1} - g_2^{-1}} \tag{12.6.5b}$$

It will be convenient to use the concept of a "magnification" of the size of the beam as it propagates from one mirror to the other. Consider the sketch shown in Fig. 12.17 where the beam of (radial) size a_1 is shown leaving mirror M_1. Since that beam (apparently) originates at P_1, the beam size will be bigger or will be magnified by $(r_1 + 1)/r_1$ by the time it reaches mirror 2. By the same token, the beam of radial size a_2 leaving M_2 and (apparently) originating at P_2 will be magnified by $(r_2 + 1)/r_2$ as it goes back to 1. The magnification factors can be related to the cavity g parameters used to describe stability or the lack thereof.*

$$\mathbf{M}_1 = \frac{r_1 + 1}{r_1} = \frac{[1-(g_1 g_2)^{-1}]^{1/2} + (1 - g_2^{-1})}{[1-(g_1 g_2)^{-1}]^{1/2} - (1 - g_1^{-1})} \tag{12.6.5c}$$

$$\mathbf{M}_2 = \frac{r_2 + 1}{r_2} = \frac{[1-(g_1 g_2)^{-1}]^{1/2} + (1 - g_1^{-1})}{[1-(g_1 g_2)^{-1}]^{1/2} - (1 - g_2^{-1})} \tag{12.6.5d}$$

The concept of dimensional magnification is most useful when computing the round trip survival factor, which is the fraction of the power in the entire beam that survives a round trip.

The effective power survival coefficient of M_2 is the product of the reflectivity of that mirror, Γ_2^2 times the fraction of the spherical wave emitted by P_1 and intercepted by M_2.

*Do not confuse the magnification parameters $\mathbf{M}_{1,2}$ with the bold face type with the names of the mirrors with the same symbol.

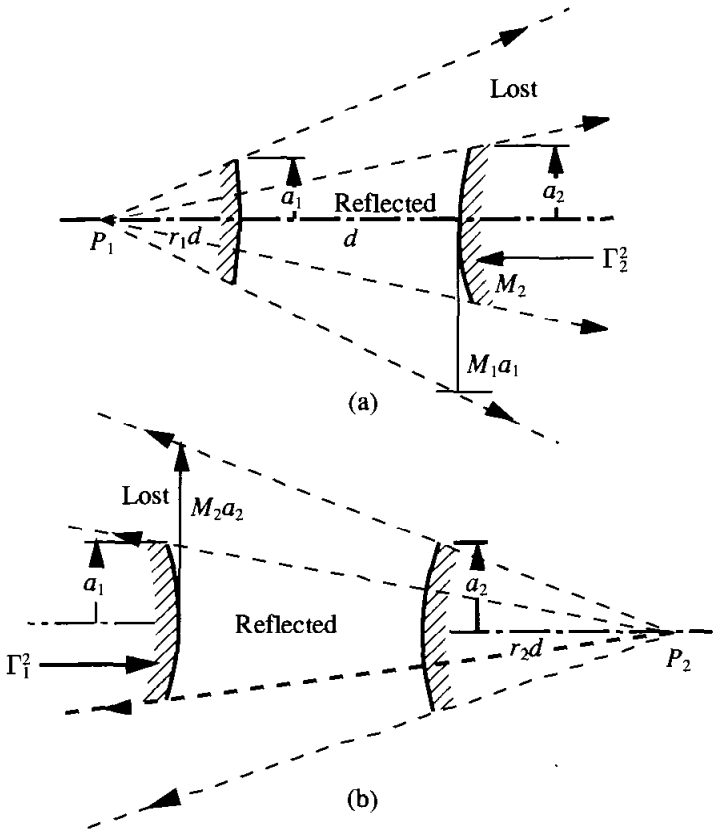


FIGURE 12.18. Power lost in an unstable cavity.

From Fig. 12.18(a), we see that

$$R_{2\text{eff}} = S_2 = \Gamma_2^2 \frac{\text{solid angle subtended by } M_2 \text{ with } P_1 \text{ as origin}}{\text{solid angle subtended by } M_1 \text{ with } P_1 \text{ as origin}}$$

$$= \Gamma_2^2 \frac{\pi a_2^2/[4\pi(r_1+1)^2 d^2]}{\pi a_1^2/[4\pi(r_1)^2 d^2]}$$

$$S_2 = \Gamma_2^2 \left[\frac{r_1}{r_1+1}\right]^2 \left[\frac{a_2}{a_1}\right]^2 = \Gamma_2^2 \left[\frac{a_2}{a_1}\right]^2 \frac{1}{\mathbf{M}_1^2} \tag{12.6.6}$$

and from Fig. 12.18(b) (or change all 2 subscripts into 1 and 1 into 2):

$$R_{1\text{eff}} = S_1 = \Gamma_1^2 \left[\frac{r_2}{r_2+1}\right]^2 \left[\frac{a_1}{a_2}\right]^2 = \Gamma_1^2 \left[\frac{a_1}{a_2}\right]^2 \frac{1}{\mathbf{M}_2^2} \tag{12.6.7}$$

Thus the round trip survival factor is given by

$$S^2 = S_1 S_2 = \Gamma_1^2 \Gamma_2^2 \left[\frac{r_1 r_2}{(r_1+1)(r_2+1)}\right]^2 = \Gamma_1^2 \Gamma_2^2 \frac{1}{\mathbf{M}_1^2 \mathbf{M}_2^2} \tag{12.6.8}$$

Note that this is the important parameter insofar as laser oscillators are concerned. The single-pass gain must be such that the net power gain exceeds the loss (i.e., $GS > 1$); see (0.1). Note also that this expression is independent of the mirror sizes within the context of this first-order theory. At first glance, this may seem strange, but recall the physical situation being studied.

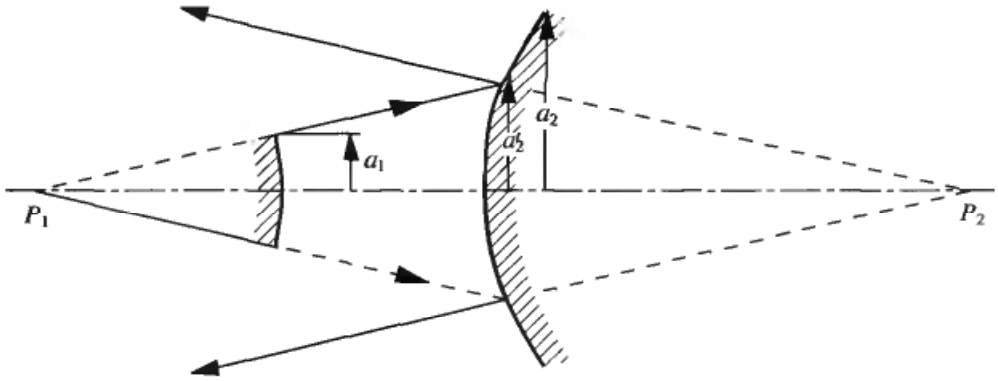


FIGURE 12.19. Extreme case of an unstable resonator.

If the size of the mirror M_2 in Fig. 12.18 were made larger, less of the power would leak out that end. But this reduction would be compensated by the increased angle of the radiation impinging on M_1 from point P_2. Fig. 12.19 illustrates a case where we should think before blindly plugging into formulas. The physical size of the mirror M_2 has nothing whatsoever to do with the radiation impinging on M_1. We should use the aperture defined by the incident radiation onto M_2 as it is shown, rather than the physical size.

The example shown in Fig. 12.20, of a cavity with a finite aperture medium between the mirrors, is even more practical than the foregoing case. If we assume that the edges of the medium (a discharge tube or a laser rod) are perfectly absorbing, then the cone angles are dictated by the combination of the mirrors, gain medium, and geometry. Here we see the use of a "beam slicer" to extract useful power output and to define the effective aperture

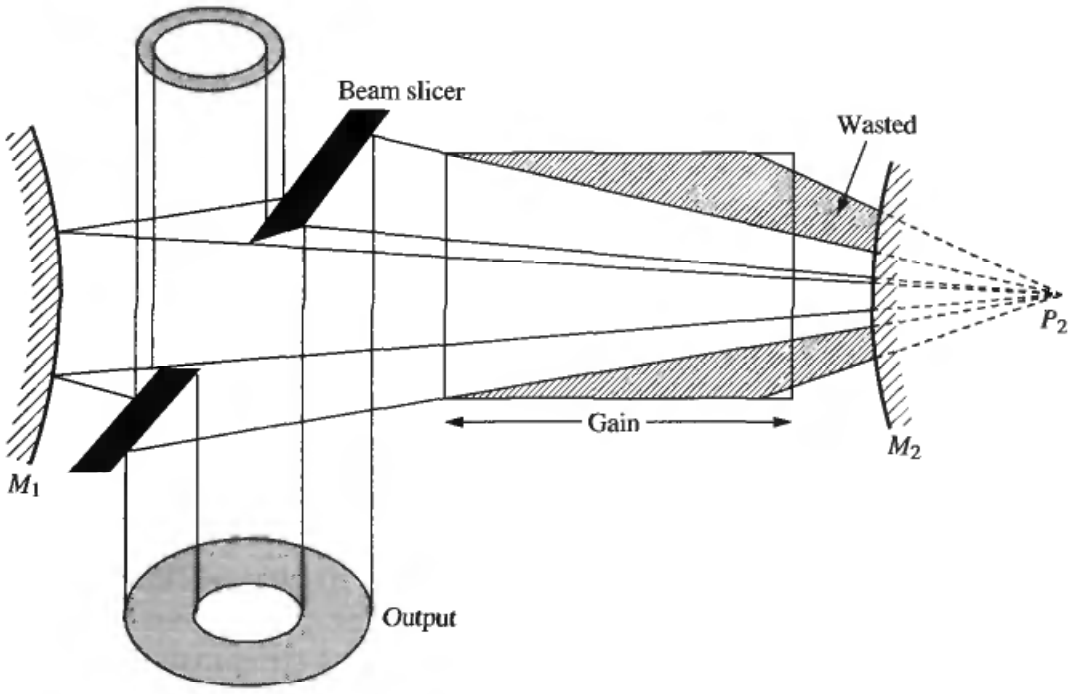


FIGURE 12.20. Laser using an unstable resonator.

of M_1. It is obvious from this figure that one should minimize the amount of power dumped uselessly into the walls of the active medium.

Let us return to the formula for the round trip transmission (12.6.8) and express the result in terms of the g parameters by utilizing the relations between r_1 and r_2 and g_1 and g_2 (12.6.5a) and (12.6.5b). The mean one-way survival factor through the cavity can be expressed as

$$S = \pm \frac{r_1 r_2}{(r_1 + 1)(r_2 + 1)} |\Gamma_1 \Gamma_2|$$

Let us ignore $|\Gamma_1 \Gamma_2|$ for a moment, as being obvious, and correlate the mean photon survival factor with the (lack of) stability of the cavity.

$$S = (\pm) \frac{1 - [1 - (g_1 g_2)^{-1}]^{1/2}}{1 + [1 - (g_1 g_2)^{-1}]^{1/2}} \tag{12.6.9}$$

Note that if $g_1 g_2 > 1$, the quantity involving the square root is less than 1. Hence, the upper choice in sign is applicable in order for the survival factor to be positive and less than 1. If $g_1 g_2 < 0$, the negative sign must be chosen to obtain sensible answers for it. Thus we have a *positive* and a *negative* branch of unstable resonators. We can solve for the contours of equiloss and plot these on a stability diagram:

positive branch	negative branch
$g_1 g_2 > 1$	$g_1 g_2 < 0$
$S = \dfrac{1 - [1 - (g_1 g_2)^{-1}]^{1/2}}{1 + [1 - (g_1 g_2)^{-1}]^{1/2}}$	$S = \dfrac{[1 - (g_1 g_2)^{-1}]^{1/2} - 1}{[1 - (g_1 g_2)^{-1}]^{1/2} + 1}$

Hence,

$$[1 - (g_1 g_2)^{-1}]^{1/2} = \frac{1 - S}{1 + S} \qquad [1 - (g_1 g_2)^{-1}]^{1/2} = \frac{1 + S}{1 - S}$$

Solving for $g_1 g_2$ in terms of S, we obtain

$$g_1 g_2 = \frac{(1 + S)^2}{S} \qquad g_1 g_2 = -\frac{(1 - S)^2}{4S} \tag{12.6.10}$$

Thus the equiloss contours are hyperbolas on the stability diagram. The positive branch corresponds to the first and third quadrants (Fig. 12.21), whereas the negative branch is in the second and fourth quadrants. If, for example, we wanted a cavity with a power loss per pass of 25%, then $S = 0.75$ and

$$g_1 g_2 = \frac{(1.75)^2}{4 \times 0.75} = 1.0208 \qquad (+ \text{ branch})$$

$$g_1 g_2 = -\frac{(0.25)^2}{4 \times 0.75} = -0.0208 \qquad (- \text{ branch})$$

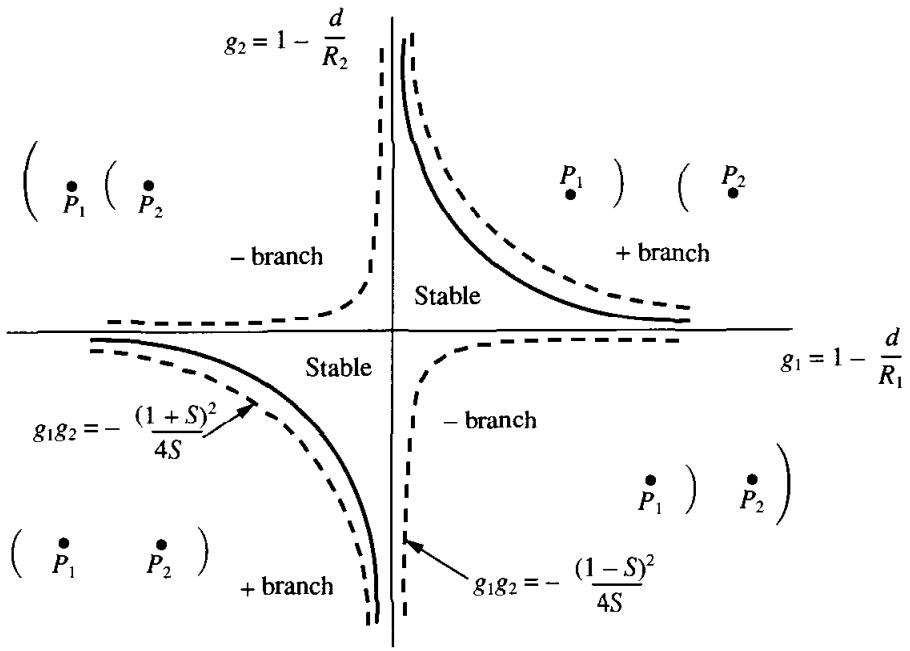


FIGURE 12.21. Equiloss contours for the unstable cavities.

12.6.2 Unstable Confocal Resonator

Consider the problem of an unstable confocal resonator shown in Fig. 12.22, where the mirrors have a common focal point at a distance $d_0/2$ to the left of M_1. Thus $|R_1| = d_0$ and $R_2 = 3d_0$ to have a common focal point.

We can return to the previous formulas for r_1 and r_2 to find an indeterminate relationship of the form $r_1 = N/D = 0/0$. This difficulty can be resolved by recognizing that perfect alignment in spacing is impossible. Thus $d = d_0(1 + \delta)$, where $d_0 = |R_1|$, and then take the limit as our accuracy improves by letting δ approach zero.

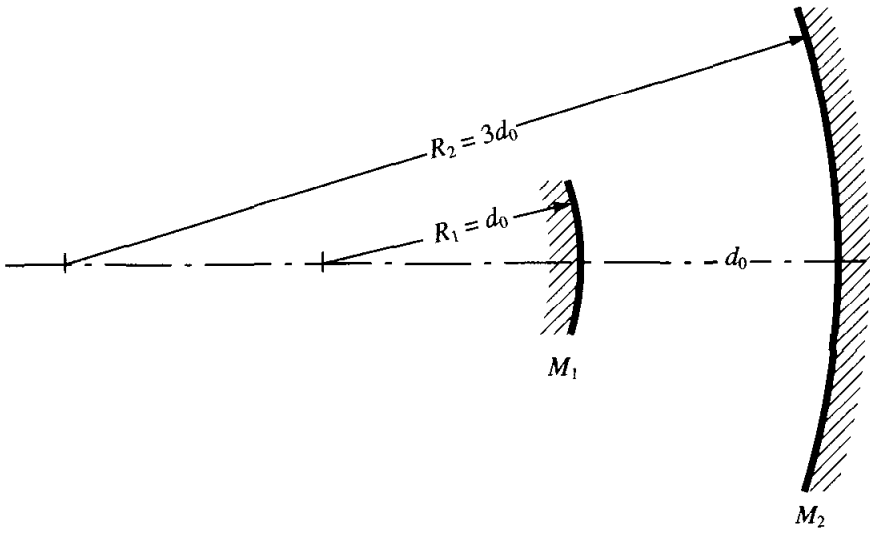
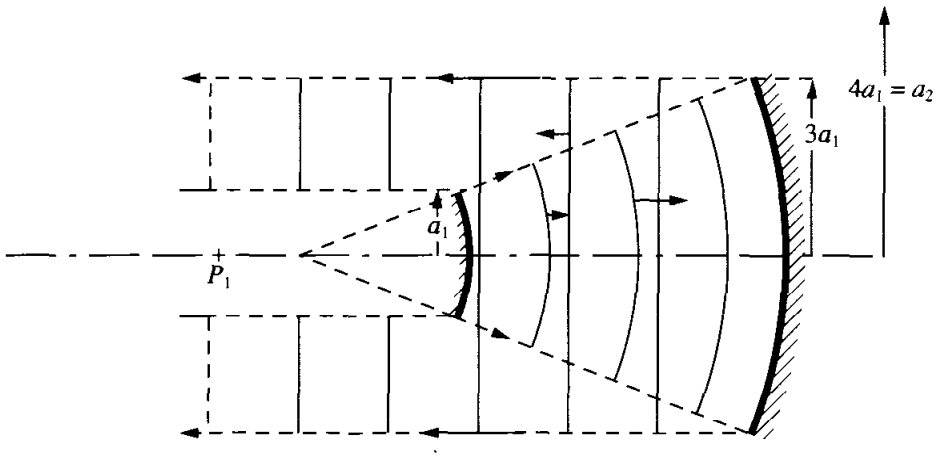
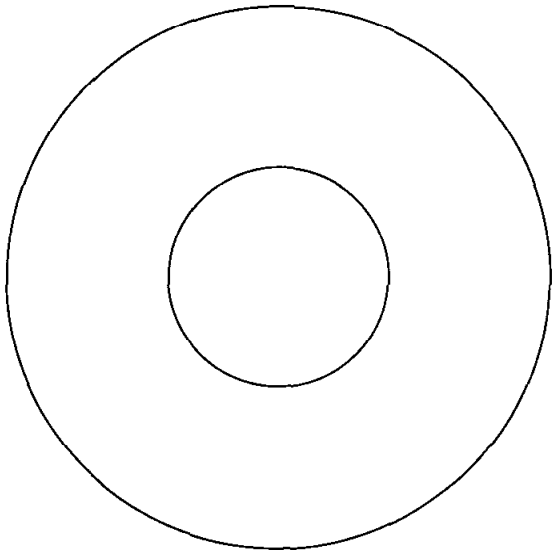


FIGURE 12.22. The unstable confocal resonator.

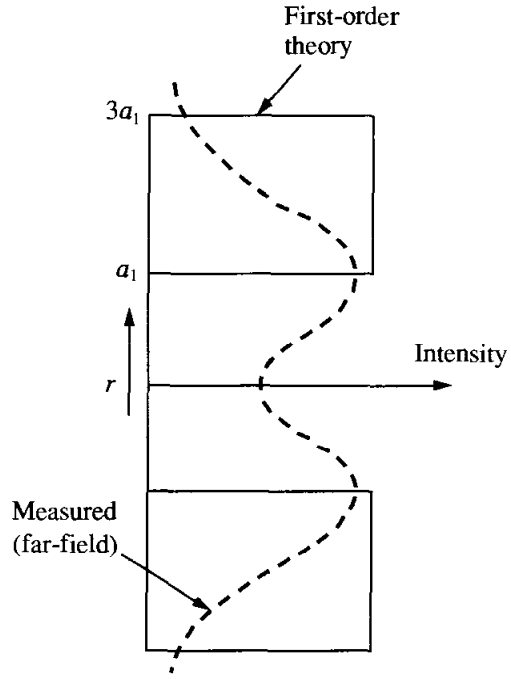
We can circumvent all the dull arithmetic by looking at the physics of the problem and using some reasonable guesswork. If we guess that point P_1 is at the focal point of M_2, the image P_2 is at $-\infty$. Now the image of P_2 in mirror M_1 is the focal point. Hence, we have achieved a self-consistent picture of the virtual sources of the radiation within the cavity. Thus the picture is as shown in Fig. 12.23. Within the context of this first-order theory, the radiation should come out in the form shown. Even though the first-order theory predicts a uniformly illuminated annular region, we know that this is impossible. Diffraction



(a)



(b)



(c)

FIGURE 12.23. (a) Radiation pattern for the cavity of Fig. 12.22. (b) Field distribution. (c) Output intensity.

effects will round off the corners and fill up the center region as will be demonstrated later (see Sec. 12.8). Indeed, an article by Frieberg, Chenausky, and Buczek [2a]* shows that the far-field pattern closely resembles a Gaussian beam.

Incidentally, Fig. 12.24 illustrates the "burn" pattern when a 14-kW CO_2 laser using a confocal unstable resonator (at the University of Illinois) is dumped into a bucket of sand (thereby making a low grade of glass). It is obvious that Fig. 12.23 closely resembles this burn pattern.

The level of sophistication of understanding of the fields in unstable resonators has advanced considerably since the publication of this first-order theory. There are obvious oversimplifications but, amazingly, it does give a good "feel" for the cavity and a reasonably good approximation for the loss per pass. For the case analyzed in this section $g_1 = 1 - [d/(-d)] = 2$, $g_2 = 1 - d/(3d) = 2/3$, $g_1 g_2 = 4/3$, and $S_2 = 1$.

$$S_1 = \frac{\pi a_1^2}{\pi (3a_1)^2} = \frac{1}{9}$$

Therefore

$$S = (S_1 S_2)^{1/2} = \frac{1}{3}$$

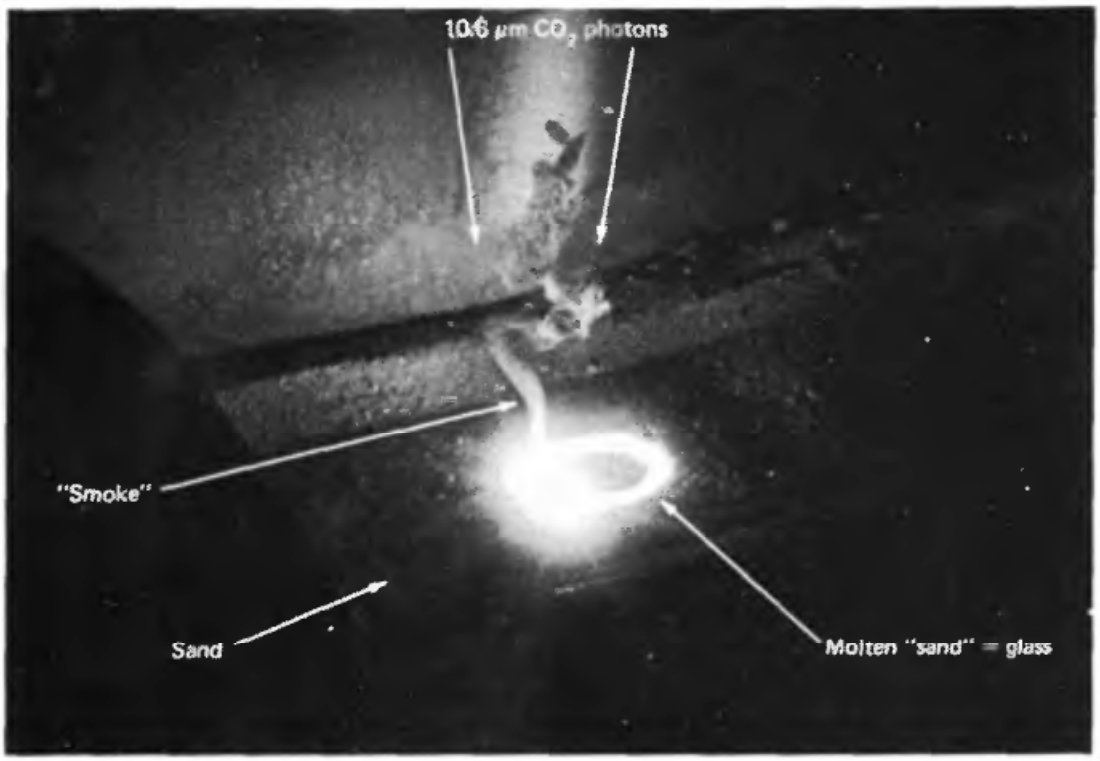


FIGURE 12.24. Burn pattern from a 14-kW CO_2 laser using an unstable resonator. The "sand" is incandescent because of the intense beam. (Photograph by S. Hutchinson.)

*See also Frieberg et al. [26], Fig. 7.

Thus there is a coupling loss of 66% per pass in this cavity. The power gain per pass should be

$$GS > 1 \qquad \text{or} \qquad G > 3 \text{ or } 4.77 \text{ dB}$$

Many lasers have small-signal gains sufficient to meet this criterion.

12.7 INTEGRAL EQUATION APPROACH TO CAVITIES

12.7.1 Mathematical Formulation

Until now, we have used three approaches to optical cavities, which are listed below in order of sophistication along with some obvious difficulties:

1. Treat the fields inside the cavity as *uniform plane waves*. However, there are two problems with this approach: We cannot permit an infinite cross section, so how "big" should it be? Once we have answered that, then how do we account for the part that misses a finite size mirror?
2. Use Hermite-Gaussian beam modes, but these are only applicable for stable cavities and even for those, diffraction effects as a result of the fields at the edges of mirrors are ignored.
3. Use the self-consistent "image" theory of unstable resonators discussed in Sec. 12.6. That theory only applies for unstable cavities and predicts sharp edges to the field distribution. Furthermore, it completely ignores the phase of the fields.

There is another approach that avoids those problems at the expense of mathematical complexity and almost certain absence of analytic solutions except for special circumstances. However, it is perfectly general. In fact, it was the first serious theoretical approach to optical cavities, and the results, which predicted high Q cavities, even for an open resonator such as two plane mirrors, had a tremendous impact on laser research.

In a perfect world, we would prefer a solution to the Helmholtz equation for each Cartesian component of the field in an optical cavity.

$$\nabla^2 E + k^2 E = 0 \tag{12.7.1}$$

satisfying the boundary conditions of a field being reasonably confined to a region close to the axis but otherwise unrestricted. As we might guess, such an ill-posed problem has a small chance of yielding a physically meaningful solution without some approximation such as that used in Chapter 3. However, Gaussian beams are not the characteristic modes of arbitrary cavities and thus we have a problem.

For our purposes, let us pose an equivalent propagation problem illustrated in Fig. 12.25 in which we presume that the field is "known" on the surface (x, y, z) with the task being to compute the field on the surface (x_0, y_0, z_0).

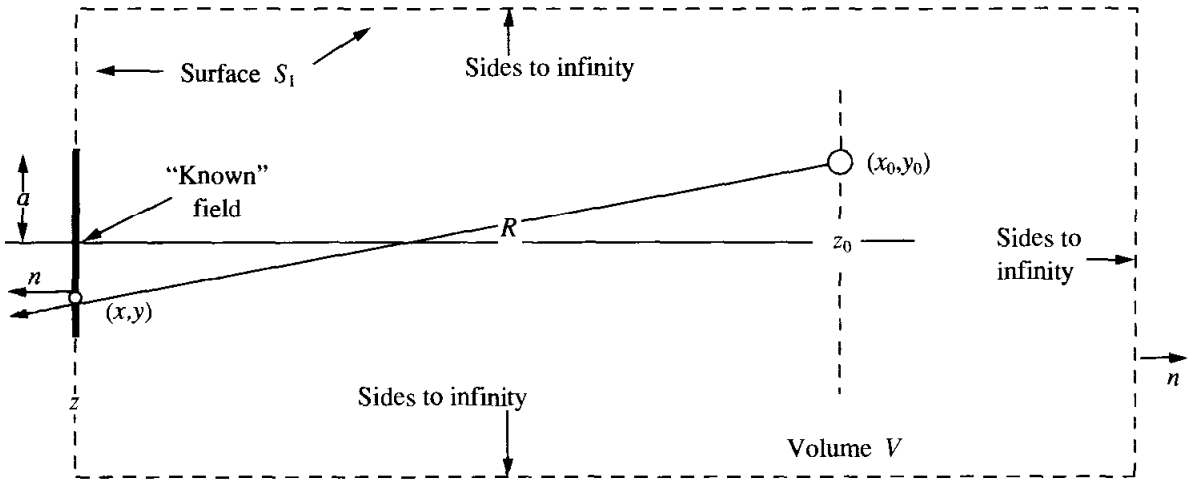


FIGURE 12.25. The geometry for the application of the integral equation.

If we can solve this equivalent problem, we will have "propagated" the known field through the distance $(x_0 - x, y_0 - y, z_0 - z)$ to the new surface, which is precisely the goal of (12.7.1).

In principle, we can always obtain a formal solution to this equivalent problem by using Green's function (i.e., the response of the system at the point $\mathbf{r}$ due to a unit source at $\mathbf{r}_0$). Green functions obey an equation similar to (but worse in complexity than) (12.7.1).

$$\nabla^2 G + k^2 G = -\delta(x - x_0)\delta(y - y_0)\delta(z - z_0) \tag{12.7.2}$$

We choose a volume V with one surface coinciding with the one on which the field is "known" and let the other surfaces of the volume go to ∞, where all fields (and G) are zero. Now we multiply (12.7.1) by G, (12.7.2) by E and integrating the difference over the volume V.

$$\iiint [G\nabla^2 E - E\nabla^2 G]\, dV = \iiint E\delta(x-x_0)\delta(y-y_0)\delta(z-z_0)\, dV = E(\mathbf{r}_0) \tag{12.7.3}$$

Because of the sampling properties of the δ function the right side of (12.7.3) is precisely the desired quantity, the field at (x_0, y_0, z_0). Unfortunately, the left side of (12.7.3) appears rather formidable. However, it can be simplified by the use of some well-known identities from vector calculus.

$$\nabla \cdot (a\mathbf{A}) = a\nabla \cdot \mathbf{A} + \mathbf{A} \cdot \nabla a \tag{12.7.4a}$$

and

$$\iiint \nabla \cdot \mathbf{B}\, dV = \oiint \mathbf{B} \cdot \mathbf{n}\, dA \tag{12.7.4b}$$

We apply (12.7.4a) twice: assign $\mathbf{A}$ to be ∇E and a being G, reverse the choices and subtract, and then recognize that the bracket in (12.7.3) can be expressed as the divergence of another vector $\mathbf{B}$. Thus, the result can be converted from a volume integral to one over

the surface of the volume by (12.7.4b) (which is a derivation of Green's theorem).

$$E(\mathbf{r}_0) = \oiint [G\nabla E - E\nabla G] \cdot \mathbf{n}\, dA \tag{12.7.5a}$$

where $\mathbf{n}$ is the outward normal to the surface enclosing the volume V. This is an important formula for it states that the field on the interior point ($\mathbf{r}_o$) of the volume V is related to G and E on the surfaces bounding that volume. Only one surface is important—where the field is known—with the other surfaces being at $\{\infty\}$ where both E and G vanish. Hence, we can drop the closed surface symbol and write

$$E(\mathbf{r}_0) = \iint_{S_1} (G\nabla E - E\nabla G) \cdot \mathbf{n}\, dA \tag{12.7.5b}$$

where $\mathbf{n}$ is the outward normal to the surface S_1. Thus to evaluate the field at the point r_0, we only need E, ∇E, G, and ∇G on the heavily shaded part of the surface S_1. The value of E and ∇E there is simple since the field is presumed to be known in the immediate neighborhood of the input plane. However, it does not appear that we have made much practical progress: in order to "propagate" the known field from $\mathbf{r}$ to $\mathbf{r}_0$ by this approach, we need G, which obeys an even more complicated equation than does E.

The utility of (12.7.5) lies in the fact that we can inject some insight into the mathematical onslaught by using the physical definition of G, which is the field at point $\mathbf{r}$ in response to a unit source at $\mathbf{r}_0$. If we are dealing with a uniform dielectric, then this is just a spherical wave of unit amplitude.

$$G = \frac{1}{4\pi |R|} e^{-j\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_0)} = \frac{1}{4\pi |R|} e^{-j\phi(r,r_0)} \tag{12.7.6}$$

where $\phi = \mathbf{k} \cdot (\mathbf{r} - \mathbf{r}_0)$ is phase shift experienced by a plane wave propagating *from* $\mathbf{r}_0$ *to* $\mathbf{r}$ and $|R| = |\mathbf{r} - \mathbf{r}_0|$ is distance between the two points.

We use (12.7.6) in (12.7.5b) with the following (minimal) restrictions.

1. The field on the surface S_1 is limited in spatial extent so that for $r \gg$ some radial distance a, $E \sim 0$. In other words, we have a "beam" incident on the heavily shaded part of Fig. 12.25.
2. We also presume that the $(z_0 - z)$ is large enough so that all angles between the z axis and R is small for any combination of (x, y) and (x_0, z_0).
3. The medium is uniform.
4. We also assume that the field E is propagating mostly in the z direction and thus $E \approx f(x, y)e^{-jkz}$.

These restrictions amount to requiring that (x_0, y_0, z_0) be in the far field of (x, y, z), which for optical frequencies is almost always true for any ruler size distances $(z_0 - z)$.

Thus

$$\mathbf{n} \approx -\mathbf{a}_z \text{(since all angles are small)}$$

For

$$E = f(x, y)e^{-jkz}$$

at the surface S_1, we have

$$\mathbf{n} \cdot \nabla E = (-\mathbf{a}_z) \cdot \nabla E = (-\mathbf{a}_z) \cdot (-jk\mathbf{a}_z)E = +jkE \tag{12.7.7}$$

The gradient of G involves a differentiation of the product of the $1/|R|$ (the amplitude factor) and the exponential term representing the phase shift $\phi(\mathbf{r}, \mathbf{r}_0)$ going from $\mathbf{r}_0$ to $\mathbf{r}$. The latter is overwhelmingly dominant because the phase goes through a complete cycle (2π) every time the distance changes by λ (thus the exponential term goes through all complex values of 1), whereas the distance $|R|$ experiences just a small fractional change of the order of $\sim \lambda/d \ll 1$.

For small angles, the phase term can be expressed as $\phi = (k \cos\theta)(z_0 - z)$ and thus ∇G is given by*

$$\nabla G = \nabla\left(\frac{1}{4\pi|R|} e^{-j\phi(r,r_0)}\right) \approx jk \cos\theta \mathbf{a}_z \left(\frac{1}{4\pi|R|}\right) e^{-j\phi(r,r_0)} \tag{12.7.8a}$$

Thus,

$$\mathbf{n} \cdot \nabla G = -jk \cos\theta \left(\frac{1}{4\pi|R|}\right) e^{-j\phi(r,r_0)} \tag{12.7.8b}$$

Substituting (12.7.7) and (12.7.8b) into (12.7.5) yields

$$\boxed{E(r_0) = jk \iint_{S_1} \frac{E(r)(1 + \cos\theta)}{4\pi|R|} e^{-j\mathbf{k}\cdot\mathbf{R}}\, dA} \tag{12.7.9}$$

Two more approximations are in order. For most of our applications the distance from the source point on S_1 to r_0 does not change by much for all possible combinations of (x, y) and (x_0, y_0). Hence we set $|R| = |z - z_0|$ in the denominator and pull it outside of the integral. The angle θ is small and hence $1 + \cos\theta \sim 2$. However, the full expression for R must be used in the exponent since small differences in $|R|$ get multiplied by a very large number, $k = 2\pi/\lambda$. With these assumptions we have

$$\boxed{E(r_0) = \frac{j}{\lambda_0(|z - z_0|)} \iint_{S_1} E(r) e^{-j\mathbf{k}\cdot\mathbf{R}}\, dA} \tag{12.7.10}$$

This is our fundamental result and is the starting point for many tasks.

A.G. Fox and T. Li [5] used (12.7.10) to determine the characteristic field distribution in an optical cavity with open sides in the manner shown in Fig. 12.26.

They overlaid a very simple logic on their analysis. They invoked an obvious physical constraint on the allowed fields or modes at the cavity mirror.

> The cavity modes must be field distributions that repeat, to within a constant scale factor, after a round trip through the cavity.

*Remember, E propagates from z to z_0 and G in the opposite direction.

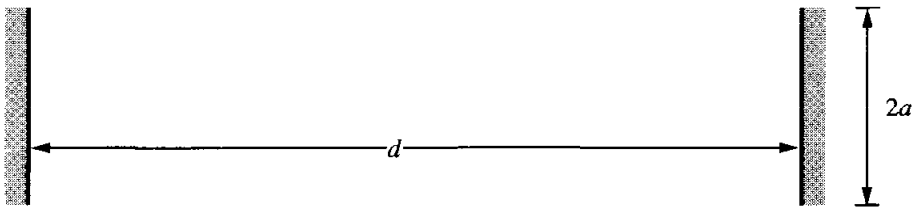


FIGURE 12.26. The strip mirror analyzed by Fox and Li [5].

This is obvious, in hindsight (but they were the first to state it explicitly) and leads to the following analytic (or numerical) procedure.

1. Assume a field on M_1, which is now the "known" field of this section with the surface S_1 being the mirror M_1. We might think of $E(r)$ in (12.7.10) as being the field reflected from M_1. This initial guess is probably wrong, but the following procedure corrects the guess.
2. Use (12.7.10) to predict the field incident on M_2 based upon the assumed distribution in step 1.
3. Multiply the field found in step 2 by the field reflection coefficient to obtain the field starting back toward M_1. If the mirror has a tapered reflectivity, insert that functional form at this point in the analysis. In any case, the mirror is of a finite size, so the field starting back is of limited extent.
4. The field reflected from M_2 becomes a new "known" field $E(r)$ to be used in (12.7.10) to compute the field incident on M_1, and from that value and the reflection coefficient of M_1 we obtain the field that has survived one round trip.

This field may have a considerably different shape than the initial choice, in which case we have *not* found a characteristic mode of the cavity. However, we have found a second guess that can be used in step 1 to 3 to generate a third guess and so on.

The critical physics is contained in the following mathematical statement. Is

$$\boxed{f^{(n+1)}(x, y) \stackrel{?}{=} S^{1/2} f^{(n)}(x, y)} \tag{12.7.11}$$

where S is a constant* (more or less) (and it must be < 1) and the superscripts on f imply the number of round trips. If the answer is yes, a characteristic mode of the cavity has been found. If the answer is no, then the field is sent back and forth (numerically) until it does obey (12.7.11).† A discussion of the results of Fox and Li is in the next section.

12.7.2 Fox and Li Results

It is just a matter of programming, computer time, and patience to obtain the characteristic field distributions of various cavity configurations using (12.7.10). A. G. Fox and T. Li [5]

*As we will see, the parameter S is our old friend, the survival factor.

†The number of round trips is not significant for if one happens to make a lucky guess then $f^{(2)} = S^{1/2} f^{(1)}$.

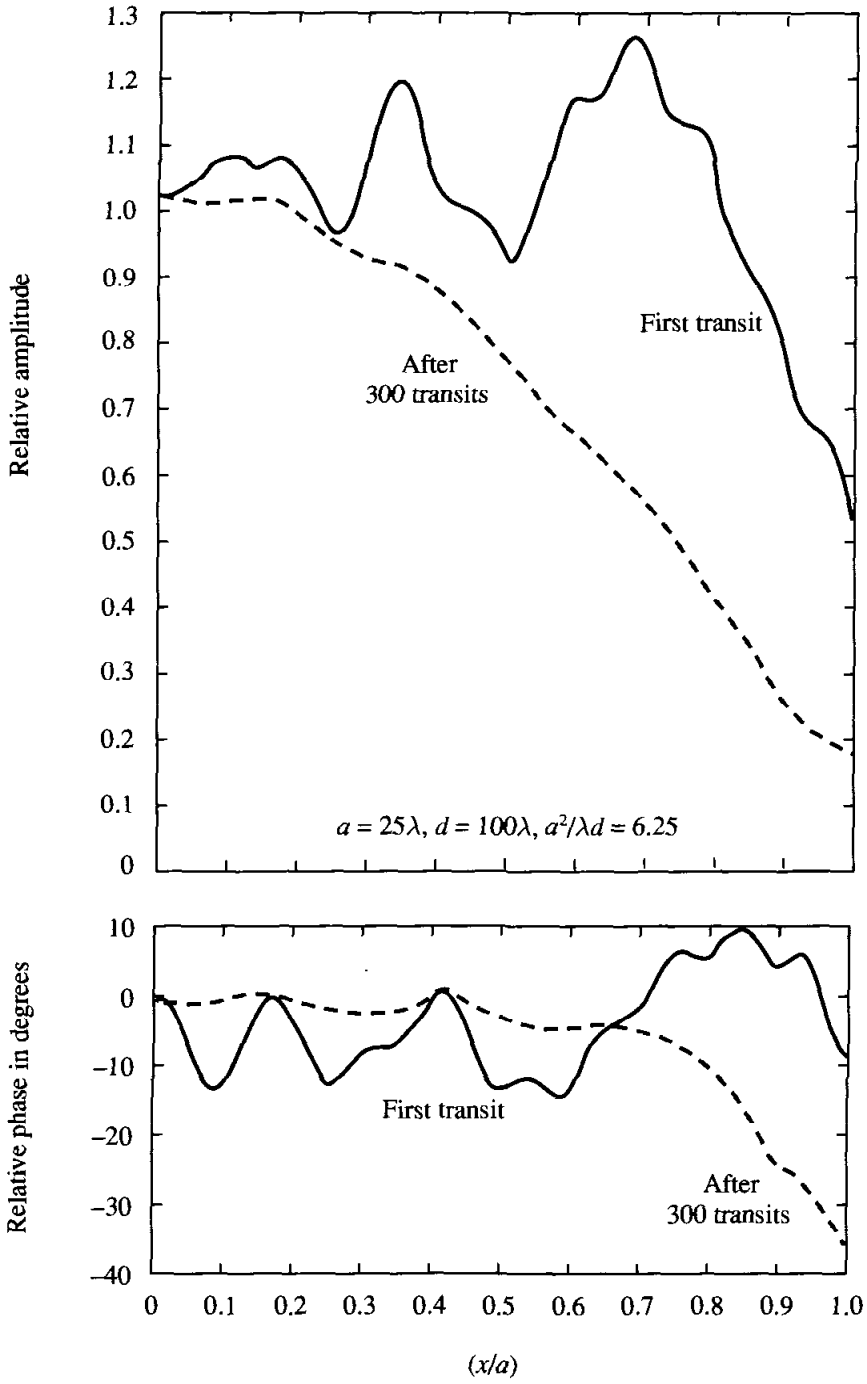


FIGURE 12.27. Amplitude and phase of the fields on infinite strip mirrors. The fields were started with a uniform amplitude $E(x) = E_0$ for $|x| < a$ (zero otherwise) and with a constant phase. (From Fox and Li [5].)

started with the easiest geometry of all, two identical strips of width $2a$ and infinitely long (in and out of paper in Fig. 12.26). The initial guess was simply a uniform field with a constant phase over the strip, zero otherwise. Fig. 12.27 is a reproduction of their Fig. 5 showing the resulting pattern with the field normalized to be 1 at the center of the strip mirror, after one transit. Note that there are some rather violent gyrations for this case. After 300 transits, the field settles down and starts looking more or less like a Gaussian beam.

The lower part of this figure illustrates the phase of the fields on one of the mirrors referenced to that at the center. The initial wave was started with constant phase independent of x. For every "bump" seen in the top for the amplitude, there is a corresponding gyration in phase for one transit. After 300 transits, and having achieved a self-consistent reproduction, the phase is just about as smooth as the amplitude with the field at edge lagging that at the center. In other words, even though the mirrors are flat, the phase front is *not*; that is, the mirrors are not surfaces of constant phase.

Given the fields, it is just a matter of computing the fraction of the power intercepted by each mirror and identifying the remainder (the part that misses) as the diffraction loss for the cavity. This fraction, taken from Fig. 8 of Fox and Li is shown in Fig. 12.28 and

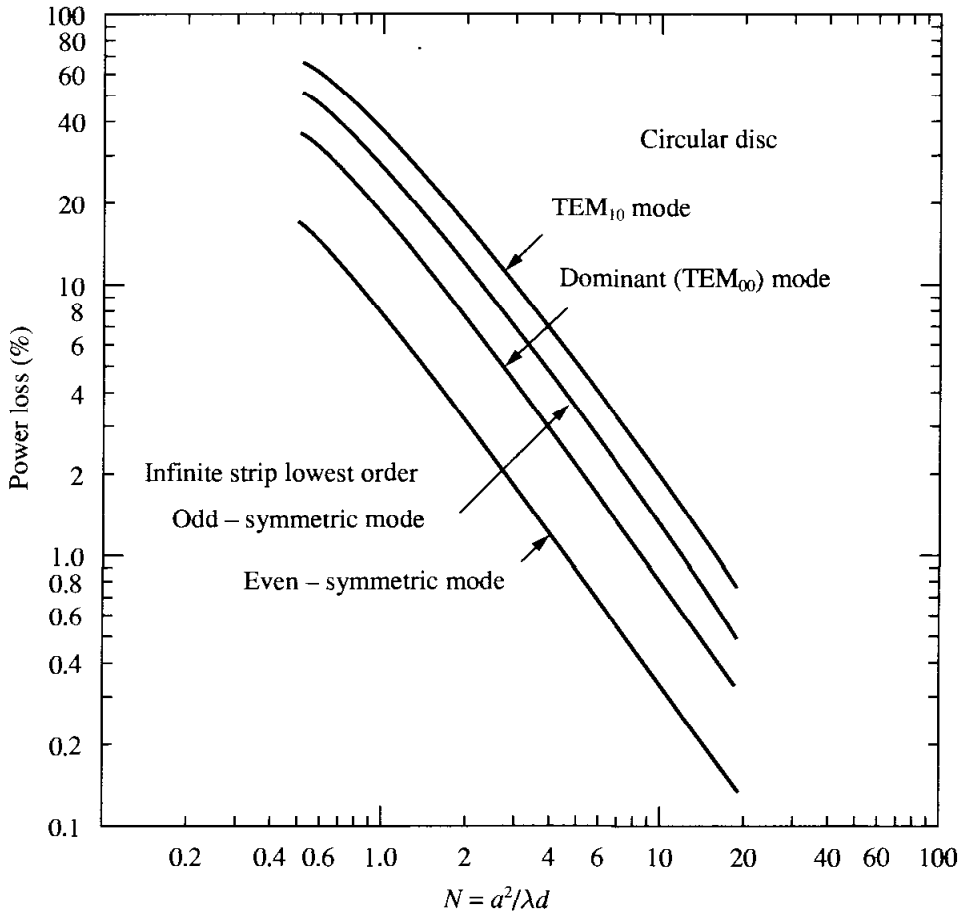


FIGURE 12.28. Losses in an open cavity per transit for a strip mirror of width $2a$ and a circular disk of radius a. (From Fox and Li [5], Fig. 8.)

confirms our intuition: large mirrors separated by reasonable distances have small losses. Now, however, it is given a scientific measure.

Let us take a numerical example to appreciate what the data of Fig. 12.28 are showing. As will become obvious in the analytic approach, the critical scaling parameter for all optical cavities is the *Fresnel* number defined by

$$N = \frac{a^2}{\lambda d} \tag{12.7.12}$$

Choose $N = 10$ and get calibrated as to what kind of cavity is being analyzed. We chose $a = 1.25$ cm, and thus the circular diameter is $\sim$ 1", about the size of a quarter of U.S. currency. Let us also choose $\lambda = 1\mu$m and solve for the distance d from the above equation to find that $d = 1563$ cm $= 15.6$ m or 51.3 ft.

> A serious student should pick up two quarters and try to imagine them as polished mirrors facing each other at this distance. We would guess that there is virtually no chance of a beam being contained between those mirrors.

Now consult Fig. 12.28 and find that the loss per pass is *only 1.2%*.

The 15.6 m cavity is excessively long. If we made our cavity with quarters a more reasonable distance—say, 200 cm—then the Fresnel number is 78.1 and the loss per pass owing to diffraction around (i.e., missing) the mirror is only 0.08%. We cannot make a reflecting surface that good. (Of course the alignment of two plane mirrors is a tough problem in itself.)

These results had a tremendous impact on laser research. Fox and Li demonstrated that it was possible to have a very high Q optical cavity with open sides, a fact that should have been known for many years prior to the invention of the laser but was not. The laser gain equation was in text books as early as 1923 (although those words were not used), and the concept of a negative temperature (or inversion) was also known. The last necessary tool was the feedback system, and Fox and Li provided a significant initial impetus there.

12.7.3 Stable Confocal Resonator

There is only one geometry that yields an analytic answer with a minimum number of approximations: This is the confocal geometry consisting of two identical "square" mirrors with a common radius of curvature $R = b$, which are separated by a distance $d = b$, and thus share a common focal point as shown in Fig. 12.29.

Also shown in Fig. 12.29 are the equations describing the surfaces of the two mirrors, which are needed to compute the phase shift of a wave originating from dx_1, dy_1, at M_1 and propagating to x_2, y_2 on M_2.

$$\phi(2, 1) = k[(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2]^{1/2} = kR \tag{12.7.13}$$

After considerable uninspiring arithmetic, we find that the distance R from (x_1, y_1, z_1) to (x_2, y_2, z_2) can be approximated by

$$R = b - \frac{x_1 x_2}{b} - \frac{y_1 y_2}{b} \tag{12.7.14}$$

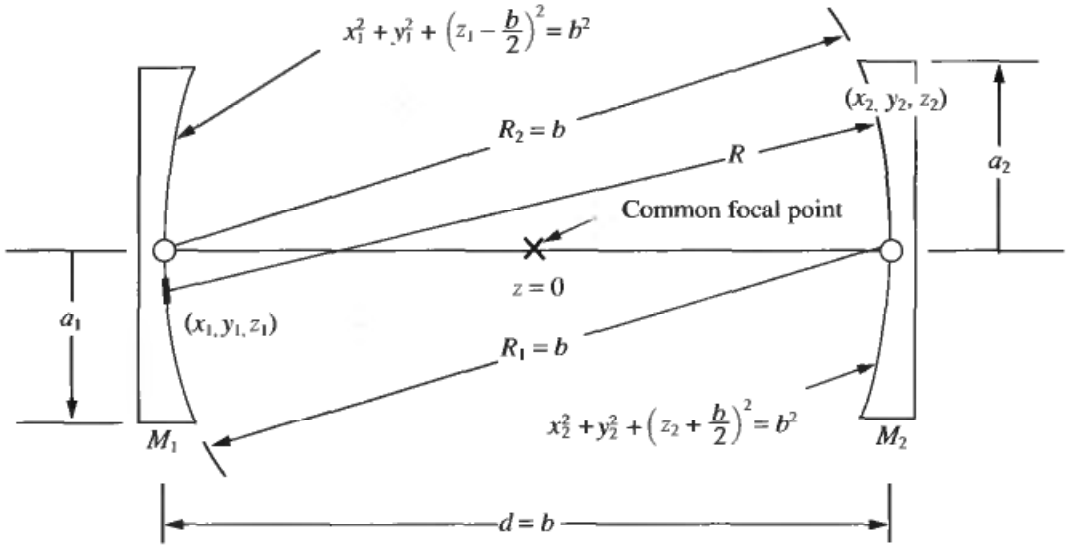


FIGURE 12.29. The confocal resonator.

which is equivalent to approximating the spherical mirror surface by a parabolic one. Thus (12.7.10) can be written as

$$E(x_2, y_2) = \frac{k}{2\pi d} e^{-j(kb-\pi/2)} \iint_{M_1} E(x_1 y_1) e^{jkx_1x_2/b} e^{jky_1y_2/b}\, dx_1\, dy_1 \tag{12.7.15}$$

where the multiplicative factor of j in (12.7.10) has been expressed as $\exp(j\pi/2)$ and combined with the axial phase shift $kd = k \cdot b$.

It is very important to keep in mind the "game plan." We are computing the field on M_2 in terms of that on M_1, which in turn can be computed in terms of that on M_2. If we had two dissimilar mirrors, we would have to complete the loop and demand self-consistency for the round trip. For the case considered here, we can cheat a bit and merely require that the field at M_2 be a scaled replication of the field on M_1. The format of (12.7.15) suggests factoring the field into a product. That is,

$$E(x_1, y_1) = f(x_1)g(y_1) \tag{12.7.16a}$$

and

$$E(x_2, y_2) = [\sigma_x f(x_2)][\sigma_y g(y_2)] \tag{12.7.16b}$$

where σ_x, σ_y are complex constants to be determined. Equation (12.7.15) can be factored and rewritten

$$\sigma_x f(x_2) = \left[\frac{k}{2\pi d} e^{-j(kb-\pi/2)}\right]^{1/2} \int_{-a_1}^{+a_1} f(x_1) e^{jkx_1x_2/b}\, dx_1 \tag{12.7.17}$$

with a corresponding equation for $g(y)$:

$$\sigma_y g(y_2) = \left[\frac{k}{2\pi d} e^{-j(kb-\pi/2)}\right]^{1/2} \int_{-a_1}^{+a_1} g(y_1) e^{jky_1y_2/b}\, dy_1 \tag{12.7.18}$$

Let us define some normalized quantities. Let

$$\boxed{N = \frac{a^2}{\lambda d} = \text{Fresnel number}} \tag{12.7.19}$$

$$C = 2\pi N \tag{12.7.20a}$$

$$U_{1,2} = \frac{x_{1,2}}{a}\sqrt{C} \qquad V_{1,2} = \frac{y_{1,2}}{a}\sqrt{C} \tag{12.7.20b}$$

Then (12.7.17) can be written in terms of normalized variables:

$$\sigma_x f(U_2) = \left[\frac{ka^2}{2\pi dC}\right]^{1/2} \left\{e^{-j(kb-\pi/2)}\right\}^{1/2} \int_{-\sqrt{C}}^{\sqrt{C}} f(U_1) e^{jU_1U_2} dU_1 \tag{12.7.21}$$

In (12.7.10) through (12.7.21) there is a "hint" of a Fourier transform creeping into view. It becomes most evident if (12.7.19) and (12.7.20) are used for the first prefactor of (12.7.21).

$$\frac{ka^2}{2\pi dC} = \frac{1}{2\pi}$$

Hence, (12.7.21) appears as a finite Fourier transform:

$$\sigma_x f(U_2) = \left\{e^{-j(kb-\pi/2)}\right\}^{1/2} \left\{\frac{1}{\sqrt{2\pi}} \int_{-\sqrt{C}}^{\sqrt{C}} f(U_1) e^{jU_1U_2} dU_1\right\} \tag{12.7.22}$$

If $C \to \infty$, then it *is* a simple Fourier transform and the solution for $f(U_1)$ is a Gaussian. In other words, suppose $f(U_1) = \exp -(U_1^2/2)$. Then (12.7.22) can be manipulated as follows:

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(U_1) e^{jU_1U_2} dU_1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\left(\frac{U_1^2}{2} - jU_1U_2\right)\right] dU_1$$

The exponent can be expressed as a perfect square plus an extra term.

$$\frac{U_1^2}{2} - jU_1U_2 = \left(\frac{U_1}{\sqrt{2}} - j\frac{U_2}{\sqrt{2}}\right)^2 + \frac{U_2^2}{2} = \frac{W^2}{2} + \frac{U_2^2}{2} \tag{12.7.23a}$$

where

$$W = U_1 - jU_2 \tag{12.7.23b}$$

Thus the integral becomes

$$I = e^{-U_2^2/2} \left\{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-W^2/2}\, dW\right\} = e^{-U_2^2/2} \tag{12.7.24}$$

Now there is a key issue that is in danger of being obscured by the smoke of mathematics. We assumed a field varying as $\exp -U_1^2/2$ on M_1 and found a field identical to it on M_2 (since U_1 and U_2 are equivalent variables merely expressing a distance along x at

the two mirrors). Because we choose a symmetrical cavity, the return trip will obviously reproduce the original field and hence we have identified a characteristic *mode* of the cavity. The assumption of $C \to \infty$, that is, infinite mirror size, leads to $|\sigma_{x,y}| = 1$. (For finite mirrors, the amplitude would be less than 1.) The total field is given by

$$E(U_2, V_2) = e^{-j(kb-\pi/2)} \exp[-(U_2^2 + V_2^2)/2] \tag{12.7.25}$$

Now it is appropriate to remove the normalizations on the spatial variables:

$$\frac{U_1^2}{2} = \frac{x_1^2 C}{2a^2} = \frac{x_1^2}{2a^2} \cdot \frac{2\pi a^2}{b\lambda} = \frac{x_1^2}{b\lambda/\pi} \equiv \frac{x_1^2}{w_s^2} \text{ and } \frac{V_1^2}{2} = \frac{y_1^2}{w_s^2}$$

where $$w_s^2 = \frac{b\lambda}{\pi} = \left(\frac{2b}{k}\right) \tag{12.7.26}$$

The parameter w_s is the spot size (e^{-1}) of the Gaussian beam at the spherical mirror and is *precisely* the result that was obtained in Chapter 5 for this geometry.

Resonance can be found directly from the phase of (12.7.25).

$$kb - \frac{\pi}{2} = q\pi$$

$$\nu = \frac{c}{2d}\left(q + \frac{1}{2}\right) \tag{12.7.27}$$

where the fact that $b = d$ has been used. This can be compared directly with (6.5.4) for $m = p = 0$:

$$\nu = \frac{c}{2d}\left[q + \frac{1 + m + p}{\pi} \cos^{-1}(g_1 g_2)^{1/2}\right] \tag{6.5.4}$$

Since

$$g_1 = 1 - \frac{d}{R_1} = 0 = g_2 = 1 - \frac{d}{R_2} \qquad \therefore \cos^{-1}(g_1 g_2)^{1/2} = \frac{\pi}{2}$$

and

$$\nu = \frac{c}{2d}\left(q + \frac{1}{2}\right)$$

and thus we arrive at precisely the same result for the resonant frequency.

For finite-sized mirrors, some of the field from M_1 "misses" M_2 and represents a power loss. A numerical evaluation of (12.7.22) leads to the fact that $|\sigma_{x,y}| < 1$. Fox and Li also evaluated this loss, and their results are plotted in Fig. 12.30. Note that the loss is very small even for a low Fresnel number. Fig. 12.31 shows the loss (per pass) as a function of the $g_{1,2}$ parameters going from a stable cavity to the unstable version covered earlier, both treated with the integral equation approach. The dashed curves for the unstable cavity represent the loss predicted by the simple geometric optics approach of Sec. 12.6.

The integral equation formulation for optical cavities is a very powerful tool. It obviously requires considerably more mathematics than does the simple Gaussian beam optics of earlier chapters, but as the earlier calculation indicates, the answers are identical, as they

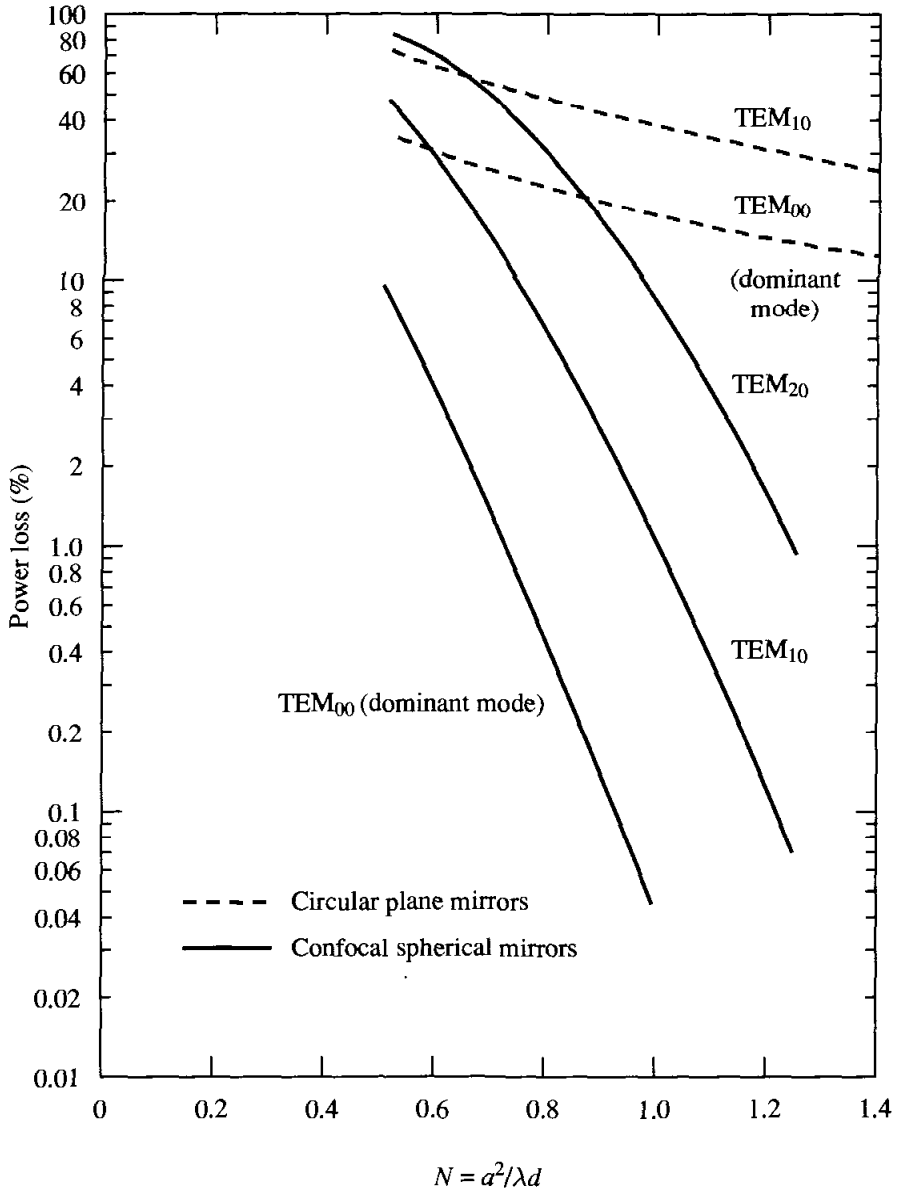


FIGURE 12.30. Power loss per transit versus $N = a^2/\lambda d$ for confocal spherical mirrors. (Dashed curves for circular plane mirrors are shown for comparison.) (From Fox and Li [5].)

should be, for cases that can be analyzed by either approach. This should inspire confidence in both approaches. Unfortunately, the integral equation approach most often requires a computer solution similar to that used by Fox and Li. Only in contrived cases, such as the confocal geometry, can an analytic solution be obtained. But let us face it, computer time is cheap, and thus we have a quite general approach to solving for the modes in *any* cavity.

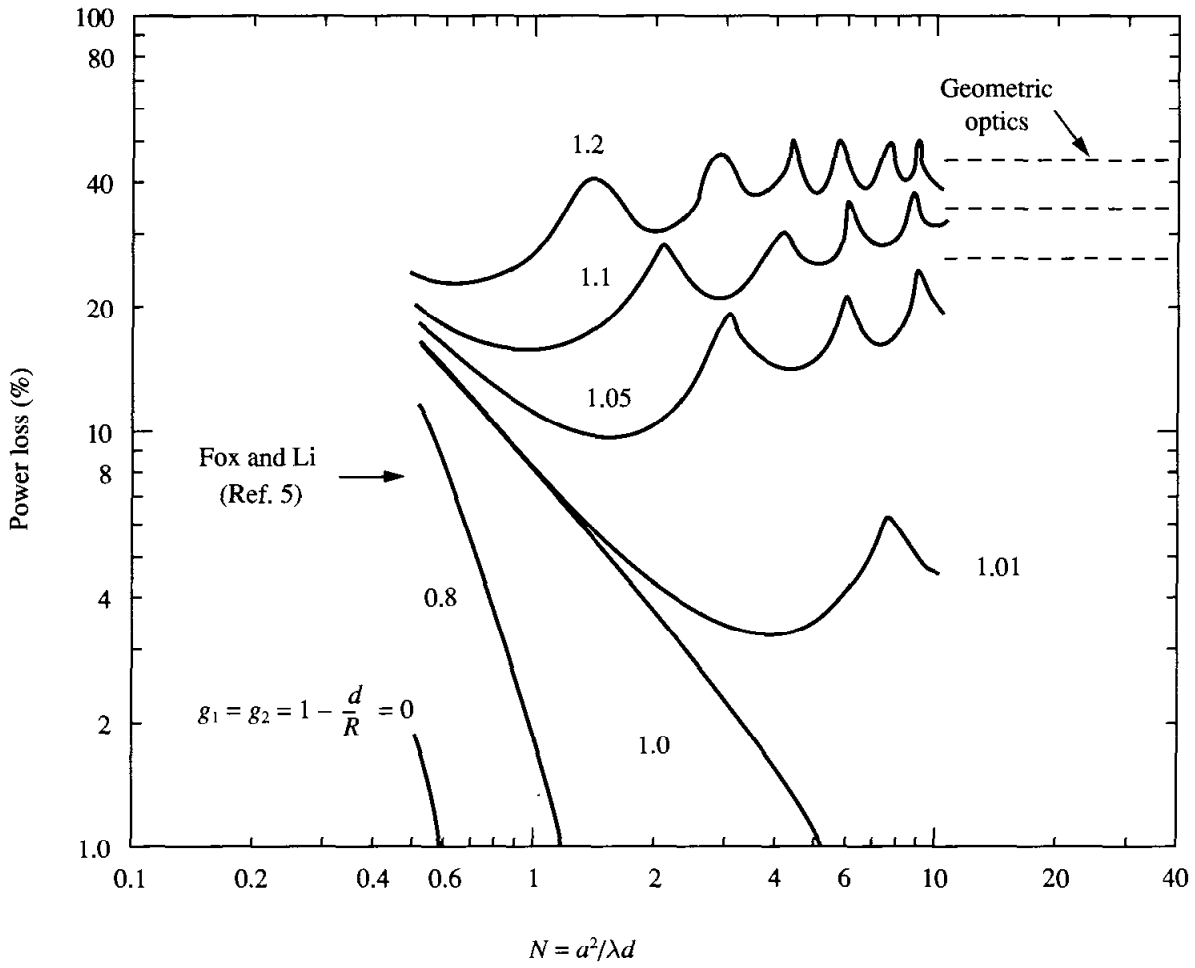


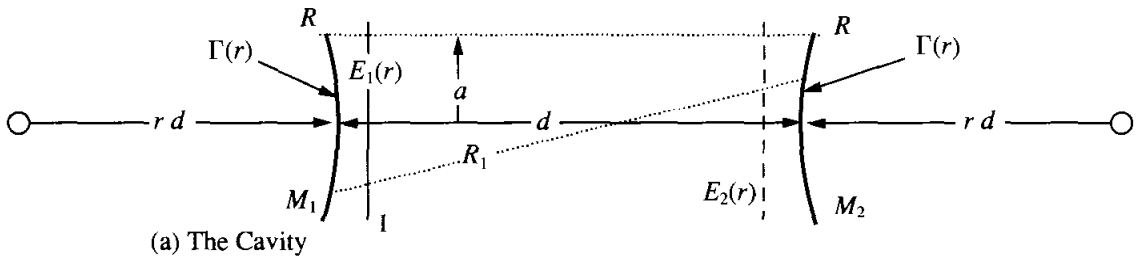
FIGURE 12.31. Loss per bounce versus Fresnel number for stable and unstable resonators. (From Siegman [1].)

12.8 FIELD ANALYSIS OF UNSTABLE CAVITIES

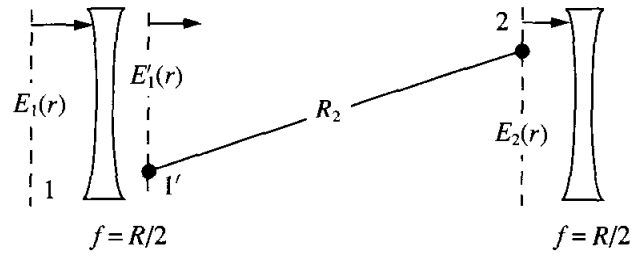
The integral equation approach to optical cavities is very general, is subject only to the minimal restrictions of open cavities (as we have formulated it), and is very "forgiving" in the sense that the DO loop will converge to the lowest loss mode. Hence, it is just a matter of computer time to generate numerical data for the fields and the loss per pass for any cavity configuration, stable or unstable. This was done by Siegman (Ref. 2c), and the percentage loss per pass is shown in Fig. 12.31 for the unstable cavities discussed earlier.

As valuable as a computer analysis is, there is considerable merit in attempting to obtain an analytical solution, or at least to reduce and compact the problem to the point where trends can be easily discerned. For instance, in Fig. 12.31, there are undulations around the geometric optic value in the loss per pass for the unstable resonators. These are real effects (not computer program instabilities), but the physical cause is not immediately obvious.

Thus, the problem to be addressed by the integral equation is shown in Fig. 12.32. It is similar to Fig. 12.17 with the added complication that the mirrors may have a (tapered)



(a) The Cavity



(b) The equivalent Lens waveguide

FIGURE 12.32. The unstable cavity and the equivalent lens waveguide.

reflectivity, which depends upon the distance from the optical axis but have identical radius of curvature. Our goal is to find the field E_2 *incident* on M_2, which is a scaled replica of the field E_1 *incident* on M_1. Because of the symmetric specification, we need only to chase the field from one mirror to the other and demand a replica.

$$E_2(r_2) = \frac{j2k}{4\pi d} \int_{M_1} E_1(r_1) e^{-jkR_1}\, dA_1$$

with $E_2(r_2) = \sigma_x \sigma_y E_1(r_1)$ for a mode.

The major arithmetic nightmare facing us is to compute the distance R_1 between two arbitrary points on the two mirror surfaces. While it can be done by brute force, there is a simple way of avoiding that task by considering the lens waveguide, which is equivalent to the cavity and is shown below it. It would simplify matters greatly if we could first relate $E_1'(r_1)$ to $E_1(r_1)$ and then find the distance R_2 between points on *planes* $1'$ and 2, which is then trivial.

The amplitude of the field is changed by the reflection coefficient, and thus the amplitude of the field at plane $1'$ is related to the incident value by $|E_{1'}(r)| = \Gamma(r)|E_1(r)|$. However, the phase is also changed by virtue of the curved reflecting surface, which can be represented by the propagation through the equivalent thin lens:

$$t = e^{j\phi_1(r)} = e^{j(kr^2/2f)} \tag{2.13.2}$$

where t is transmission coefficient of a lossless lens. Thus,

$$E_{1'}(r_1) = \Gamma(r) e^{j(kr^2/2f)} E_1(r_1) \tag{12.8.1}$$

and our integral equation becomes

$$E(r_2) = j \frac{2k}{4\pi d} \int_{M_1} [\Gamma(r) e^{+j(kr^2/2f)}] E_1(r_1) e^{-jkR_2}\, dA_1 \tag{12.8.2}$$

where R_2 is the distance between the two planes 2 and 1′ as shown in Fig. 12.32(b) and is given by

$$R_2 = \sqrt{d^2 + (x_2 - x_1)^2 + (y_2 - y_1)^2} \approx d + \frac{(x_2 - x_1)^2}{2d} + \frac{(y_2 - y_1)^2}{2d} \tag{12.8.3}$$

Now comes the task of compacting (12.8.2) to make it manageable. Our earlier studies in Chapters 3 to 6 indicated the importance of the g parameters, and the geometric optic approach to unstable cavities introduced the normalized distances r, which were related to g [cf., (12.6.5)], or equivalently the magnification **M**. Let us try to fold in the insight provided by the geometric optic approach but keep the rigor of the integral equation.

Using (12.6.5) with $g_2 = g_1 = g$, we obtain

$$r = \frac{[g^2 - 1]^{1/2} - (g - 1)}{2(g - 1)} \tag{12.6.5}$$

to obtain

$$\mathbf{M} = \frac{r + 1}{r} = \frac{\sqrt{g^2 - 1} + g - 1}{\sqrt{g^2 - 1} - g - 1} \tag{12.8.4}$$

Hence

$$\frac{\mathbf{M} + 1}{\mathbf{M} - 1} = \frac{2(\sqrt{g^2 - 1})}{2(g - 1)} = \sqrt{\frac{g + 1}{g - 1}}$$

or

$$g = \frac{1}{2} \cdot \frac{\mathbf{M}^2 + 1}{\mathbf{M}} \tag{12.8.5}$$

Assume that the fields can be factored into a product $f(x)g(y)$ and likewise for the reflectivity $\Gamma(r) = \Gamma_x(x)\Gamma_y(y)$. Equation (12.8.2) becomes

$$\begin{aligned}[\sigma_x f(x_2)] \cdot [\sigma_y g(y_2)] &= e^{-j(kd - \pi/2)} \\ &\times \left\{ \left(\frac{1}{\lambda d} \right)^{1/2} \int_{M_{1x}} \Gamma_x(x_1) \cdot [f(x_1) e^{-j(k/2f)x_1^2}] \cdot e^{-j(k/2d)(x_2 - x_1)^2} \, dx_1 \right\} \\ &\times \left\{ \left(\frac{1}{\lambda d} \right)^{1/2} \int_{M_{1y}} \Gamma_y(y_1) \cdot [f(y_1) e^{-j(k/2f)y_1^2}] \cdot e^{-j(k/2d)(y_2 - y_1)^2} \, dy_1 \right\}\end{aligned} \tag{12.8.6}$$

Now there is no reason to keep both (x, y) in our integral equation since all manipulations on x apply to y. We also anticipate that the fields will resemble those predicted by the geometric optics approach of Sec. 12.6, and thus we express $f(x)$ in terms of a new function $V(x)$ by

$$f(x) = V(x) \exp\left[-j \frac{k(x^2 + y^2)}{(r + 1)d} \right] = V(x) \{\text{WCF}\}$$

where wavefront curvature factor (WCF) is

$$\exp\left[-j \frac{k(x^2 + y^2)}{2(r + 1)d} \right] = \exp\left[-j \left\{ \frac{k}{2d} \left(\frac{\mathbf{M} - 1}{\mathbf{M}} \right) (x^2 + y^2) \right\} \right] \tag{12.8.7}$$

The wavefront curvature factor expresses the geometric optic idea of a limited extent spherical wave originating at the point P of Fig. 12.16, and (12.8.4) has been used to express $(r + 1) = \mathbf{M}/(\mathbf{M} - 1)$. Discard the factor $\exp -j(kd - \pi/2)$, as being an obvious term to be reinserted at the end, and the x part of (12.8.6) becomes:

$$[\sigma_x V(x_2)] \cdot \exp -j\left\{\frac{k}{2d}\left(\frac{\mathbf{M}-1}{\mathbf{M}}\right)x_2^2\right\} = \left\{\left(\frac{1}{\lambda d}\right)^{1/2}\int_{\mathbf{M}_{1x}} \Gamma_x(x_1)\cdot V(x_1) \times \exp\left[-j\frac{k(\mathbf{M}-1)}{2d\cdot\mathbf{M}}x_1^2\right]\cdot \exp -j\left[\frac{kx_1^2}{2f}\right]\cdot \exp -j\frac{k(x_2-x_1)^2}{2d}\,dx_1\right\} \tag{12.8.8}$$

We would like to get rid of the wavefront curvature term by having it as a common factor appearing on both sides of (12.8.8). Hence we rewrite the sum of the terms in the exponent of (12.8.8) with a common factor of $k/(2d)$:

$$\text{sum} = -x_2^2\left(\frac{\mathbf{M}-1}{\mathbf{M}}\right) + x_2^2\left(\frac{\mathbf{M}-1}{\mathbf{M}}\right) + [x_2^2 - 2x_1x_2 + x_1^2] + x_1^2\left(\frac{\mathbf{M}-1}{\mathbf{M}}\right) - \frac{d}{f}x_1^2 \tag{12.8.9a}$$

Now use the fact that $g = 1 - d/2f$ or $d/f = 2(1-g) = -(\mathbf{M} - 2 + 1/\mathbf{M})$ to obtain

$$\text{sum} = x_2^2\left(\frac{\mathbf{M}-1}{\mathbf{M}}\right) + \mathbf{M}\left[x_1 - \frac{x_2}{\mathbf{M}}\right]^2 \tag{12.8.9b}$$

Hence, (12.8.8) becomes (after canceling the common factor on both sides of the equation)

$$\sigma_x V(x_2) = \frac{1}{(\lambda d)^{1/2}}\int \Gamma(x_1)\cdot V(x_1)\exp\left\{-jk\mathbf{M}\left[x_1 - \frac{x_2}{\mathbf{M}}\right]^2\right\}dx_1 \tag{12.8.10}$$

We can make the formula more compact by introducing a parameter a, which expresses the "size" of a mirror. It may be the physical radius of the mirror as it is drawn in Fig. 12.32(a) or it may be a taper parameter.

For instance, if the mirror were a simple one of uniform reflectivity, then

$$\Gamma(x, y) = \Gamma_0 \qquad |x| \text{ or } |y| < a \qquad \Gamma(x, y) = 0 \text{ otherwise} \tag{12.8.11a}$$

For a tapered mirror of the next section

$$\Gamma(x, y) = \Gamma(x)\Gamma(y) = \left\{\Gamma_0^{1/2}\exp\left[-\frac{x^2}{2a^2}\right]\right\}\cdot\left\{\Gamma_0^{1/2}\exp\left[-\frac{y^2}{2a^2}\right]\right\} \text{ for all } x, y \tag{12.8.11b}$$

We need not be specific other than to recognize that there is some characteristic radial size a involved. Equation (12.8.10) becomes

$$\sigma_x V(x_2) = e^{-j(kd-\pi/2)} \left(\frac{a^2}{\lambda d}\right)^{1/2} \int_{M_1} \Gamma(x_1) \cdot V(x_1) \times \exp\left[-j\frac{ka^2\mathbf{M}}{2d}\left[\frac{x_1}{a} - \frac{x_2}{\mathbf{M}a}\right]^2\right] d(x_1/a) \tag{12.8.12}$$

Introducing the conventional Fresnel number $N = a^2/\lambda d$ of Sec. 12.7, we have

$$\boxed{\sigma_x V(x_2) = e^{-j(kd-\pi/2)}\sqrt{N} \int_{M_1} \Gamma(x_1) \cdot V(x_1) \times \exp\left[-j\pi N\mathbf{M}\left[\frac{x_1}{a} - \frac{x_2}{\mathbf{M}a}\right]^2\right] d(x_1/a)} \tag{12.8.13}$$

This equation can be made even more compact by a change of variable (with an accompanying change in limits of integration)

Let

$$u = \sqrt{\pi N\mathbf{M}}\left[\frac{x_1}{a} - \frac{x_2}{\mathbf{M}a}\right] \qquad du = \sqrt{\pi N\mathbf{M}}\, d(x_1/a) \tag{12.8.14}$$

$$\therefore \qquad \sigma_x V(x_2) = \frac{e^{-j(kd-\pi/2)}}{\sqrt{\mathbf{M}}} \left\{\frac{1}{\sqrt{\pi}} \int_{M_1} \Gamma(x_1) V(x_1) e^{-ju^2} du\right\} \tag{12.8.15}$$

Finally, our integral has reached a manageable size for evaluation—possibly by a computer—which is the only general means for further progress. There are some trends that can be observed by considering two special cases with one requiring a judicious approximation. For the remainder of this section, we restrict our attention to cavities with a uniform reflectivity $\Gamma(r) = \Gamma_0$ for $r \leq a$, $\Gamma = 0$ otherwise.

The form of (12.8.13) suggests that the main contribution to the integral comes from the regions near $x_1 = x_2/\mathbf{M}$, because for intervals of x_1 much different than that value, the exponential term oscillates violently through all complex values of 1. This observation suggests a Taylor series expansion of $V(x_1)$ (around $x_1 = x_2/\mathbf{M}$ or $u = 0$) is in order.

$$V(x_1) = V(x_2/\mathbf{M}) + V'(x_2/\mathbf{M})[x_1 - (x_2/\mathbf{M})] + \frac{V''(x_2/\mathbf{M})}{2!}[x_1 - (x_2/\mathbf{M})]^2 + \cdots$$

and then we ignore all terms except the first, recognize that $V(x_2/\mathbf{M})$ does not depend upon the variable of integration x_1, discard the phase factor as being ignorable for the moment, and write (12.8.15) in the following form:

$$\sigma_x V(x_2) = \left(\frac{\Gamma_0}{\mathbf{M}}\right)^{1/2} V\left(\frac{x_2}{\mathbf{M}}\right) \left\{\frac{1}{\sqrt{\pi}} \int_{M_1} e^{-ju^2} du\right\} \tag{12.8.16}$$

If the Fresnel number N is allowed to be infinite, then the magnitude of the quantity in the braces is 1, and we have the simple approximate solution

$$\boxed{\sigma_x V(x_2) = \left(\frac{\Gamma_0}{\mathbf{M}}\right)^{1/2} V\left(\frac{x_2}{\mathbf{M}}\right)} \tag{12.8.17}$$

This is precisely the result obtained from the geometric imaging analysis. For instance, suppose we choose a simple polynomial power expansion for $V(x) = x^n$. Then, $V(x_2/\mathbf{M}) = (x_2/\mathbf{M})^n$, and thus a solution for σ_x is:

$$\sigma_x x_2^n = \frac{\sqrt{\Gamma_0}}{\mathbf{M}^{n+1/2}} x_2^n \qquad \text{or} \qquad \sigma_x = \frac{\sqrt{\Gamma_0}}{\mathbf{M}^{n+1/2}}$$

Now the factor σ_x describes the survival of the field as a function of the x coordinate for a single pass between the mirrors, and if we include the variation in y by the equation y^m, then we have the survival factor for the (n, m) mode with a field variation of the form $x^n y^m$.*

$$\boxed{s = \sigma_x \sigma_y = \frac{\Gamma_0}{\mathbf{M}^{m+n+1}}} \tag{12.8.18}$$

If $m = n = 0$, the denominator is $\mathbf{M}$ and we obtain precisely the result inferred by the geometric imaging approach. In addition, we also see that the fraction of the field surviving for the higher order modes (with $m + n > 0$) is considerably smaller, which indicates a high degree of mode discrimination. This high mode discrimination is one of the main features of the unstable resonators.

The ploy of letting the Fresnel number $N \to \infty$ is on awkward grounds, however, since then there would be no opening in the cavity for the radiation to escape. If we keep a finite mirror size, but still use the Taylor series, then we obtain

$$\begin{aligned}\sigma_x V(x_2) &= \frac{\sqrt{\Gamma_0}}{\sqrt{\mathbf{M}}} V\left(\frac{x_2}{\mathbf{M}}\right) \cdot \left\{\frac{1}{\sqrt{\pi}} \int_{u_2}^{u_1} e^{-ju^2}\, du\right\} \\ &= \frac{\sqrt{\Gamma_0}}{\sqrt{\mathbf{M}}} V\left(\frac{x_2}{\mathbf{M}}\right) \cdot \frac{1}{\sqrt{2}} \left\{[C(u_1) - C(u_2)] - j[S(u_1) - S(u_2)]\right\} \end{aligned} \tag{12.8.19}$$

where $$\left(\frac{2}{\pi}\right)^{1/2} \int_0^x e^{-ju^2}\, du = C(x) - jS(x)$$

The integral is related to the Fresnel cosine and sine integrals where the values u_1 and u_2 corresponding to the appropriate limits on x_1 by (12.8.14). We will not go further with this other than to indicate that this same integral arises in the case of diffraction around sharp ("knife") edges and yields fields (or fringes) in the shadow of that edge. It is these fringes that are responsible for the undulations of the loss around the geometrically predicted value

*Much of the book has used S as the survival factor for the power. Hence, we will use s as the survival for the field.

in Fig. 12.17 that occur (approximately) at an "equivalent Fresnel number" given by

$$N_{\text{eq}} = \frac{1}{2}\left(\mathbf{M} - \frac{1}{\mathbf{M}}\right)N \tag{12.8.20}$$

If the reflection coefficient is an arbitrary function of the radius, we must retreat to (12.8.15) and use a computer. A simple case that can be handled with elementary functions is that of a mirror with a Gaussian tapered reflection coefficient dependent upon the radial distance from axis.

That is,

$$\Gamma(r) = \Gamma_0 \exp\left[-\frac{(x^2 + y^2)}{2a^2}\right] \tag{12.8.21a}$$

and thus

$$\Gamma_x(x) = \Gamma_0^{1/2} \exp\left[-\frac{1}{2}\left(\frac{x}{a}\right)^2\right] \tag{12.8.21b}$$

Although such a mirror may be difficult to fabricate, it avoids the sharp edges of the mirror that produce the fringes in the shadow zone.

With this functional form of the reflection coefficient, all limits of integration on all integrals go to $\pm\infty$, and (12.8.13) becomes

$$\sigma_x V(x_2) = \frac{\sqrt{\Gamma_0}}{\sqrt{\mathbf{M}}}\left\{\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty} \exp\left[-\left(\frac{x_1}{a}\right)^2\right]\cdot V(x_1) \times \exp\left[-j\pi N\mathbf{M}\left[\frac{x_1}{a} - \frac{x_2}{\mathbf{M}a}\right]^2\right]\right\} d\left[\sqrt{\pi N\mathbf{M}}\,\frac{x_1}{a}\right] \tag{12.8.22}$$

There is a hint of a Fourier transform appearing in this equation, and this suggests that we pick a Gaussian for $V(x)$; that is,

assume

$$V(x) = \exp\left[-\left(\frac{x}{w}\right)^2\right] \tag{12.8.23}$$

where w is an unknown spot size to be determined by the requirement of fields on the mirrors being scaled replicas of each other. Insert this guess into (12.8.22), complete the square in the exponent inside the integral, extract terms involving x_2 from the integral, and find that the debris left can be integrated with an elementary result. There are so many pitfalls and factors in such a procedure that it is easy to make a mistake. Let us see if we cannot use a simpler technique and one that provides additional insight.

If the guess expressed by (12.8.23) is correct, then we merely have a Gaussian beam propagating through an infinite sequence of lenses. But if this guess is correct (and it is and can be verified with a bit of work), then we already know how to handle the problem without the necessity of the integral equation: Use the $ABCD$ law of Sec. 4.5 and Chapter 5 to express the propagation of the beam and demand self consistency. Such an approach is much less taxing on our abilities to solve integral equations and is much easier from a conceptual standpoint. The significant difference encountered here is that the equivalent

lenses are not uniformly lossy as was assumed in the earlier chapters and so our first task is to find the $ABCD$ matrix for a "tapered" lens. This can be done almost by inspection and the application of this approach is the subject of the next section.

12.9 *ABCD* LAW FOR "TAPERED MIRROR" CAVITIES

A general expression for a Gaussian beam *incident* (at $z = 0^-$) on a mirror is [cf., (3.3.2)]

$$E(0^-) = E_0 \exp\left[-j\frac{k(x^2+y^2)}{2q_1}\right] \tag{3.3.2}$$

where any z dependent phase factors are absorbed in E_0 and q_1 is the complex beam parameter. The *field* transmission factor of the equivalent tapered *thin* lens is thus given by

$$t = \Gamma_0 \exp\left[-\left(\frac{x^2+y^2}{2a^2}\right)\right] \exp\left[+j\frac{k(x^2+y^2)}{2f}\right] \tag{12.9.1}$$

where Γ_0 is the field reflection coefficient at $x = y = 0$. Hence, the field transmitted through an equivalent tapered lens (or reflected from the mirror) is

$$E(0^+) = \Gamma_0 E_0 \exp\left[-j\frac{k(x^2+y^2)}{2q_1}\right] \exp\left[-\frac{x^2+y^2}{2a^2}\right] \exp\left[+j\frac{k(x^2+y^2)}{2f}\right] \tag{12.9.2}$$

Now the $ABCD$ law would indicate that the output (i.e., the field reflected from the mirror) would be a Gaussian with a new complex beam parameter q_2.

$$E(0^+) = \Gamma_0 E_0 \exp\left[-j\frac{k(x^2+y^2)}{2q_2}\right] \tag{12.9.3}$$

Since we are dealing with Gaussian beams, the complex beam parameter must transform according to the $ABCD$ law:

$$\frac{1}{q_2} = \frac{C + D(1/q_1)}{A + B(1/q_1)} \tag{3.6.3}$$

The term $B = 0$ because a "thin" lens (or mirror) has no longitudinal distance, and thus (3.6.3) becomes

$$\frac{1}{q_2} = \frac{C}{A} + \frac{D}{A} \cdot \frac{1}{q_1} \tag{12.9.4}$$

Now $A = D = 1$ because the element has a plane of symmetry midway between $z = 0^+$ and $z = 0^-$. Equate the fields expressed by (12.9.2) and (12.9.3) to obtain the value for C:

$$-j\frac{k}{2q_2} = -j\frac{k}{2q_1} - \frac{1}{2a^2} + j\frac{k}{2f} \qquad k = \frac{2\pi}{\lambda}$$

Since $$\frac{1}{q_2} \triangleq \frac{1}{R_2} - j\frac{\lambda}{\pi w_2^2} = C + \frac{1}{q_1} = \frac{1}{R_1} - \frac{1}{f} - j\frac{\lambda}{\pi}\left(\frac{1}{w_1^2} + \frac{1}{2a^2}\right)$$

Thus,
$$\boxed{C = -\frac{1}{f} - j\frac{\lambda}{2\pi a^2}} \tag{12.9.5}$$

If $a \to \infty$, then we have a uniform mirror or lens, and we recover the elementary result used in Chapters 3 to 6.

A word of caution. We would *never* obtain this result by "ray tracing"; it is applicable to Gaussian beams alone, as is the $ABCD$ law.

The physical interpretation of (12.9.5) is that the curved tapered mirror changes the phase front of the incident Gaussian beam, per usual, and it also changes the spot size because of the nonuniform transmission. It transmits the field at large radii and reflects that at the smaller values. The reflected Gaussian has a new spot size given by

$$\boxed{\frac{1}{w_2^2} = \frac{1}{w_1^2} + \frac{1}{2a^2}} \tag{12.9.6}$$

Having the $ABCD$ matrix means that all of the prior work of Chapters 3 to 6 is directly applicable to both stable *and* unstable cavities with the proviso that at least one mirror has a taper. Unfortunately, the presence of a complex number in the $ABCD$ matrix complicates matters enormously and precludes obtaining a general expression for an arbitrary geometry with a general taper specification. However, a particular case illustrates the procedure to be followed and identifies the pitfalls to be avoided.

Example

Consider the geometry shown in Fig. 12.33 with output coupler M_2 having a negative radius of curvature ($-|R_2|$ with the quantity $|R_2|$ being a positive dimension) with R_1 concave (and also a positive dimension) with an infinitely large aperture. Let us

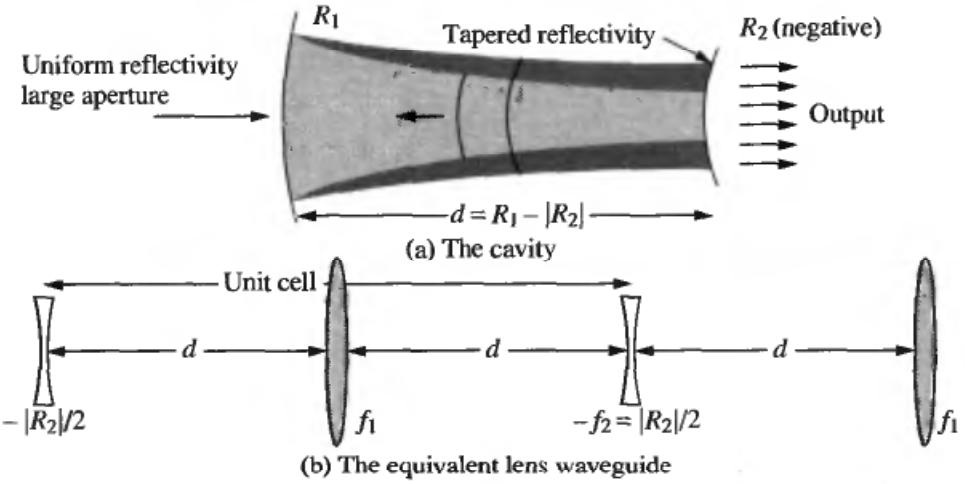


FIGURE 12.33. A cavity with a tapered mirror.

choose the particular geometry such that $g_1 g_2 = 1$ and thus $d = R_1 - |R_2|$, which is on the border line of stability with neither the analysis of Chapters 3 to 6 nor that of Sec. 12.2 being applicable. The geometric imaging approach would suggest that *all* of the field survives ($S = 1$)—see (12.6.9)—whereas the minimum spot size of the Gaussian beam at the beam waist given by (5.2.8) would predict $w_0^2 = 0$. (Be careful: Substitute $-|R_2|$ in that equation and use $d = R_1 - |R_2|$). If $w_0^2 = 0$—where ever it occurs—the beam would expand infinitely fast, and an insignificant fraction would be intercepted by any finite size mirror and thus the survival factor would be zero. Clearly, the answer lies in between the two approaches.

As a result of the assumption of $g_1 g_2 = 1$, the magnification parameters have a particular simple form

$$\mathbf{M}_1 = \frac{|R_2|}{R_1} < 1$$

and

$$\mathbf{M}_2 = \frac{R_1}{|R_2|} > 1$$

and thus $\mathbf{M}_1 \mathbf{M}_2 = 1$. To save a bit on notation, set $\mathbf{M} = \mathbf{M}_2 = R_1/|R_2|$ and thus $\mathbf{M}_1 = 1/\mathbf{M}_2 = 1/\mathbf{M}$, which expresses (rather loosely, as we shall see) the idea that the beam expands by the ratio $R_1/|R_2|$ in going from 2 to 1 and contracts by this same ratio going the other way as is sketched in the diagram.

The ray matrix for the unit cell chosen in Fig. 12.33(b) is

$$T = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\dfrac{1}{f_1} & 1 \end{bmatrix} \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \dfrac{1}{f_2} - j\dfrac{X}{d} & 1 \end{bmatrix} \tag{12.9.7a}$$

where $X = \lambda_0 d/(2\pi a^2) = 1/(2\pi N)$, a convenient abbreviation

$$T = \begin{bmatrix} \left(1 - \dfrac{d}{f_1}\right) + \left(2 - \dfrac{d}{f_1}\right)\left(\dfrac{d}{f_2} - jX\right) & d\left(2 - \dfrac{d}{f_1}\right) \\ -\dfrac{1}{f_1} + \left(1 - \dfrac{d}{f_1}\right)\left(\dfrac{1}{f_2} - j\dfrac{X}{d}\right) & \left(1 - \dfrac{d}{f_1}\right) \end{bmatrix} \tag{12.9.7b}$$

which is a general description of the unit cell specified and is not dependent upon the restriction of $g_1 g_2 = 1$. For the case $g_1 g_2 = 1$, use $d = R_1 - |R_2|$, and express all ratios of $d/f_{1,2}$ in terms of the magnification $\mathbf{M} = |R_1/R_2|$ to obtain what follows:

$$T = \begin{bmatrix} \dfrac{(3\mathbf{M} - 2)}{\mathbf{M}} - j\dfrac{2X}{\mathbf{M}} & \dfrac{2d}{\mathbf{M}} \\ -\dfrac{1}{d}\left[\dfrac{2(\mathbf{M} - 1)^2}{\mathbf{M}} - j\dfrac{\mathbf{M} - 2}{\mathbf{M}} X\right] & \dfrac{2 - \mathbf{M}}{\mathbf{M}} \end{bmatrix} \tag{12.9.8}$$

[The serious student will check the above and verify that $AD - BC = 1$ for the determinant of (12.9.8).]

Now the complex beam parameter q is determined from the $ABCD$ law (3.6.1) by forcing q to repeat after a round trip in exactly the same fashion as was done in Chapter 5 with the same general result.

$$\frac{1}{q} \triangleq \frac{1}{R_{\text{inc}}} - j\frac{\lambda}{\pi w_{\text{inc}}^2} = \frac{1}{2B}\left\{-(A-D) \pm [(A+D)^2 - 4]^{1/2}\right\} \tag{5.3.5}$$

where R_{inc} and w_{inc} are the beam parameters of the Gaussian at the defining plane of the chosen unit cell, which for Fig. 12.33 is at the tapered mirror 2.

It is not as easy as it was in Chapter 5 to identify the "parts" of the complex beam parameter, the radius of curvature, and the spot size because both $(A - D)$ and the radical can generate imaginary numbers. For our particular case, we have

$$\frac{1}{q} = -\frac{(\mathbf{M}-1)}{d} + j\frac{X}{2d} \pm \frac{\mathbf{M}}{2d}\left[-j\frac{2X}{\mathbf{M}} - \frac{X^2}{M^2}\right]^{1/2} \tag{12.9.9}$$

This equation illustrates a general issue that must always be addressed on a case-by-case basis.

> The sign of the square root in (12.9.9) is always chosen such that $1/q$ has a negative imaginary part so as to describe a beam varying as $\exp[-(r/w)^2]$.

For the specific problem considered here, the complex number inside the square root is in the third quadrant, and thus the two square roots are in the second and fourth quadrants of the complex plane. If the second quadrant value is chosen, we must use the negative sign, and use the positive value for the fourth quadrant choice, with either leading to the same answer. If we limit our attention to cavities with reasonably large Fresnel numbers $N > 1$, then $X = 1/2\pi N$ is small, $(X/\mathbf{M})^2$ is extremely small, which suggests neglecting the $(X/\mathbf{M})^2$ term in (12.9.9) to obtain

$$\frac{1}{q} \approx -\frac{(\mathbf{M}-1) - \sqrt{\mathbf{M}X}/2}{d} - j\frac{\sqrt{\mathbf{M}X}/2 - X/2}{d} \tag{12.9.10}$$

Thus

$$R_{\text{inc}} \approx -\frac{d}{(\mathbf{M}-1) - \sqrt{\mathbf{M}X}/2} \tag{12.9.11a}$$

and

$$\frac{\pi w_{\text{inc}}^2}{\lambda} \approx \frac{d}{\sqrt{\mathbf{M}X}/2 - X/2} \tag{12.9.11b}$$

Let us take a set of typical parameters:

> Let $d = R_1 - |R_2| = 100$ cm and $\lambda_0 = 10.6$ μm (CO_2 wavelength); $R_1 = 10$m, with a very large aperture, uniform reflectivity of $\Gamma_1^2 = 0.98$; $R_2 = -9$ m, with a Gaussian tapered field reflection coefficient $\Gamma_2(r) = \Gamma_0 e^{-(r^2/2a^2)}$; the geometric magnification $\mathbf{M} = R_1/R_2 =$

$10/9 = 1.11$; and choose $X = 1/45$, an arbitrary but a reasonable choice. Thus the Fresnel number $N = a^2/\lambda_0 d = 7.162$ and the tapering distance $a = 0.8713$ cm. To put this into a common perspective, consider a 1.5 inch diameter substrate with a central reflectivity of Γ_0^2. At $r = 3/8$ inch, the reflectivity is only 29.6% of the peak and down to less than 1% at the edge.

The spot size and radius of curvature of the beam incident on M_2 (from the left) is found from (12.9.11) to be $w_{\text{inc}} = 0.7072$ cm and $R_{\text{inc}} = -30.73$ m, which indicates a slightly converging wavefront. The amplitude of the field of the incident beam (at the surface of M_2) is

$$E_{\text{inc}} = E_0 \exp[-(r/w_{\text{inc}})^2] \tag{12.9.12a}$$

The amplitude of the field reflected from M_2 is given by

$$|E_{\text{ref}}| = \Gamma_0 E_0 \exp\left[-r^2\left(\frac{1}{w_{\text{inc}}^2} + \frac{1}{2a^2}\right)\right] \triangleq \Gamma_0 E_0 \exp[-(r/w_{\text{ref}})^2] \tag{12.9.12b}$$

and thus the spot size of the beam reflected is given by

$$\frac{1}{w_{\text{ref}}^2} = \frac{1}{w_{\text{inc}}^2} + \frac{1}{2a^2} \qquad a = 0.8713 \text{ cm} \qquad \therefore w_{\text{ref}} = 0.6134 \text{ cm} \tag{12.9.12c}$$

The Gaussian beam reflected from M_2 has its curvature changed to

$$\frac{1}{R_{\text{ref}}} = \frac{1}{R_{\text{inc}}} - \frac{1}{f_2} = \frac{1}{(-30.73 \text{ m})} - \frac{1}{(-4.5 \text{ m})} \qquad \text{or} \qquad R_{\text{ref}} = +527.21 \text{ cm} \tag{12.9.12d}$$

which indicates an expanding beam returning to M_1.

A significant parameter for a laser is the fraction of the incident power (not intensity) reflected by mirror 2 and is given by

$$\Gamma_2^2(\text{eff}) = \frac{\text{power reflected}}{\text{power incident}} = \Gamma_0^2 \frac{\iint e^{-2(r/w_{\text{ref}})^2} r\, dr}{\iint e^{-2(r/w_{\text{inc}})^2} r\, dr} = \Gamma_2^2 \left(\frac{w_{\text{ref}}}{w_{\text{inc}}}\right)^2 \tag{12.9.13a}$$

yielding $\Gamma_2^2(\text{eff}) = 0.7522\Gamma_0^2$ (12.9.13b)

for the numerics chosen here.

The output intensity as a function of r is the difference between the incident and the reflected values:

$$I_{\text{out}}(r) = \frac{E_{\text{inc}}^2}{2\eta_0}\left\{\exp[-2(r/w_{\text{inc}})^2] - (\Gamma_0^2 \exp[-(r/a)^2]) \exp[-2(r/w_{\text{inc}})^2]\right\}$$

or

$$I_{\text{out}}(r) = \frac{E_{\text{inc}}^2}{2\eta_0}\left\{\exp[-2(r/w_{\text{inc}})^2] - \Gamma_0^2 \exp[-2(r/w_{\text{ref}})^2]\right\} \tag{12.9.14}$$

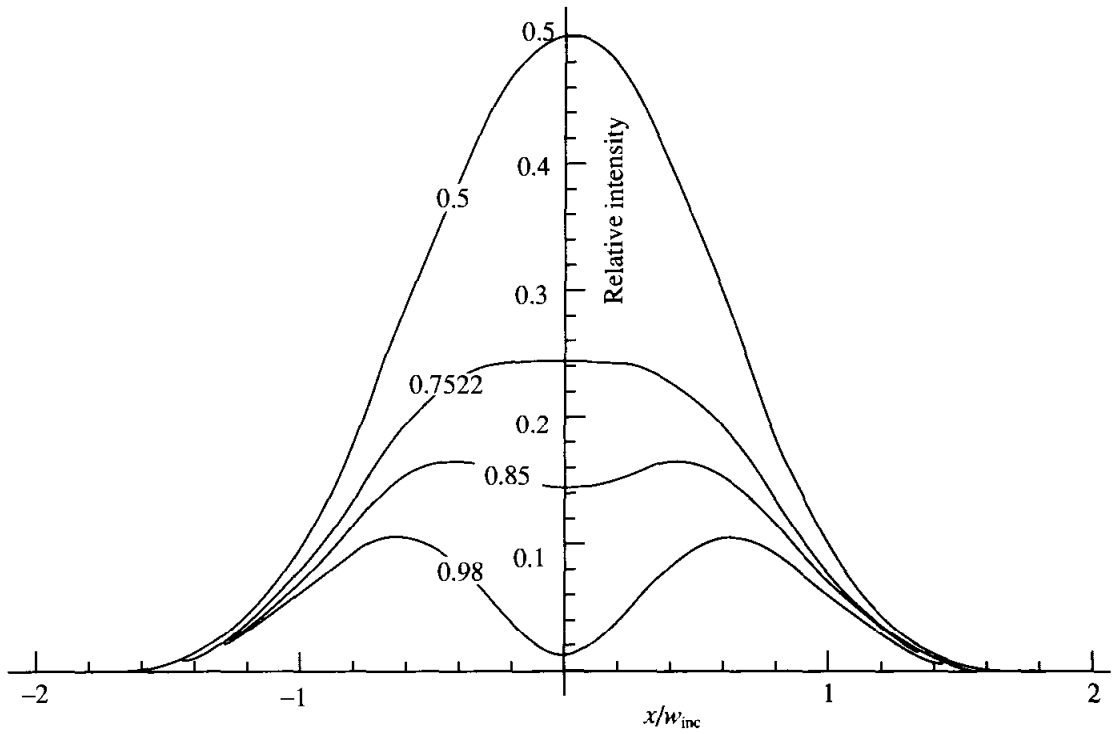


FIGURE 12.34. The spatial distribution of the output for different central reflectivities of 0.98, 0.85, 0.7522, and 0.5 assuming constant power incident from $\mathbf{M}_1$.

A sketch of the output intensity as a function of x/w_{inc} is shown in Fig. 12.34. If the peak reflectivity at the center Γ_0^2 is high, then there is a central dip and the output resembles a partially filled donut. If Γ_0^2 is low, the output resembles the incident Gaussian with a spot size w_{inc}. It is left for a problem to show that if

$$\Gamma_0^2 = \left[\frac{w_{ref}}{w_{inc}}\right]^2 \tag{12.9.15}$$

then the output beam is maximally flat at $r = 0$ [i.e, $(d^2I/dr^2) = 0$] and thus the output does not have a central dip which is a desirable feature. For such a mirror, the effective reflectivity of M_2 is $\Gamma_2^2(\text{eff}) = (w_{ref}/w_{inc})^4 = 0.5658$. If the reflectivity of the other mirror M_1 is high—say, 98%—then the threshold for the laser is readily attained for this flat top beam.

$$\gamma_{th} = \frac{1}{2l_g} ln\left\{1/[\Gamma_1^2\Gamma_0^2(\text{eff})\right\} = 0.295\%/\text{cm} \tag{12.8.16}$$

for $l_g = 100$ cm.

This is a high quality beam originating from an effective volume encompassing a significant number of active atoms.

It is left for a problem to compute the spot size and radius of curvature of the beam incident on $\mathbf{M}_1$ with the answer being $w_1 = 0.7316$ cm and $R_{\leftarrow} = 609.07$ cm (found by using the Gaussian beam expansion formulas). This means that the spot size has expanded by a ratio of 1.193 in passage from mirror 2 back to 1 and is slightly greater than the

geometric magnification ratio of 10/9. There is not any reason to expect that **M** should be equal to the first spot-size ratio but it is a good guess as to the beam shape.

This analysis and thus all of the formulas beyond (12.9.6) are specific to the case of $g_1 g_2 = 1$, but the procedure is similar for any combination of mirrors. The important issue brought out by the example is that a high quality beam can be generated by a large volume of atoms.

12.10 LASER ARRAYS

12.10.1 System Considerations

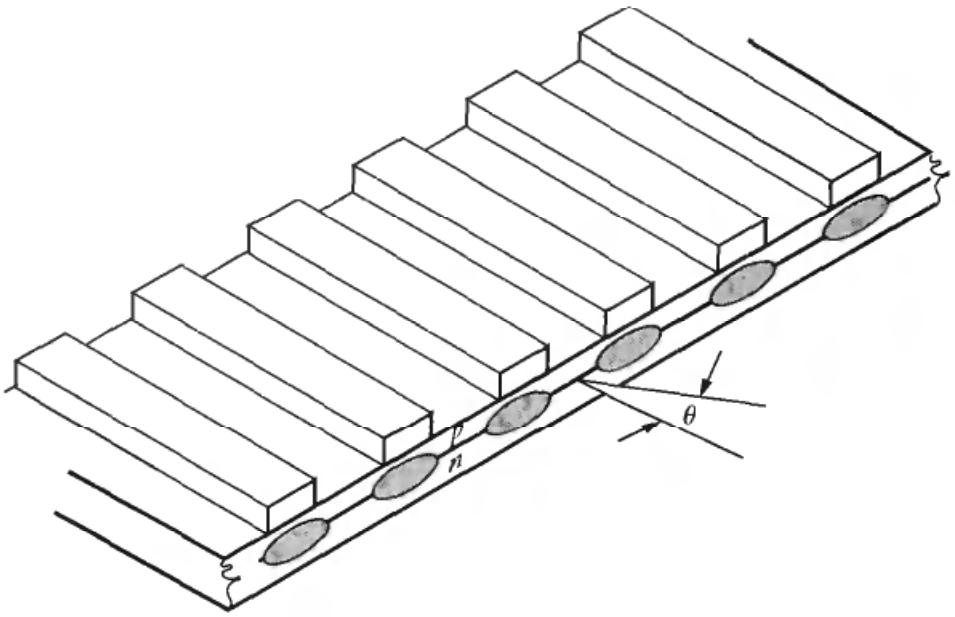
We cannot increase the excitation to any one given laser cavity to arbitrarily high values so as to generate high optical power and expect it to survive. There are many laser physics issues that negate this approach and practical considerations that forbid it. For instance, if we attempt to load a CO_2 with too much power, the "waste heat" raises the rotational temperature, which in turn lowers the gain. In semiconductor lasers, the internal optical electric field may become so large that it literally self-destructs because of optical breakdown at any minor imperfection at the output facet. (This presumes, of course, that the device did not melt because of its waste heat.)

One approach to obtaining high power is to utilize the coherence properties of lasers to "add" the fields generated by many parallel sources, each of which could be working at its optimum operating point. A similar situation is often used in the microwave domain with a master oscillator driving many (N) amplifiers with each of those exciting an element in an array antenna. Ideally, we can obtain N times the power radiated by a single element but into the pattern of the much larger array, which implies a much narrower radiated beam. As a bonus, the beam pattern can be electronically scanned by controlling the relative phase between the elements. There has been a major effort to adapt this older microwave technology to the optical domain with the majority of the effort being directed toward semiconductor arrays, but some work has also been devoted to coupling the larger CO_2 or chemical lasers also.

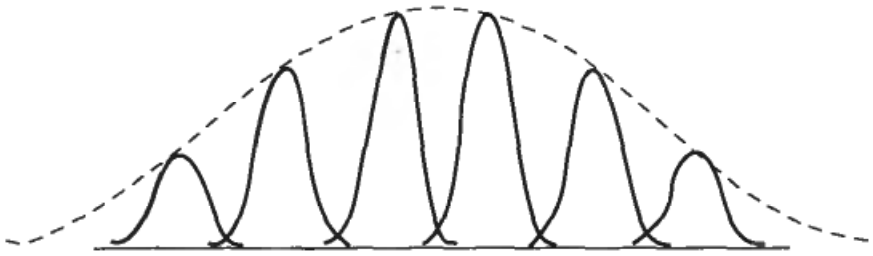
12.10.2 Semiconductor Laser Array: Physical Picture

The problem to be addressed (slowly) is shown in Fig. 12.35(a) and differs from the microwave scheme in an important issue: There are N oscillators, not amplifiers, which are coupled and radiate at the output aperture. Each laser oscillates at the same frequency (or wavelength) and is locked in a specific phase relationship with respect to each other, which need not be zero or even the same between adjacent stripes. The theoretical questions to be addressed are, "What are the allowed *phases* of the coupled oscillators, and, if more than one solution is possible, then which is the most desirable?"

It is as important to realize that the question is real and makes a significant difference to us in the outside world as it is to answer it with pages of mathematics. Suppose that each stripe produces a nice smooth mode profile with a central amplitude of each laser under

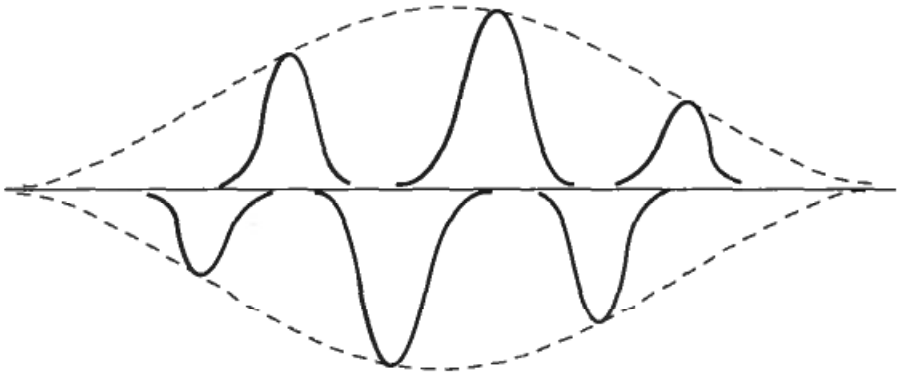


(a)



0°- phase shift between elements

(b)



180° - phase shift between elements

(c)

FIGURE 12.35. (a) A typical six-stripe semiconductor array. (b) An in-phase field distribution. (c) The same field elements as (b) but with a 180° phase shift between adjacent stripes.

another smooth envelope as shown in Fig. 12.35(b). To compute the far field or radiation pattern of this array, the observable, we need the specification of the relative phase of each element in addition to the amplitude specification. If, for instance, all oscillators had a common phase (which can be assigned to be zero), then the radiation from this array would be a maximum along the axis because each differential source at $\pm x$ would contribute an equal-in-phase radiated field there. If, however, the phase of each adjacent oscillator differed by 180°, the field along the central axis would be 0 because of the phase cancellation of contributions from each symmetrically placed but out-of-phase element. If we avoid the mathematics for a moment, we would guess that the far-field radiation pattern of the in-phase distribution would be a single lobe with an angle determined (roughly) by the bigger envelope, whereas the 180° distribution would be a two-lobe pattern with a zero on the axis. Detailed calculation using antenna theory will verify that guess. Obviously, we would prefer the in-phase distribution, but alas, a vote is not taken on the matter: The laser physics *decides* the issues. Unless special techniques are used, the laser prefers the out-of-phase distribution because the total field is minimized in the regions between the stripes, which are absorptive.*

Inasmuch as there are $N = 6$ elements in the example shown in Fig. 12.35, we also suspect that there are $N = 6$ different combinations of phase and amplitude assignments among the elements. These are the so-called "supermodes," the coherent combination of the individual modes of each strip. The next section describes those combinations.

12.10.3 Supermodes of the Array

The laser array must oscillate at a common frequency (wavelength) so that a coherent addition of the fields is possible. Hence, each laser must couple to and lock with a particular phase relationship to its adjacent neighbor. Because of this coupling, the propagation constant of the modes in the array may differ somewhat from that of an isolated element, but not by much. Further, we expect the transverse field configuration of the individual element to be hardly changed from that of normal mode for an isolated stripe as found in Sec. 12.2, which is expressed in symbolic form by

$$\mathbf{E}_l = E_l(x, y)e^{-j\beta_0 z}\hat{\mathbf{a}}_t \tag{12.10.1}$$

where the fact that there is gain in the stripe is ignored in the interest of simplicity and β_0 is the propagation constant for an isolated element. We presume that (12.10.1) satisfies the *uncoupled wave equation* in all of its glory and gory detail with the appropriate boundary conditions. With nearest neighbor coupling, the amplitude of the field for the l^{th} element obeys

$$\frac{dE_l}{dz} = -j\beta_0 E_l - j\kappa E_{l+1} - j\kappa E_{l-1} \tag{12.10.2}$$

with l running from 1 to N and κ indicating how much of the neighbor is injected into the l^{th} cavity. The functional dependence of the fields on the transverse coordinates has also

*There is no current in the gaps. Therefore, the semiconductor is not inverted, and thus there is absorption.

been suppressed so as to concentrate on the amplitudes of each stripe. Our task is to find a nontrivial solution to the *entire* array, which can be expressed in matrix form by

$$\frac{d}{dz}\begin{bmatrix} E_1 \\ E_2 \\ . \\ . \\ . \\ . \\ E_N \end{bmatrix} = -j\begin{bmatrix} \beta_0 & \kappa & 0 & . & . & . & 0 \\ \kappa & \beta_0 & \kappa & 0 & . & . & 0 \\ 0 & \kappa & \beta_0 & \kappa & . & . & 0 \\ . & . & \kappa & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & . & . & . & \kappa & \beta_0 & \kappa \\ 0 & . & . & . & 0 & \kappa & \beta_0 \end{bmatrix} \cdot \begin{bmatrix} E_1 \\ E_2 \\ . \\ . \\ . \\ . \\ E_N \end{bmatrix} \tag{12.10.3a}$$

or

$$\frac{d}{dz}\bar{\bar{\mathbf{E}}} = -j\bar{\bar{\mathbf{M}}} \cdot \bar{\bar{\mathbf{E}}} \tag{12.10.3b}$$

where $\bar{\bar{\mathbf{E}}}$ is a column matrix and $\mathbf{M}$ is a tridiagonal square array. If we assume that the fields can be expressed as

$$E_l^{(k)} = A_l^{(k)} \exp[-j\beta_k z] \tag{12.10.4}$$

then (12.10.2) becomes a simple eigenvalue problem.

$$-j\begin{bmatrix} \Delta\beta & \kappa & 0 & . & . & . & 0 \\ \kappa & \Delta\beta & \kappa & . & . & . & 0 \\ 0 & \kappa & \Delta\beta & \kappa & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & . & . & . & \kappa & \Delta\beta & \kappa \\ 0 & . & . & . & . & \kappa & \Delta\beta \end{bmatrix} \cdot \begin{bmatrix} A_1^{(k)} \\ A_2^{(k)} \\ . \\ . \\ . \\ . \\ A_N^{(k)} \end{bmatrix} = 0 \tag{12.10.5}$$

with $\Delta\beta = \beta_0 - \beta_k$. For a nontrivial solution (i.e., $A_l^{(k)} \neq 0$), the determinant must vanish, and this procedure guarantees N solutions for the propagation constants in the array with each solution corresponding to a different amplitude distribution among the elements. We can solve those equations without recourse to the beauties and subtleties of matrix theory by recognizing the close analogy between difference and differential equations.

Consider one of the equations of (12.10.5):

From the l^{th} row:

$$(\beta_0 - \beta_k)A_l^{(k)} + \kappa(A_{l-1}^{(k)} + A_{l+1}^{(k)}) = 0 \tag{12.10.6}$$

We guess that

$$A_l^{(k)} = A_0^{(k)} \sin l\theta$$

where θ is a parameter to be determined. Expand the coefficients A_{l-1} and A_{l+1} by the standard trigonometric formulas, add, and multiply by κ as dictated by (12.10.6):

$$A_{l-1}^{(k)} = A_0^{(k)}[\sin l\theta \cos\theta - \cos l\theta \sin\theta]$$

$$A_{l+1}^{(k)} = A_0^{(k)}[\sin l\theta \cos\theta + \cos l\theta \sin\theta]$$

$$\kappa(A_{l+1}^{(k)} + A_{l-1}^{(k)}) = 2\kappa A_0^{(k)} \sin l\theta \cos\theta$$

Insert these expansions of the guess into (12.10.6) and cancel common factors of $A_0^{(k)} \sin l\theta$.

$$\begin{aligned}(\beta_0 - \beta_k)A_0^{(k)} \sin l\theta &= -2\kappa A_0^{(k)} \sin l\theta \cos\theta \\ (\beta_0 - \beta_k) &= -2\kappa \cos\theta\end{aligned} \tag{12.10.7}$$

This equation should be interpreted as determining the propagation constants β_k in terms of the parameter θ, which is not known but which can be found from a bit of physics: The array is the same whether viewed from the front or back and thus the N^{th} element must look like the first if you go to the other side of the page.

$$A_N^{(k)} = A_0^{(k)} \sin N\theta = \pm A_1^{(k)} = \pm A_0^{(k)} \sin\theta$$

$$N\theta = k\pi - \theta \ .$$

or

$$\theta = \frac{k\pi}{N+1} \tag{12.10.8}$$

The pertinent data for the various combinations of fields of the supermodes are now available:

$$A_l^{(k)} = A_0^{(k)} \sin l\left\{\frac{k\pi}{N+1}\right\} \tag{12.10.9}$$

$$\beta_k = \beta_0 + 2\kappa \cos\left[\frac{k\pi}{N+1}\right] \tag{12.10.10}$$

It is easy to show that there are only N values of $k = 1, 2, \cdots N$ that yield different field configurations, each with a slightly different propagation constant. The laser oscillates at a frequency such that the round trip phase shift, $\beta_k 2d = q2\pi$, for each supermode configuration. If β_0 is approximated by $\omega n/c$, then the equation for the frequency of the k supermode is

$$\nu_k = \frac{c}{\lambda_0}\left[1 - \frac{\kappa\lambda_0}{\pi n}\cos\left(\frac{k\pi}{N+1}\right)\right] \tag{12.10.11}$$

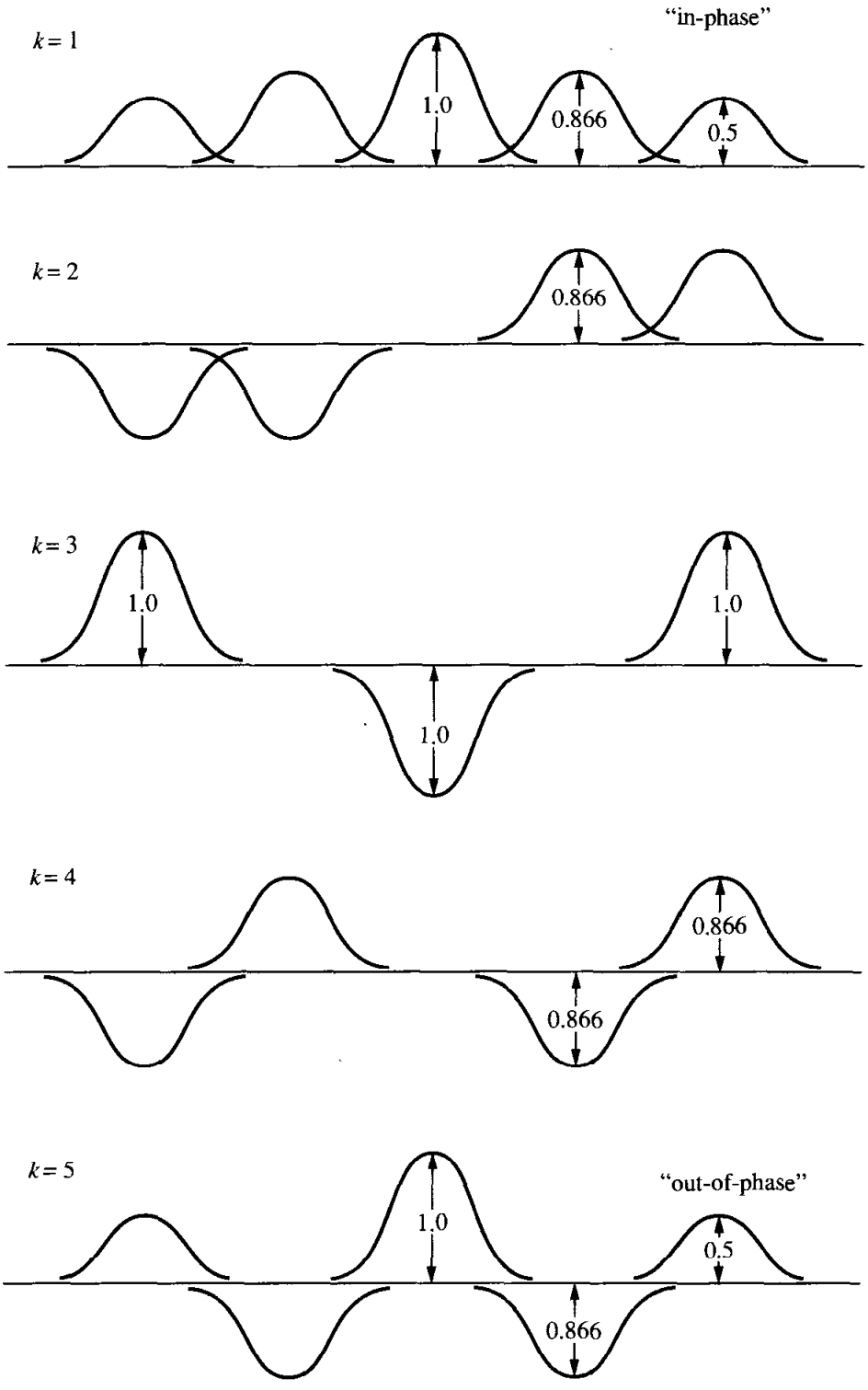


FIGURE 12.36. The supermode field distribution for $N = 5$.

where the axial mode number q is approximated by $2nd/\lambda_0$, and λ_0 is the free-space wavelength from each isolated element. For $\kappa\lambda_0/2\pi n$ small, the supermodes generate a range of wavelengths centered around λ_0.

$$\frac{\Delta\lambda}{\lambda_0} = \frac{\lambda_{k=N} - \lambda_{k=1}}{\lambda_0} = \frac{\kappa\lambda_0}{\pi n}\left[\cos\left(\frac{\pi}{N+1}\right) - \cos\left(\frac{N\pi}{N+1}\right)\right] \tag{12.10.12}$$

Fig. 12.36 on page 573 illustrates the "bottom line" of these calculations: the field distribution for an $N = 5$ array.

As mentioned earlier, the physics of the laser controls which of these supermodes will be chosen, and, unfortunately, the 180° phase shift between stripes $k = 5$ is usually the answer because it has the highest gain-to-loss ratio since the total field is a minimum in the regions between the stripes. We can coax the laser into the other supermodes by external feedback or careful control of the current injection. It should be clear, however, that the far-field pattern of each of these supermodes is considerably different.

12.10.4 Radiation Pattern

The field radiated by this array can be found from the application of (12.7.10) of this chapter:

$$\mathbf{E}(r_0) = \frac{j}{\lambda_0 d}\int_{x_1}\int_{y_1} \mathbf{E}(x_1, y_1)e^{-j\phi(2,1)}\,dx_1\,dy_1 \tag{12.7.10}$$

where the limits of integration extend over the entire aperture of the array, and $\phi(2, 1)$ is the phase shift from the points on the facet to the observation point. It is not much fun to make these calculations, and the only saving grace comes from making the far-field approximation. (Fortunately, the semiconductor laser, even an array, is so small that almost everything is in the far field.) Then the far-field pattern is just the Fourier transform of the aperture distribution, and all of the history of it can be brought to bear on this problem. From an operational standpoint, we merely collect the radiation with a lens with its focal plane at that of the array. The output of the lens is the desired Fourier transform.

PROBLEMS

12.1. Read the article by A.E. Siegman and R. Arrathoon, "Modes in Unstable Optical Resonators and Lens Waveguide," *IEEE J. Quantum Elect. QE-3*, No. 4, 156-163, April 1967. Answer the following questions by simple equations, a sketch, or a paragraph.

(a) What is the parameter M and how is it related to the geometry of Fig. 12.23.

(b) Show how (12.7.10) leads to Equation (4) of the Siegman and Arrathoon paper. Why are the subscripts on g no longer necessary?

(c) Why is "the region centered around a point $y - x/M$" so important in Equation (5)? What is a "Fresnel integral"?

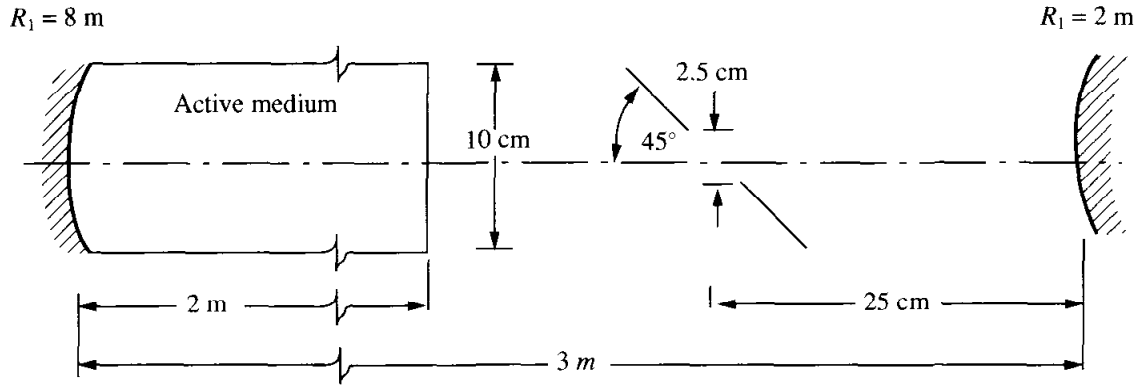
(d) Show the "obvious" development of Equation (8) from Equation (7).

(e) What is the meaning of the Fresnel number N?

(f) Verify the phase-shift curve plotted for $N = 2.7$ in Fig. 11 using the geometrical analysis.

(g) If a cavity were specified by $L = 50$ cm, $g = 1.10$, $N_{eq} = 2.5$, what is the size of the mirrors and the radius of curvature? Compare the loss predicted by the geometric analysis to that plotted in Fig. 11.

12.2. Consider the unstable resonator shown in the accompanying diagram. The reflection coefficient of R_1 is 0.98, that of R_2 is 0.95.



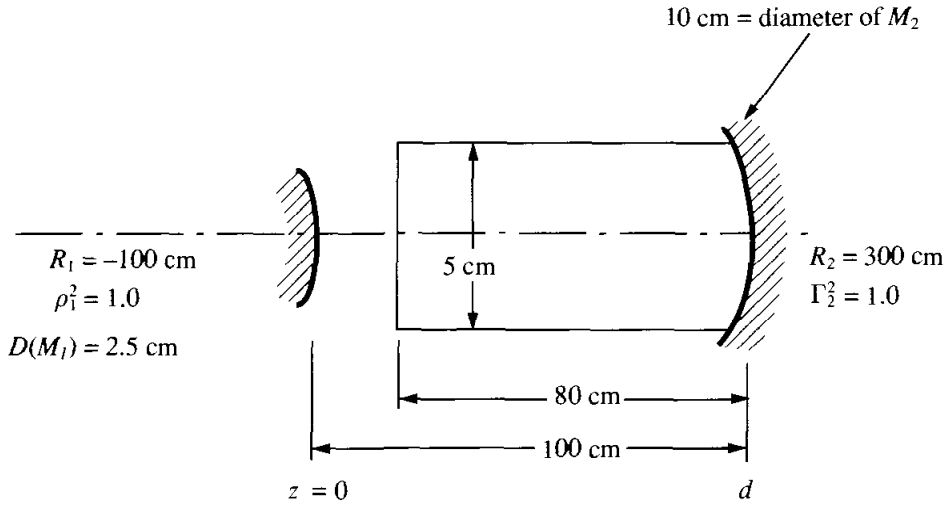
(a) Find the location of points P_1 and P_2 and shown them on a sketch of the optical cavity.

(b) Use Siegman's simple approach to sketch the field variation within the optical cavity. Find and label all pertinent dimensions. Show the wavefront coming from the beam slicer.

(c) What is the mean loss per pass through the empty cavity?

(d) What is the gain required of this laser?

12.3. Consider a laser with a confocal unstable resonator as shown on the accompanying diagram. Sketch the amplitude of the field at $z = 0^-$ and $z = d$, and describe why this field pattern fits the definition of a mode. What should be the gain coefficient of the active cell to obtain oscillation?



12.4. The starting point for the analysis of a finite aperture mirror system is (12.7.10). If we apply this expression to nonconfocal geometry with $d \neq R$ and express the field as a product in the form $E = u(x) \cdot v(y)$, then the integral equation becomes

$$\sigma u(x_2) = \int_{-a}^{a} K(x_1, x_2) u(x_1)\, dx_1$$

where

$$K(x_1, x_2) = \sqrt{\frac{j}{\lambda d}} \cdot \exp\left\{-j \frac{k}{2d}[g(x_1^2 + x_2^2) - 2x_1 x_2]\right\}$$

and $g = 1 - d/R$ (same as in Chapter 2), $k = 2\pi/\lambda$, d is mirror spacing, R is radius of curvature of the two mirrors, and σ contains the attenuation and phase shift experienced by the wave in propagating from M_1 to M_2. This is (74) in the paper by H. Kogelnik and T. Li, "Laser Beams and Resonators," *Appl. Opt. 5*, 1550-1567, 1967. Show the essential steps in arriving at this answer from (12.7.10).

12.5. The mirrors of the confocal cavity shown on the diagram below have a tapered field reflection coefficient given by

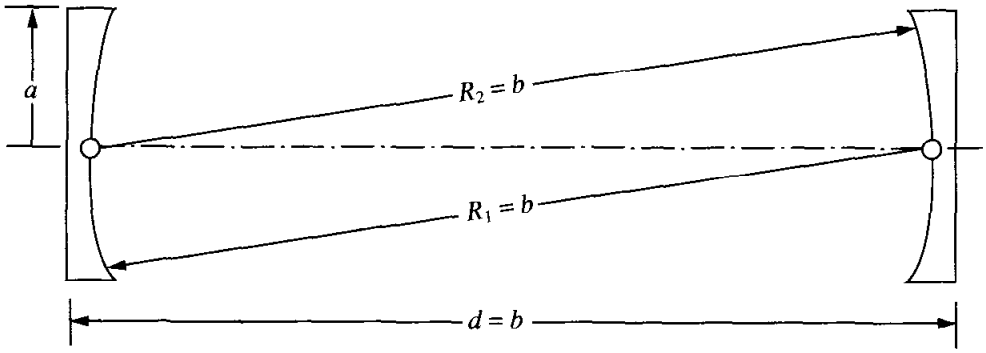
$$\Gamma(r) = \Gamma_0 \exp[-(tr)^2]$$

where r = distance from the axis

Γ_0 = field reflection coefficient at $r = 0$ (use 0.98 for a numerical value)

t = a taper rate that can be expressed as a multiple of the normal spot size on the mirrors; $t = m/w_s$

w_s = spot size on $M_{1,2}$ (for $t = 0$) and is equal to $(2b/k)^{1/2}$



(a) Use the integral equation formulation of the problem to find a self-consistent set of fields reflected from each mirror. (Assume $a \to \infty$.) The functional form of the solution is a Gaussian with a spot size w: Find the ratio $(w/w_s)^2$.

(b) Find the single pass gain G for the laser to oscillate.

(c) Plot the relative output *intensity* (i.e., $\mathbf{E}^2$) at M_2 as a function of the normalized distance (r/w_s) assuming the transmission is $1 - \Gamma^2(r)$. Label the position of peak intensity.

(d) Use the above results and the expansion laws for a Gaussian mode to sketch the variation of the fields *inside* the cavity. Identify the planes where the wave front of the positive and negative going waves are planar and the spot sizes are a minimum. Evaluate w_0^2 at those planes.

12.6. In the example considered for gain guiding, the following parameters were chosen: $\lambda_0 = 8400$ Å, $n = 3.6$, $\gamma_0 = 100\ \text{cm}^{-1}$, $L_x = 10\ \mu\text{m}$, and $L_y = 5\ \mu\text{m}$, where the length parameters refer to the parabolic gain and index profiles. The spot sizes and the radius of curvature of the mode had the following values: $w_x = 0.862\ \mu\text{m}$, $w_y = 5.22\ \mu\text{m}$, $R_y = 3.66.9\ \mu\text{m}$, and $\gamma_{\text{net}} = 72.75\ \text{cm}^{-1}$ for the field parameters *inside* the semiconductor cavity shown below.

(a) If the cavity length was 500 μm, the facet reflectivity $R_{1,2} = 0.36$, and the current density was 1000 A/cm^2, what is the gain coefficient?

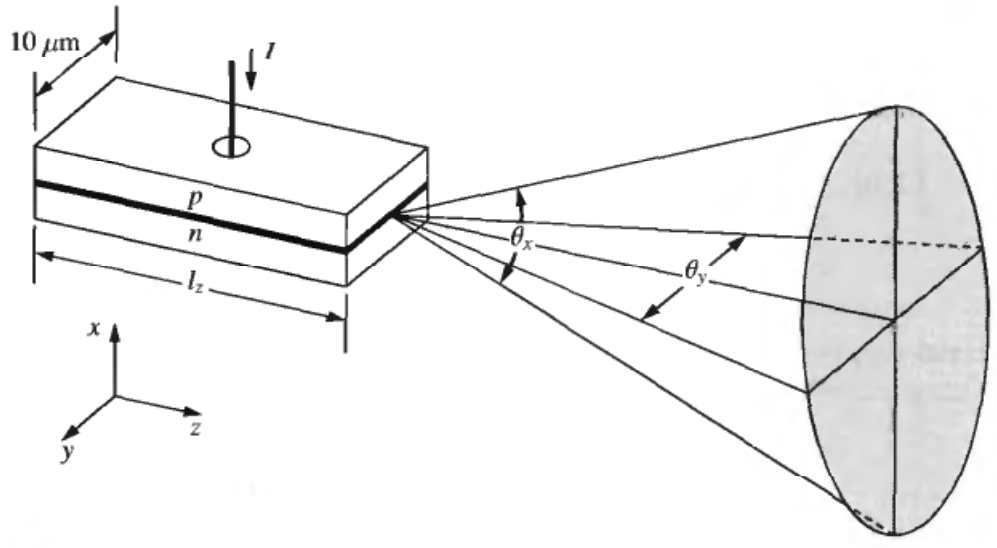
(b) Assume that the gain coefficient scales directly with current, and compute the threshold current.

(c) The parameters given above describe the field inside the laser. The observable is the external field. Use the $ABCD$ law for Gaussian beams to compute what follows:

(1) The beam spot sizes and the radius of curvature on the air side of the facet.

(2) Find the far-field divergence angle of the beam as it spreads in the x and y directions. Assume $\theta/2 = dw/dz$ (in radians).

(3) If the beam were *not* astigmatic, $R_y = \infty$, but with the same spot sizes, what would be the angle θ_y?

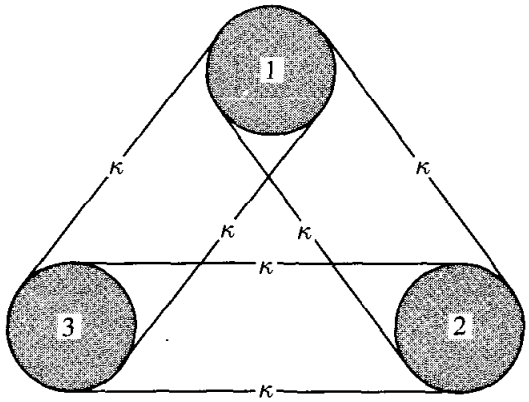


12.7. Read the article by D.D. Cook and F.R. Nash, "Gain Induced Guiding and Astigmatic Output Beam of GaAs Laser," *J. Appl. Phys. 46*, 1660-1672, 1975. The following questions refer to this paper.

(a) Show the development of Equations (2.2) and (2.3) from (2.1).

(b) Derive Equations (2.27) and (A7) using the $ABCD$ law rather than by using the procedure followed.
(c) Explain (with words and a few equations) the logic of Equation (2.11).
(d) Which equation is the *integral* definition of the modal gain, and why does it make sense in terms of stimulated emission and absorption?
(e) What is the origin of Equation (C1)?
(f) Explain the notation and development of Equation (B5).

12.8. Consider the coupled triangular array of lasers shown in the diagram below. Assume that the propagation constant of an isolated channel is β_0 and there is a coupling coefficient of κ between each element.
(a) Write the system of equations describing the propagation of a supermode of the array.
(b) Find the characteristic values for all of the allowed values of the propagation constant for the supermodes.
(c) Indicate the relative amplitudes of the fields for these supermodes.

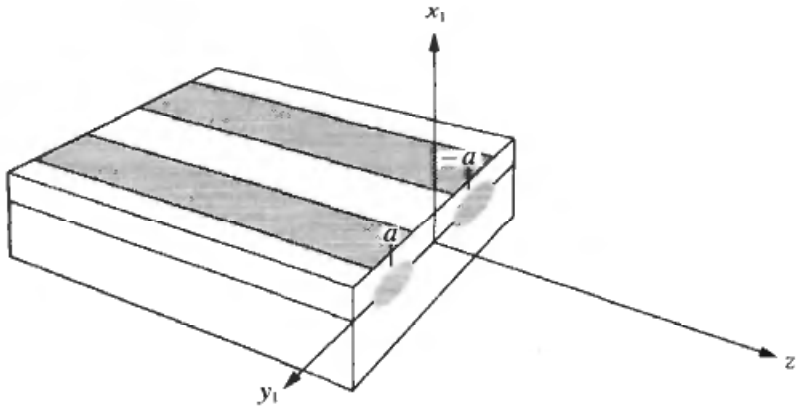


12.9. Find the propagation constants and the relative field amplitudes for the $N = 4$ linear array.

12.10. Consider a semiconductor "array" of two coupled lasers that produce a near field (or facet) field distribution described by an elliptical Gaussian beam associated with each stripe.

$$E_{1,2}(x_1, y_1, z = 0) = A_{1,2} \exp[-(x_1/w_x)^2] \exp[-(y_1 \pm a)/w_y]^2$$

where $w_x \ll w_y$.
(a) Find and sketch the total fields of the two supermodes of this array assuming $a = w_y$.
(b) Use the Fresnel integral with $d = z_2$ to compute the total field of the supermodes in the far-field plane. The following steps are offered to help in the formulation of the mathematics.
(1) Expand the distance between (x_2, y_2, z_2) and $(x, y, z = 0)$ in a Taylor series.
(2) If $k(x_1^2 + y_1^2)/2z \ll \pi/2$ (which defines the far field), we can neglect this term. Do so, but explain why this is permissible.



(3) The remaining debris can be interpreted as a wave apparently originating at $z = 0$ with a spherical phase front times an array integral. Identify those terms. What is the radius of curvature of the phase front?

(4) Evaluate the array integral by using standard (nonnumerical) mathematical techniques.

(c) If only one stripe were oscillating, then the envelope of the far-field intensity pattern would also be an elliptical Gaussian beam. Sketch the e^{-2} contour of this supermode beam pattern showing the contrast between it and the facet field.

(d) Plot the relative intensity (as a function of y_2) of the two supermodes in the far-field zone, and compare this sketch to the intensity at the facet. Define characteristic spot size w_{x2}, w_{y2} appropriate for this far field and relate them to those at the facet.

(e) Identify characteristic distances along y_2 where the field vanishes.

(f) Use the answers to (e) to evaluate the angular spread of the two supermodes.

12.11. The phenomenon of gain guiding was first discovered and analyzed in the large solid state or gas lasers. Due to nonuniform pumping, the gain coefficient is maximum on axis and is assumed to decrease in a parabolic fashion as

$$1 - \left(\frac{r}{L_2}\right)^2$$

where L_2 is the characteristic radial distance where the gain crosses over to absorption. An additional complication arises from the fact that the local pumping heats the medium, which usually makes the dielectric constant increase with radius according to

$$1 + \left(\frac{r}{L_1}\right)^2$$

an effect sometimes referred to as index *antiguiding*. The problem is to analyze a situation in which both effects appear simultaneously. Start with the wave equation

$$\nabla^2 E + \frac{\omega^2}{c^2}\left(\epsilon'[1 + \left(\frac{r}{L_1}\right)^2] + j\epsilon''[1 - \left(\frac{r}{L_2}\right)^2]\right) E = 0$$

(a) Assume

$$E = \Psi(r, z) \exp\left[\frac{\gamma_0}{2} - j\beta\right] z$$

and derive an equation for Ψ by using the paraxial wave approximation.

(b) Assume $\Psi = \exp -j[P(z) + \beta r^2/2q(z)]$, and find two ordinary differential equations for $P(z)$ and $q(z)$.

(c) Assume $q'(z) = 0$ and solve for q and $P(z)$. Express your answer in terms of

$$\frac{1}{a^2} = \left\{\left[\frac{1}{L_1^2}\right]^2 + \left[\frac{\gamma_0}{\beta L_2^2}\right]^2\right\}^{1/2}$$

and

$$\theta = \tan^{-1}\left\{\frac{\gamma_0}{\beta}\left[\frac{L_1}{L_2}\right]^2\right\}$$

(d) Find an expression (in terms of a and θ) for the spot size, the radius of curvature of the phase front, and the net or modal gain coefficient and phase constant.

(e) Evaluate these expressions for $\gamma_0 = 0.1$ cm^{-1}, $\lambda_0 = 1.06$ μm, $n = 1.73$, $L_1 = 100$ cm, and $L_2 = 0.1$ cm, parameters that might be found in a small YAG laser pumped axially with another laser. Make sure you compute those values that emerge from the rod *into air*. This means that you will have to apply the $ABCD$ law for the index-air interface.

(f) Make a careful sketch of the variation of the real index $((\epsilon')^{1/2})$, the gain coefficient, and the electric field as a function of r. Indicate the change in the real index between $r = 0$ and $r = 1$ mm, and compare the electric field at $r = 1$ mm to that at $r = 0$.

(g) If this mode emerged from a rod at $z = l_g$, then we would hope to reflect it back into the medium and re-establish that same mode propagating back to the left. Show that a perfect mode match is *not* possible with a simple curved mirror as a reflector. Does this mean that the laser will not oscillate if a simple mirror configuration is used? What part of the above development must be changed to accommodate the curved mirror (theory be darned). Do not resolve the problem; just indicate what has to be done.

12.12. Suppose three stripe semiconductor lasers are close enough to be coupled and thus locked in a supermode of the array and hence there are three different amplitudes of each laser strip.

(a) Derive the eigenvalue matrix for this case.

(b) What are the relative peak amplitudes of three solutions, and sketch the relative amplitude distribution as a function of the x coordinate. Assume that each individual stripe produces a Gaussian with the same spot size w_{0y}.

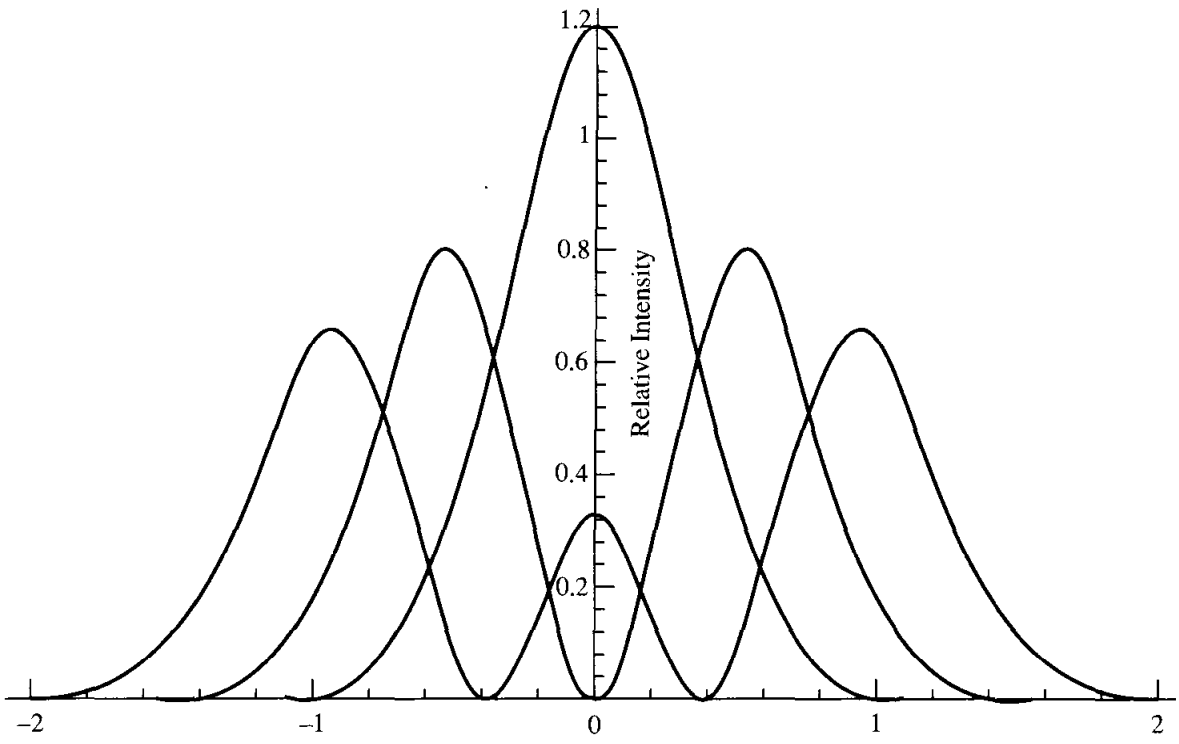
(c) What is the difference in wavelength between the various supermodes.

(d) What follows assumes that each stripe oscillated in a Gaussian field distribution around its central location at the output facet where the wavefront is

planar.

$$E_{1,2,3}(x, y) = A_{(1,2,3)}^{(k)} \exp\left[\frac{-(y - y_i)^2}{w_{0y}^2}\right] \exp -\frac{x^2}{w_{0x}^2}$$

with $y_1 = -a$, $y_2 = 0$, $y_3 = +a$, and $w_{0y} = a$. The following three graphs show the *far-field* intensity distribution for each supermode as a function of y_2 normalized to the spot size from an individual element in the *far field*. (There is nothing to be learned from the x direction, so it is ignored.) The vertical axis has been scaled so that each mode radiates equal total powers. Identify which far-field pattern is associated with which solution [found in (b)] and explain the logic for your choice. Let $w_{0y} = 25$ μm and $\lambda_0 = 0.85$ μm. Use the graphical data to estimate the far-field spreading angle for the various supermodes. Which one is most desirable, but which is most likely to be obtained?



12.13. Some semiconductor gain guided lasers produce a curved wavefront with an internal field incident on the facet which can be approximated by

$$E^+ = E_0 \exp[-x^2/w_x^2] \exp[-y^2/w_y^2] \exp\left[-j\frac{ky^2}{2R_y}\right] \exp[-j\beta z]$$

where $w_x \ll w_y$ and R is independent of z within the semiconductor cavity. Assume $n = 3.6$.

(a) Prove that the spot sizes on the air side of the facet are equal to those on the inside.

(b) The radius of curvature of the emerging beam (in air) is *not* equal to internal value at the facet. What is the relationship?

(c) Notice that the phase front is planar with respect to x in the example of Sec. 12.3 but curved with respect to y. Thus the expansion law (in air) for w_x uses the output facet plane for $z = 0$ (i.e., $R = \infty$, and w_x = a minimum). The corresponding plane for the expansion in the y direction is behind the output facet. Derive a formula for this astigmatic distance.

(d) What are the formulas for far-field expansion angles θ_x and θ_y in terms of w_x, w_y, and R_y at the output facet?

12.14. Consider a *uniform* medium $(-L/2 < z < L/2)$ with a field amplification factor g_0 and a phase constant $\beta_0 (g_0 \ll \beta_0)$ with a reflector Γ (real) placed at $z = L/2$. Assume a uniform plane wave incident from the left.

(a) Find an expression for the total E and H fields for $-L/2 < z < L/2$ in terms of the incident wave E_0.

(b) Find an expression for the real part of the Poynting vector $\mathbf{E} \times \mathbf{H}^*$.

(c) Evaluate the power reflection and transmission coefficients; that is, $R = P_{\text{ref}}(z = -L/2)/P_{\text{inc}}(z = -L/2)$ and $T = P_t(z = +L/2)/P_{\text{inc}}(z = -L/2)$.

(d) For what value of Γ will the power reflection coefficient exceed the transmission coefficient?

12.15. The transmission through a passive Bragg reflector is very small for $\kappa L \geq 3$ and $\delta = 0$ (see Fig. 12.10). If $\delta > \kappa$, then p becomes complex and the transmission changes rapidly with δ.

(a) Show that (12.5.15) and (12.5.16) lead to

$$|S_{11}|^2 = R = \frac{(\kappa L)^2 \sin^2 y}{y^2 \cos^2 y + [(\kappa L)^2 + y^2] \sin^2 y}$$

and

$$|S_{12}|^2 = T = \frac{y^2}{y^2 \cos^2 y + [(\kappa L)^2 + y^2] \sin^2 y}$$

where $pL = jy$.

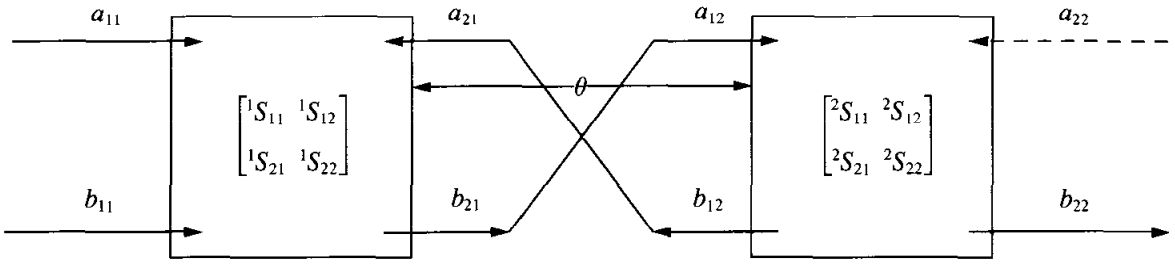
(b) What are the values of δL for which $|S_{12}|^2 = 1$?

(c) Find the transcendental equation that determines the value of the detuning at which $|S_{12}|^2 = 0.5$ (the 3 dB points).

(d) Show that $\delta \cdot L/\pi = (\nu - \nu_B)/(c/2nL)$, a convenient normalized frequency scale.

(e) Determine the FWHM value of $\delta L/\pi$ for which $|S_{11}|^2 > 0.5$ and also the bandpass. Assume $\kappa L = 3$ and $Ga_xIn_{1-x}P_yAs_{1-y}$ with $n = 2.27$.

12.16. Consider the geometry illustrated below with two Bragg mirrors separated by an optical phase shift $\theta = \omega n d/c$. Use the scattering matrix formalism of Appendix I to derive the net transmission and reflection coefficients for the two-mirror system.



[NOTE: The first subscript on (a, b) refers to the "side" of the mirror (left=1, right=2), and the second refers to the mirror. The prior superscript on S refers to the mirror.]

12.17. Consider the case discussed in the text in Sec. 12.9 and shown in Fig. 12.33.

(a) Show that if $\Gamma_0^2 = (w_r/w_i)^2$, there is no "dip" in the output beam.

(b) The parameters of the Gaussian beam *incident* on M_2 were given in the example in Sec. 12.9. Use the Gaussian beam expansion (3.3.10) and (3.3.11) to find the plane $z = 0$ for the reflected beam. What is the radius of curvature and spot size of the beam at M_1?

(c) Compare w_1/w_r to the magnification factor $\mathbf{M} = 1.111$ assumed for the example discussed in the text.

(d) Sketch the variation of the spot size and wavefront curvature for the field going from $1 \rightarrow 2$ and also for that going from $2 \rightarrow 1$.

(e) Repeat the case covered in Sec. 12.9 for $2\pi N = 22.5$, a choice that makes the radius of curvature at M_2 equal to ∞.

12.18. This problem is a repeat of the case examined in the text in which we use the integral equation to predict the field on mirror M_2 in terms of the assumed field on mirror M_1. The cavity is confocal (i.e., $d = b = R_1/2 = R_2/2$) with square mirror of dimension $2a$ as shown in Fig. 12.29. Rather than choosing a Gaussian, we pick a [cosine]^2 variation for the field; that is, let

$$E(x_1) = \cos^2 \frac{\pi x_1}{2w_s} \qquad \text{for} \quad x_1 < w_s$$

$$E(x_1) = 0 \qquad \text{for} \quad x_1 > w_s$$

with a similar variation for y. This resembles a beam. Assume that the dimensions of $M_{1,2}$ are such that $a/w_s = 2$ and that the Fresnel number of the cavity is $\pi/2$. Define

$$U_1 = \frac{x_1\sqrt{C}}{a}$$

$$C = 2\pi N, \; N = \text{Fresnel number} = \pi/2$$

$$\therefore E(U_1) = \cos^2 U_1 = \frac{1}{2}(1 + \cos 2U_1)$$

(a) Find an analytic expression for the field incident on M_2 in the following format. Caution: $E(U_2)$ will *not* be a cosine^2 function, but it will "look" like one even

though the function will be different. It will have the following form:

$$E(U_2) = \left[\frac{1}{2}\,\frac{(\qquad\qquad)}{(\qquad\qquad)}\right]\left\{\left[\frac{\sin\theta_2}{\theta_2}\right]\cdot\frac{\pi^2}{(\qquad\qquad)}\right\}$$

where $\theta_2 = \dfrac{\pi U_2}{2}$ $\qquad$ $U_2 = \dfrac{\pi}{2}\cdot\dfrac{x_2}{w_s}$

(b) Make a careful graph of the variation of the intensity as a function of $x_{1,2}/a$. What is the relative amplitude of the field at the edge of M_2 (i.e., $x_2 = a$)?

(c) Part (b) will indicate that the spatial extent of the field on M_2 is much larger than that on M_1. Thus, we now change our specifications to $N = \pi/2$ as before, but let $a/w_s = 2\text{m}$. Find a new expression for the field on M_2 similar to that found in (a).

$$E(U_2) = \left[\frac{1}{2}\,\frac{(\qquad\qquad)}{(\qquad\qquad)}\right]\left\{\left[\frac{\sin\theta_2}{\theta_2}\right]\cdot\frac{\pi^2}{(\qquad\qquad)}\right\}$$

where $\theta_2 = (?)$

(d) Let $m = 1/2$. Construct a careful graph of the field on M_2 and compare it to that on M_1. You should find that the spatial extent of the field on M_2 is smaller than on M_1.

(e) Choose the value of m to make the FWHM of the fields on both mirrors equal. What is the value of m? (Ans: $m = 0.627$.)

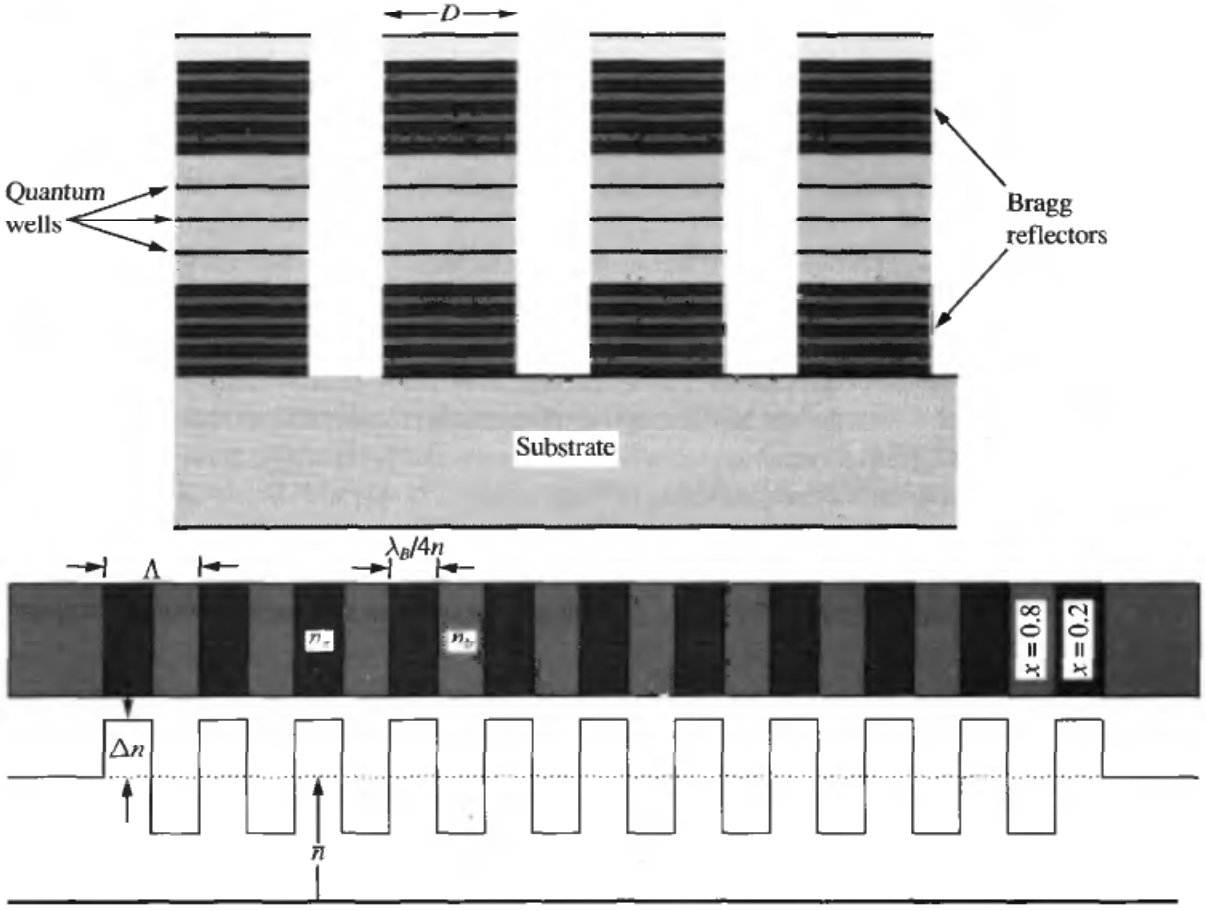
(f) For this value of m found in (e), make a careful graph of the two fields (on M_1 and M_2) as a function of $x_{1,2}/a$. (You will not be able to tell the difference between the two fields and thus have found (approximately) a mode.) What is the value of $E(x_2 = a)$?

(g) Use the above to estimate the loss per pass owing to diffraction, and compare with the numerical solution presented in Fig. 12.28. This requires you to extrapolate to $N = 1.57$.

12.19. If the index is a function of (x, y), we have the possibility of $\mathbf{E}\cdot\nabla n(x, y) \neq 0$. What changes, if any, should be made to (12.2.2) and (12.2.4a) and (12.2.4b)?

12.20. Shown below is a schematic of a vertical-cavity surface emitting semiconductor laser (VCSEL), where a thin gain medium—usually a quantum well (or two or three)—is placed at the maximum of the electric field of a standing wave between two highly reflecting mirrors. Lasing is to take place in a direction perpendicular to the wafer surface. It is imperative to maintain near-perfect crystal quality throughout the structure, and thus the mirrors must be grown onto the substrate using materials that are lattice matched but still have different indices of refraction for use as a dielectric mirror. The $Al_xGa_{1-x}As/GaAs$ system is suited for this purpose.

Find the reflection coefficient (at the Bragg wavelength) of the dielectric stack of $(N + 1/2)$ pairs of alternating layers of index $n_{a,b}$ with each being a quarter wavelength thick at the center (Bragg) wavelength imbedded in a medium of index $(n_a + n_b)/2$ using two theories:



(a) Use the standard transmission line theory for a $\lambda/4$ transformer and cascade the results.

(b) Use the theory of the distributed Bragg mirror.

For numerical purposes, assume $Al_xGa_{1-x}As$ as being the material system and use Fig. 11.12 to evaluate the indices of refraction. Let $N = 20$ pairs.

REFERENCES AND SUGGESTED READINGS

1. A.E. Siegman, "Unstable Optical Resonators for Laser Applications," *Proc. IEEE 53*, 277–287, Mar. 1965.
2. (a) R.J. Freiberg, P.P. Chenausky, and C.J. Buczek, "An Experimental Study of Unstable Confocal CO_2 Resonators," *IEEE J. Quant. Electron. QE-8*, 882–892, Dec. 1972.
   (b) R.J. Freiberg, P.P. Chenausky, and C.J. Buczek, "Asymmetric Unstable Traveling-Wave Resonators," *IEEE J. Quant. Electron. QE-10*, 279–289, Feb. 1974.
   (c) A. Siegman and R. Arrathoon, "Modes in Unstable Optical Resonators and Lens Waveguides," *IEEE J. Quant. Electron. QE-3*, 156, 1967.
3. Yu. A. Ana'ev, "Unstable Resonators and Their Applications," *Sov. J. Quant. Electron. 1*, 565–586, 1972.
4. H.A. Haus, *Waves and Fields in Optoelectronics* (Englewood Cliffs, N.J.: Prentice Hall, 1984).
5. A.G. Fox and T. Li, "Resonant Modes in a Maser Interferometer," *Bell Syst. Tech. J. 40*, 453–488, Mar. 1961.

6. G.D. Boyd and J.P. Gordon, "Confocal Multimode Resonator for Millimeter through Optical Wavelength Masers," *Bell Syst. Tech. J. 40*, 489–508, Mar. 1961.
7. G.D. Boyd and H. Kogelnik, "Generalized Confocal Resonator Theory," *Bell Syst. Tech. J. 41*, 1347–1369, July 1962.
8. H. Kogelnik, "Imaging of Optical Modes-Resonators with Internal Lenses," *Bell Syst. Tech. J. 44*, 455–494, Mar. 1963.
9. G.H.B. Thompson, *Physics of Semiconductor Laser Devices* (New York: John Wiley & Sons, 1980).
10. H.K. Kressel and J.K. Butler, *Semiconductor Lasers and Heterojunction LED* (New York: Academic Press, 1977).
11. H.C. Casey Jr. and M.P. Panish, *Heterostructure Lasers* (New York: Academic Press, 1978).
12. R.G. Hunsperger, *Integrated Optics: Theory and Technology*, Springer Series in Optical Science (New York: Springer-Verlag, 1984).
13. D.D. Cook and F.R. Nash, "Gain-induced Guiding and Astigmatic Output of GaAs Laser," *J. Appl. Phys. 46*, 1660–1672, 1975.
14. F.R. Nash, "Mode Guidance Parallel to the Junction Plane of Double-Heterostructure GaAs Lasers," *J. Appl. Phys. 44*, 4696, 4707, 1973.
15. E. Kapon, J. Katz, and A. Yariv, "Supermode Analysis of Phase-Locked Arrays of Semiconductor Lasers," *Opt. Lett.* Vol. 10, No. 4, 1984.
16. H. Kogelnik and C.V. Shank, "Stimulated Emission in a Periodic Structure," *Appl. Phys. Lett. 18*, 152–154, 1971.
17. H. Kogelnik and C.V. Shank, "Coupled Wave Theory of Distributed Feedback Lasers," *J. Appl. Phys. 43*, 2327–2335, 1975.
18. W. Streiffer, D.R. Scifres, and R.D. Burnham, "Coupled Wave Analysis of DFB and DBR Lasers," *IEEE J. Quant. Electron. QE-13*, 134–140, 1977.
19. T. Ikegami, "Reflectivity at Facet and Oscillation Mode in Double-Heterostructure Injection Lasers," *IEEE J. Quant. Electron. QE-8*, 470–476, 1972.
20. S. Chinn, "Effects of Mirror Reflectivity in a Distributed Feedback Laser," *IEEE J. Quant. Electron. QE-9*, 574–580, 1973.
21. M. Nakamura, K. Aiki, J. Umceda, and A. Yariv, "CW Operation of Distributed-Feedback GaAs-GaAlAs Diode Lasers at Temperatures Up To 300 K," *Appl. Phys. Lett. 27*, 403, 1975.
22. H. Haus and C.V. Shank, "Antisymmetric Taper of DFB Laser," *IEEE J. Quant. Electron. QE-12*, 534, 1976.
23. S.L. McCall and P.M. Platzman, "An Optimized $\pi/2$ DFB Laser," *IEEE J. Quant. Electron. QE-21*, 1899, 1985.
24. H. Kogelnik and C.V. Shank, "Coupled-Wave Theory of DFB Laser," *J. Appl. Phys. 43*, 2327, 1972.
25. W.H. Steier, *The Laser Handbook*, Vol. 3, Article A5, Ed. M.L. Stitch (New York: North Holland, 1979).
26. S. DeSilvestri, P. Laporta, V. Magni, and O. Svelto, "Solid State Laser Unstable Resonators With Tapered Reflectivity Mirrors: The Super-Gaussian Approach," *IEEE J. Quant. Electron. QE 24*, 1172, 1988.
27. M.L. Tilton, G.C. Dente, A.H. Paxton, J. Cser, R.K. DeFreeze, C.E. Moeller, and D. Depatic, "High Power, Nearly Diffraction-Limited Output From a Semiconductor Laser With an Unstable Resonator," *IEEE J. Quant. Electron. 27*, 2098, 1991.

28. S. DeSilvestri, V. Magni, O. Svelto, and G. Vatentini, "Lasers With Super-Gaussian Mirrors," *IEEE J. Quant. Electron. 26*, 1500, 1990.

29. A. Yariv and P. Yeh, "Confinement and Stability in Optical Resonators Employing Mirrors With Gaussian Reflectivity Tapers," *Opt. Comm. 13*, 370–374, 1975.

30. T. Schrans and A. Yariv, "Semiconductor Lasers with Uniform Longitudinal Intensity Distribution," *Appl. Phys. 16*, 1526, 1990.

31. L.M. Miller, J.T. Verdeyen, J.J. Coleman, R.P. Bryan, J.J. Alwan, K.J. Beernik, J. Hughes, and T.M. Cockerill, "A Distributed Feedback Ridge Waveguide Quantum Well Heterostructure Laser," *Photon. Tech. Lett. 3*, 6–8, 1991.

32. Y. Tohmori, Y. Suematsu, H. Tsushima, and S. Arai, "Wavelength Tuning of GaInAsP/InP Integrated Laser with Butt-jointed Built-in Distributed Bragg Reflector," *Electronics Lett. 19*, 656–657, 1983.

33. Y. Yoshikuni, K. Oe, G. Motosugi, T. Matsuoka, "Broad Wavelength Tuning under Distributed Feedback Lasers," *Electronics Lett. 22*, 1153–1154, 1986.

34. Kotaki, Y., M. Matsuda, M. Yano, H. Ishikawa, and H. Imai, "1.55 μm Wavelength Tunable FBH-DBR Laser," *Electron. Lett., 23*, 325–327, 1987.

35. Murata, S., I. Mito, K. Kobayashi, "Tuning Ranges for 1.55 μm Wavelength Tunable DBR Lasers," *Electron. Lett. 24*, 577–579, 1988.

36. Koch, T.L., U. Koren, and B.I. Miller, "High Performance Tunable 1.5 μm InGaAs/InGaAsP Multiple Quantum Well Distributed Feedback Bragg Reflector Lasers," *Appl. Phys. Lett. 53*, 1036–1038, 1988.

37. Koch, T.L., U. Koren, R.P. Gnall, C.A. Burrus, and B.I. Miller, "Continuously Tunable 1.5 μm Multiple-Quantum-Well GaInAs/GaInAsP Distributed-Bragg-Reflector Laser," *Electron. Lett., 24*, 1431–1433, 1988.

38. X. Pan, H. Olesen and B. Tromborg, "A Theoretical Model of Multielectrode DBR Lasers," *IEEE J. Quantum. Electron. QE-24*, 2423–2432, 1988.

39. M. Oberg, S. Nilsson, T. Klinga, and P. Ojala, "A Three-Electrode Distributed Bragg Reflector Laser with 22 nm Tuning Range," *IEEE Photonics Lett. 3*, 299–310, 1991.

40. R.C. Alferness, T.L. Koch, L.L. Buhl, F. Storz, F. Heismann, M.J.R. Martyak, "Grating-Assisted InGaAsP/InP Vertical Codirectional Coupler Filter," *Appl. Phys. Lett. 55*, 2011–2013, 1989.

41. G. Griffel and A. Yariv, "Frequency Response and Tunability of Grating-Assisted Directional Couplers," *IEEE J. Quantum Electron., QE-27*, 1115–1118, 1991.

42. S. Illek, W. Thulke, M.C. Amann, "Codirectionally Coupled Twin-Guide Laser Diode for Broadband Electronic Wavelength Tuning," *Electron. Lett. 27*, 2207–2210, 1991.

43. R.C. Alferness, L.L. Buhl, U. Koren, B.I. Miller, M.G. Young, T.L. Koch, C.A. Burrus, and G. Raybon, "Tunable InGaAsP/InP Buried Rib Waveguide Vertical Coupler Filter," *Appl. Phys. Lett., 60*, 980–982, 1992.

44. R.C. Alferness, U. Koren, L.L. Buhl, B.I. Miller, M.G. Young, T.L. Koch, G. Raybon and C.A. Burrus, "Broadly Tunable InGaAsP/InP Based on a Waveguide Vertical Coupler Filter," *Appl. Phys. Lett. 60*, 3209–3211, 1992.

45. Z.M. Chuang and L.A. Coldren, "On the Spectral Properties of Widely Tunable Lasers Using Grating-Assisted Codirectional Coupling," *IEEE Photonics Technology Letters, 5*, 7–9, 1993.

46. I. Kim, R.C. Alferness, L.L. Buhl, U. Koren, B.I. Miller, M.A. Newkirk, M.G. Young, T.L. Koch, G. Raybon and C.A. Burrus, "Broadly Tunable InGaAsP/InP Vertical Coupler Filtered Laser with Low Tuning Current," *Electron. Lett. 29*, 664–666, 1993.

47. Y. Tohmori, Y. Yoshikuni, T. Tamamura, H. Ishii, Y. Kondo, and M. Yamamoto "Broad-Range Wavelength Tuning in DBR Lasers with Super Structure Grating (SSG)," *IEEE Photonics Technology Letters, 5*, 126–129, 1993.

48. Y. Tohmori, Y. Yoshikuni, H. Ishii, F. Kano, T. Tamamura and Y. Kondo, "Over 100 nm Wavelength Tuning in Superstructure Grating (SSG) DBR Lasers," *Electron. Lett. 29*, 352–354, 1993.

49. H. Ishii, Y. Tohmori, T. Tamamura, and Y. Yoshikuni, "Super Structure Grating (SSG) for Broadly Tunable DBR Lasers," *IEEE Photonic Technology Letters, 4*, 393–395, 1993.

50. V. Jayaraman, D.A. Cohen, and L.A. Coldren, "Demonstration of Broadband Tunability in a Semiconductor Laser Using Sampled Gratings," *Appl. Phys. Lett., 60*, 2321–2323, 1992.

51. V. Jayaraman, A. Mathur, L.A. Coldren, and P.D. Dapkus, "Extended Tuning Range in Sample Grating DBR Lasers," *IEEE Photonic Technology Letters, 5*, 489–491, 1993.



Maxwell's Equations and the "Classical" Atom

13.1 INTRODUCTION

The starting point for any electromagnetic problem is to find a solution to Maxwell's equations, which was the subject of the first six chapters, and more of the esoteric issues for optical fields will be addressed later. The astute student will realize, however, that we have dealt with intensities for the laser problems rather than with the fields themselves. The key simplification that allowed us to bypass (E, H) and go directly to the intensity $\mathbf{E} \times \mathbf{H}$, was the use of the Einstein coefficients and the rate equations.

There are limitations to this approach that can only be appreciated by a semiclassical quantum description of field and atom interaction, a subject for the next chapter. In that approach, the field remains classical, but the atom is quantized. This chapter provides some introductory concepts pertaining to the field-atom interaction while still keeping a classical viewpoint of both.

Although the classical approach is much simpler than the quantum one, simplicity is not the main issue. The steps to be followed are precisely those to be employed in the quantum approach (with a different equation), many of the results are independent of the approach, and finally some of the historical terminology was "invented" to make the classical model "fit" the experimental facts. For instance, terms such as "the classical A coefficient," the oscillator strength, anomalous dispersion, and the Hamiltonian of the field energy all come from the classical approach.

13.2 POLARIZATION CURRENT

The atoms enter into Maxwell's equations via the polarization current term.

$$\nabla \times \mathbf{H} = \mathbf{i}_c + n^2\epsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \frac{\partial \mathbf{P}_a}{\partial t} \tag{13.2.1}$$

where $\mathbf{i}_c$ is the conduction current carried by free carriers, $n^2\epsilon_0\partial\mathbf{E}/\partial t$ is displacement current, which includes the vacuum value and that contributed by the other atoms of the host lattice in which the active atoms are imbedded, and $\partial\mathbf{P}_a/\partial t$ is the polarization current contributed by the active atoms. Let us ignore the conduction current and combine (13.2.1) with the other Maxwell equation to obtain the wave equation.

$$\begin{aligned} \nabla \times [\nabla \times \mathbf{E}] \equiv \nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E} &= -\mu_0 \frac{\partial}{\partial t} \nabla \times \mathbf{H} \\ &= -n^2\mu_0\epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} - \mu_0 \frac{\partial^2 \mathbf{P}_a}{\partial t^2} \end{aligned} \tag{13.2.2}$$

The divergence of (13.2.1) is identically zero and, if the index is homogeneous, then we obtain

$$\nabla \cdot (\nabla \times \mathbf{H}) \equiv 0 = n^2\epsilon_0 \frac{\partial}{\partial t}(\nabla \cdot \mathbf{E}) + \frac{\partial}{\partial t} \nabla \cdot \mathbf{P}_a$$

Thus

$$\nabla \cdot \mathbf{E} = -\frac{1}{n^2\epsilon_0} \nabla \cdot \mathbf{P}_a$$

$$\nabla^2\mathbf{E} - \frac{n^2}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = \mu_0 \frac{\partial^2 \mathbf{P}_a}{\partial t^2} - \frac{1}{n^2\epsilon_0} \nabla(\nabla \cdot \mathbf{P}_a) \tag{13.2.3}$$

The last term is usually neglected—for reasons which will become clear—and thus we have a fundamental equation for the electric field of the laser.

$$\boxed{\nabla^2\mathbf{E} - \frac{n^2}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = \mu_0 \frac{\partial^2 \mathbf{P}_a}{\partial t^2}} \tag{13.2.4}$$

This equation states something very important: The polarization of the atoms is the *source* for the electric field. But it is also incomplete. What is the polarization? It is not something you look up in a handbook. Rather it is the induced (or forced) response of the atom to the electric field, and consequently we need a set of equations to obtain the response $\mathbf{P}_a(t)$.

There are always two essential parts in the computation of this response: (1) a constitutive (or defining) equation relating $\mathbf{P}_a$ to the number of active atoms and their response with due attention paid to averaging over some distribution parameter and (2) a dynamical equation relating the temporal behavior of a key atomic parameter. There are always these two parts irrespective of whether we use classical or quantum theory. Because much of the notation was invented with an all classical approach, we will use it to illustrate the procedure and to introduce some notation.

Polarization is aptly named; it is the averaged dipole moment induced by the electric field; that is, the field displaces charges (polarizing the medium) and thereby inducing a dipole. Hence, our constitutive equation is

$$\mathbf{P}_a = \langle N_a \boldsymbol{\mu} \rangle = -\langle N_a e \Delta \mathbf{x}_a \rangle \tag{13.2.5}$$

where μ is the electric dipole moment, which is equal to the electronic charge $(-e)$ times the displacement $\Delta\mathbf{x}_a$ of that charge from its equilibrium position caused by the electric field, and N_a is the number of active bound electrons addressed by the electric field. (We are presuming that only the electrons of the atoms can respond to an optical electric field. However, "holes" in a semiconductor or in an incomplete shell of molecule can also respond and then one would use $+e$ for the charge.) Equation (13.2.5) is correct for any theory with the differences in a quantum approach appearing in the averaging process, the computation of $\Delta\mathbf{x}_a$, and the definition of N_a.

The major distinction comes in the prescription for the dynamical equation of motion for $\Delta\mathbf{x}_a$, and here we limit our field of view to a classical atom obeying Newton's equation but "doctored" to include known effects. For instance, we know that atoms will have a characteristic transition frequency ω_{21}, and we also know that any kind of internal displacement of an active electron would eventually damp out in the absence of an external electric field. Thus we formulate Newton's equation of motion to read as

$$\frac{d^2\Delta\mathbf{x}_a}{dt^2} + \frac{1}{\tau}\frac{d\Delta\mathbf{x}_a}{dt} + \omega_{21}^2 \Delta\mathbf{x}_a = -\frac{e}{m}\mathbf{E} \tag{13.2.6}$$

Thus our classical model of an atom is a charged mass m tied to two brick walls by springs and some sort of a sliding friction as shown in Fig. 13.1. However farfetched such a model may appear, it will predict much of the observed interactions and much of the historical approach used it to introduce the notation.

Let us compute the response of this atom to the electrical field varying as $\mathbf{E} = \mathbf{E}_0 \exp(j\omega t)$ by recognizing that the response must have the same time history expressed by $\Delta\mathbf{x}_a(t) = \Delta\mathbf{x}_0 \exp(j\omega t)$. Thus, we find the relation between $\Delta\mathbf{x}_0$ and $\mathbf{E}_0$ to be

$$\Delta\mathbf{x}_0 = -\frac{e}{m}\frac{\mathbf{E}_0}{(\omega_{21}^2 - \omega^2) + j\omega/\tau} \tag{13.2.7}$$

Ignoring, for the moment, the averaging symbols in (13.2.5), we obtain $\mathbf{P}_a(t)$ to be

$$\mathbf{P}_a(t) = \epsilon_0 \left\{ \frac{N_a e^2}{m\epsilon_0} \left[\frac{\omega_{21}^2 - \omega^2}{(\omega_{21}^2 - \omega^2)^2 + (\omega/\tau)^2} - j\frac{(\omega/\tau)}{(\omega_{21}^2 - \omega^2)^2 + (\omega/\tau)^2} \right] \right\} \mathbf{E}_0 \exp[j\omega t] \tag{13.2.8}$$

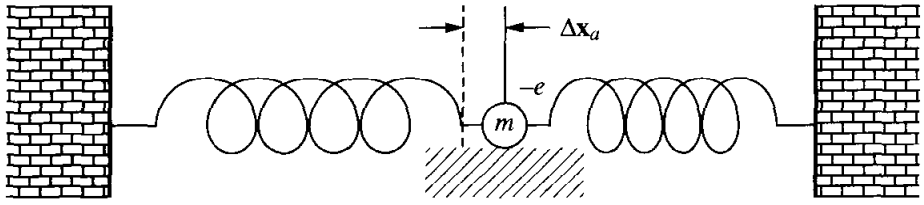


FIGURE 13.1. The mechanical model of an atom.

Notice that there are real and imaginary parts to the relation between $\mathbf{P}_a$ and $\mathbf{E}_a$, a fact of importance to lasers: You cannot have one without the other. The functional form of (13.2.8) is so important that we name the quantity in the curly braces as the complex susceptibility of the atom and write a compact representation for the relationship between $\mathbf{P}_a$ and $\mathbf{E}_a$

$$\mathbf{P}_a(t) \triangleq \epsilon_0(\chi_a' - j\chi_a'') \cdot \mathbf{E}_0 \exp[j\omega t] \tag{13.2.9}$$

where, by convention, the single prime indicates the real part and the double prime indicates the imaginary one. This is quite general and will hold true for the quantum calculation in the next chapter. It can be made as precise as is desired by due attention to averaging and even holds for negative N, which, as we will see, correspond to a population inversion.

We will have the occasion to use the expression for the ratio χ_a'/χ_a''

$$\frac{\chi_a'}{\chi_a''} = \frac{\omega_{21}^2 - \omega^2}{(\omega/\tau)} \approx \frac{2\omega_{21}(\omega_{21} - \omega)}{\omega_{21}/\tau}$$

or

$$\boxed{\frac{\chi_a'}{\chi_a''} = \left(\frac{\nu_{21} - \nu}{\Delta\nu/2}\right)} \tag{13.2.10}$$

where $\Delta\nu = 1/2\pi\tau$. This is only true for the case considered here, but (13.2.10) is a very good approximation for almost every bell-shaped response for χ_a''.

13.3 WAVE PROPAGATION WITH ACTIVE ATOMS

We return to (13.2.4) and (13.2.9) on the right side and look for uniform plane waves varying as

$$\mathbf{E}(z, t) = \mathbf{E}_0 \exp[(\gamma/2 - jk)z] \exp[+j\omega t] \tag{13.3.1}$$

where the gain coefficient $\gamma/2$ and phase constant of k are determined by a self-consistent solution to (13.2.4).

$$[(\gamma/2 - jk)^2 + (\omega n/c)^2)]E_0 \exp[j\omega t] = -\frac{\omega^2}{c^2}(\chi_a' - j\chi_a'')E_0 \exp[j\omega t] \tag{13.3.2}$$

Cancel common factors and equate real and imaginary parts to obtain

$$k^2 - (\gamma/2)^2 = \frac{\omega^2}{c^2}(n^2 + \chi_a') \tag{13.3.2}$$

$$-k\gamma = \frac{\omega^2}{c^2}\chi_a'' \tag{13.3.3}$$

Except for pathological cases, $(\gamma/2)^2 \ll k^2$ and also the term χ_a' is much less than n^2. A judicious use of the binomial theorem leads to

$$\boxed{k = \frac{\omega n}{c}\left(1 + \frac{\chi_a'}{2n^2}\right)} \tag{13.3.4}$$

which enables us to solve for the gain coefficient

$$\gamma = -\frac{\omega n}{c}\left(\frac{\chi_a''}{n^2}\right)$$

or

$$\boxed{\frac{\gamma}{k} = -\frac{\chi_a''}{n^2}} \tag{13.3.5}$$

where we have neglected the correction to k in (13.3.4) in obtaining (13.3.5).

These boxed equations are very important and are perfectly general. The first obvious conclusion to be made is that χ_a'' must be negative so that the gain coefficient γ can be positive. This presents us with a dilemma: Every term in the square bracket of (13.2.8) is positive, and thus we can only obtain gain if N is negative. Thus, our first ad hoc assumption will be to force our classical theory to agree with experiment by replacing N by

$$N \to \left(\frac{g_2}{g_1}N_1 - N_2\right) \tag{13.3.6}$$

and gain is possible if $N_2 > (g_2 g_1)N_1$. Before proceeding too far in that direction, let us make a connection with experiment and introduce some notation.

Suppose a normal state of affairs prevails with most of the atoms in the lower state, and thus $\gamma(\nu)$ is negative (i.e., absorption). With a bit of experimental care, we can measure the absorption as a function of frequency and integrate the area under the curve.

$$\text{integrated spectral absorption} = \int_0^\infty [-\gamma(\nu)]\, d\nu \tag{13.3.7}$$

where

$$\gamma(\omega) = -\left(\frac{\omega n}{c}\right)\cdot\left(\frac{\chi_a''(\omega)}{n^2}\right)$$

and

$$\frac{\chi_a''(\omega)}{n^2} = \frac{1}{n^2}\cdot\frac{Ne^2}{m\epsilon_0}\frac{(\omega/\tau)}{(\omega_{21}^2 - \omega^2)^2 + (\omega/\tau)^2} \tag{13.2.8}$$

Most of the contribution to this integral comes from frequencies close to ω_{21}. Hence we can safely replace ω by ω_{21}, except where the difference appears. Convert from ω to ν, and define a linewidth $\Delta\nu = 1/(2\pi\tau)$.

$$-\gamma(\omega) = \frac{\omega n}{c}\cdot\frac{1}{n^2}\cdot\frac{Ne^2}{m\epsilon_0}\frac{\omega/\tau}{4\omega^2[(\omega_{21} - \omega)^2 + (1/2\tau)^2]}$$

$$-\gamma(\nu) = \frac{\pi}{n} \cdot \frac{Ne^2}{mc} \cdot \left(\frac{1}{4\pi\epsilon_0}\right) \cdot \left\{ \frac{1/2\pi\tau}{2\pi[(\nu_{21} - \nu)^2 + (1/4\pi\tau)^2]} \right\}$$
$$= \frac{\pi}{n} \cdot \frac{Ne^2}{mc} \cdot \left(\frac{1}{4\pi\epsilon_0}\right) \cdot \left\{ \frac{\Delta\nu}{2\pi[(\nu_{21} - \nu)^2 + (\Delta\nu/2)^2]} \right\} \tag{13.3.8}$$

We should recognize the expression in the curly braces as the homogeneous Lorentzian line shape $g_h(\nu)$ used in Chapter 7 onward and recall the fact that its integral over all frequencies is 1. Here the integrated spectral absorption is

$$\text{integrated spectral absorption} = \int_0^\infty -\gamma(\nu)\, d\nu = \frac{\pi}{n} \frac{Ne^2}{mc} \cdot \left(\frac{1}{4\pi\epsilon_0}\right) \tag{13.3.9}$$

Equation (13.3.9) *almost* worked for the early scientists, but not quite. It appeared as if only a fraction of the active electrons, N, participated in the transition and to account for that fact and to ensure agreement, (13.3.9) was multiplied by f_{12}, the absorption oscillator strength, so as to force agreement with experiment.

$$\text{integrated spectral absorption} = \frac{\pi}{n} \frac{(Nf_{12})e^2}{mc} \cdot \left(\frac{1}{4\pi\epsilon_0}\right) \tag{13.3.10}$$

This is an exact equation—the rules were bent to make it so. Now we can use our "known" result for the absorption coefficient under the condition of $N_2 = 0$ to obtain an "exact" expression for the Einstein A_{21} coefficient.

$$\int_0^\infty [-\gamma(\nu)]\, d\nu = A_{21} \frac{\lambda_0^2}{8\pi n^2} \frac{g_2}{g_1} N_1 \int_0^\infty g(\nu)\, d\nu = A_{21} \frac{\lambda_0^2}{8\pi n^2} \frac{g_2}{g_1} N_1$$

after equating this to (13.3.10), we obtain an expression for the A coefficient.

$$A_{21} = 8\pi^2 n \cdot \frac{\nu_{21}^2}{c^3} \cdot \left(\frac{e^2}{m}\right) \cdot \frac{g_1}{g_2} \cdot f_{12} \left(\frac{1}{4\pi\epsilon_0}\right) \tag{13.3.11}$$

Thus, knowing the absorption oscillator strength is equivalent to knowing the A coefficient (and thus the B coefficients). The above derivation is not very satisfying, but it is exact because we forced it to be so.

Thus, we make our "classical" expression for the complex susceptibility correct by replacing N with a sequence of steps:

$$N \to N_1 f_{12} \to f_{21} \left[\frac{g_2}{g_1} N_1 - N_2 \right] \tag{13.3.12}$$

where $f_{21} = (g_1/g_2) f_{12}$ is the emission oscillator strength.

We can return to (13.2.10) and use (13.3.4) and (13.3.5) to show that in the vicinity of a "transition" the phase velocity of the wave also undergoes some wondrous gyrations. In a length l_g, a wave experiences a geometric phase shift of $\omega n l_g/c$ and the presence of the active atoms contributes an excess, which is given by (13.3.4). The ratio of the excess

phase shift to the line integrated gain is

$$\frac{(k - \omega n/c)l_g}{\gamma l_g} = \frac{(\omega n/c)(\chi_a'/2n^2)l_g}{-(\omega n/c)(\chi_a''/n^2)l_g} = -\frac{\chi_a'}{2\chi_a''} = -\left(\frac{\nu_{21} - \nu}{\Delta\nu}\right) \tag{13.3.13}$$

The numerator is the phase shift experienced by the electromagnetic wave in "excess" of the geometrical value and the denominator is the line integrated gain. If γl_g is negative (absorption), then the excess phase shift is positive for $\nu < \nu_{21}$ and negative for $\nu > \nu_{21}$ (i.e., an "anamolous" behavior when it was first discovered) and has the opposite behavior if we have gain. This affects the laser oscillation frequency as the following example demonstrates.

Example "Mode pulling" in a laser

A laser oscillates at a frequency such that the round trip phase shift is an integral multiple of 2π. For the geometry shown on the insert of Fig. 13.2, we have

$$\underbrace{\frac{\omega n_g}{c} \cdot 2l_g}_{\text{geometric phase shift in } l_g} + \underbrace{(k - \frac{\omega n_g}{c}) \cdot 2l_g}_{\text{excess phase shift in gain medium } l_g} + \underbrace{\frac{\omega n_2}{c} \cdot 2(d - l_g)}_{\text{geometric phase shift for rest of cavity}} = q \cdot 2\pi$$

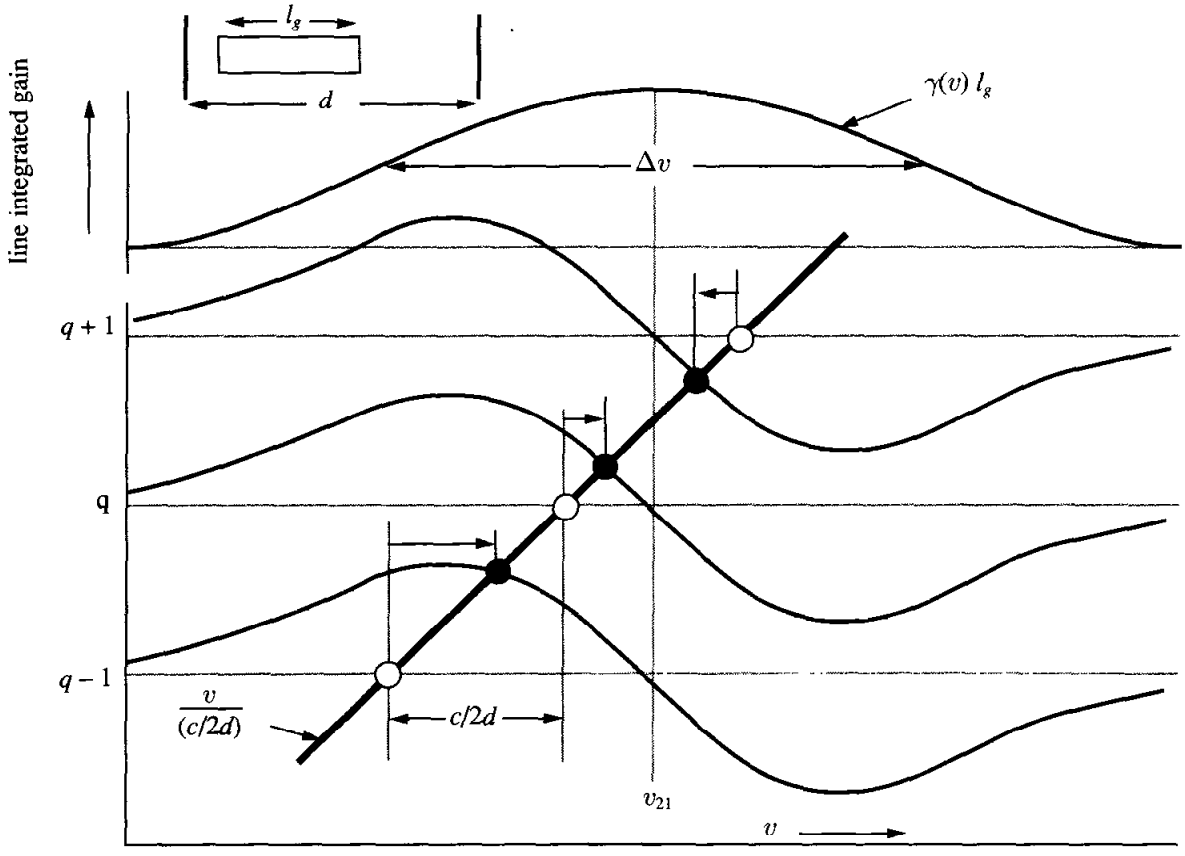


FIGURE 13.2. The graphical solution for the resonant frequencies in a cavity is shown by the solid circles and includes the dispersion contributed by the active atoms. The open circles represent the solution if the dispersion is neglected. Note that the frequencies are always pulled toward line center.

or

$$\frac{2\pi\nu(2l_g)}{c} \pm \left(\frac{\nu_{21}-\nu}{\Delta\nu}\right)\cdot\gamma(\nu)\cdot 2l_g + \frac{2\pi\nu}{c}2(d-l_g) = q\cdot 2\pi$$

where all refractive indexes have been set equal to 1 (to minimize the clutter of symbols) and the excess phase shift has been replaced by (13.3.13). Usually, the resonant frequency is given by $\nu = q(c/2d)$, and thus we solve the above to emphasize the change from the usual.

$$\frac{\nu}{(c/2d)} = q + \left(\frac{\nu_{21}-\nu}{\Delta\nu}\right)\cdot\left(\frac{\gamma(\nu)\cdot 2l_g}{2\pi}\right) \tag{13.3.14}$$

Figure 13.2 shows a "graphical solution" to (13.3.14) where the left side is shown as a straight line with a slope of $c/2d$ and the right side is a series of constants $q-1$, q, and $q+1$ with the correction term—the excess phase shift—added to each integer. (The line integrated gain $\gamma(\nu)l_g$ is presumed to be positive and is shown at the top.)

If the line integrated gain is small (or neglected), then the spacing between modes is the normal free spectral range or $c/2d$. If the gain is large, then the pulling becomes significant (and can be measured). We should also realize that we should use the saturated gain coefficient and thus the oscillation frequency of one mode will be dependent upon the power and proximity (in frequency) of the other modes.

13.4 THE CLASSICAL A_{21} COEFFICIENT

The classical model of the atom with a charge displaced from its equilibrium position by $\Delta\mathbf{x}_a$ and exhibiting simple harmonic motion at an angular frequency ω_{21} can be used to "derive" a classical value of the A coefficient without the subterfuge of using the mechanical model of the atom. Let us neglect damping in (13.2.6) and assume that something started the electron oscillating about its equilibrium position with a peak displacement $\Delta\mathbf{x}_0$.

$$\Delta\mathbf{x}_a(t) = \Delta\mathbf{x}_0 \sin\omega_{21}t \tag{13.4.1}$$

Thus, the polarization current $\partial P_a/\partial t$ from this one "atom" is

$$\frac{\partial P_a}{\partial t} = -e\cdot\Delta\dot{\mathbf{x}}_a(t) = -e\cdot\omega_{21}\cdot\Delta\mathbf{x}_0\cos\omega_{21}t = I_0\Delta\mathbf{x}_0\cos\omega_{21}t \tag{13.4.2}$$

Thus, we have an elementary electric dipole antenna of length $\Delta\mathbf{x}_0$ carrying a current $I_0 = e\omega_{21}$ (along the linear direction specified by $\Delta\mathbf{x}_0$), which has been analyzed in every elementary electromagnetic book. Such an antenna radiates a total power of

$$P_{\text{rad}} = \left(\frac{\mu_0}{\epsilon_0}\right)^{1/2}\cdot\frac{\pi}{3}\cdot\frac{(I_0\Delta\mathbf{x}_0)^2}{\lambda_{21}^2} = \frac{1}{3}\left(\frac{1}{4\pi\epsilon_0}\right)\frac{\omega_{21}^4}{c^3}(e\Delta\mathbf{x}_0)^2 \tag{13.4.3}$$

Make the identification of $e\Delta\mathbf{x}_0 = \mu_{21}$, the dipole moment associated with the transition, convert to frequency units, and assume a density of N_2 randomly oriented (in space) dipoles with uncorrelated phases to obtain the total power radiated by this group of atoms.

$$P_{\text{rad}} = \frac{16}{3}\left(\frac{1}{4\pi\epsilon_0}\right)\frac{\pi^4\nu_{21}^4}{c^3}|\mu_{21}|^2 N_2\cdot(\text{volume}) \tag{13.4.4}$$

Our prior approach using the Einstein A coefficient would indicate that power radiated by this collection of atoms would be

$$P_{\text{rad}} = h\nu_{21}A_{21}N_2 \cdot (\text{volume}) \tag{13.4.5}$$

Equating the two expressions leads to

$$A_{21} = \frac{16}{3}\left(\frac{1}{4\pi\epsilon_0}\right)\frac{\pi^4\nu_{21}^3}{c^3}\frac{|\mu_{21}|^2}{h} \tag{13.4.6}$$

This, as we will find in the next chapter, is very close to the exact answer and points out the necessity of computing the expectation value of the dipole moment.

This is the last of the "doctoring" of the physical world, and we return to Maxwell's equations and prepare them to accept a quantum calculation of the susceptibility.

13.5 (SLATER) MODES OF A LASER

We have argued that we need $\mathbf{P}_a$ to compute $\mathbf{E}$ and, of course, we need $\mathbf{E}$ to find $\mathbf{P}_a$ from the equation of motion. Both solutions must be self-consistent. One of the first glaring problems is that the fields are manipulated by both spatial and time derivatives whereas only time derivatives appear in the right side of (13.2.4). For the case of the active medium contained within a cavity, there is a formalism, initiated by J.C. Slater in World War II in connection with microwave electronics, which is applicable here and permits a self-consistent solution for the time behavior of the amplitudes of the modes in a very transparent form. This formalism describes the fields in terms of a time dependent amplitude of the characteristic modes of the cavity and then finds an "equation of motion" for the mode amplitude involving only time.

Consider a lossless ideal cavity bounded by perfectly reflecting walls. We assert that the mode functions $\mathbf{E}_m(x, y, z)$ and $\mathbf{H}_m(x, y, z)$ form a complete orthogonal set if they obey

$$\boxed{k_m\mathbf{E}_m = \nabla \times \mathbf{H}_m} \tag{13.5.1a}$$

$$\boxed{k_m\mathbf{H}_m = \nabla \times \mathbf{E}_m} \tag{13.5.1b}$$

where

$$\boxed{k_m = \omega_m\sqrt{\mu\epsilon}} \tag{13.5.1c}$$

with $\mathbf{n} \times \mathbf{E}_m = 0 = \mathbf{n} \cdot \mathbf{H}_m$ on the boundaries where $\mathbf{n}$ is the outward normal to the cavity volume.* The parameter m is a short-hand notation for mode indices, say, the triplet (m, p, q) of the Hermite-Gaussian beam modes of Chapters 3 to 6. It is easy to fall into a trap of thinking of $(\mathbf{E}_m, \mathbf{H}_m)$ as the fields inside this ideal cavity. They are, but remember that these functions have *no* time variations and, as we shall see, have "dimensions" of $(L)^{-3/2}$ not volts/meter or amps/meter. Furthermore, notice the symmetry of (13.5.1) and

*The quantities ϵ and μ (without subscripts) are shorthand for $\epsilon_r\epsilon_0$ and $\mu_r\mu_0$.

the absence of the minus sign expected in Maxwell's equations. It is better to consider the functions $\mathbf{E}_m$ and $\mathbf{H}_m$ as being the spatial part of a field much like you would consider the characteristic modes of a string clamped at $z = 0$ and $z = d$, and thus an arbitrary displacement can be expressed as a linear combination of terms of the form

$$\sin k_m z \qquad \text{or} \qquad \cos k_m z \qquad \text{where } k_m d = m\pi \tag{13.5.2}$$

In other words, $\mathbf{E}_m$ and $\mathbf{H}_m$ are spatial components of a Fourier series whose time dependent amplitudes are to be determined.

If we combine (13.5.1a) and (13.5.1b), we find that $\mathbf{E}_m$ and $\mathbf{H}_m$ obey the familiar Helmholtz equation

$$\nabla^2 \mathbf{E}_m + k_m^2 \mathbf{E}_m = 0 \tag{13.5.2a}$$

$$\nabla^2 \mathbf{H}_m + k_m^2 \mathbf{H}_m = 0 \tag{13.5.2b}$$

We also require $(\mathbf{E}_m, \mathbf{H}_m)$ to be normalized when integrated over the volume of the cavity.

$$\iiint \mathbf{E}_m \cdot \mathbf{E}_m \, dV = 1 = \iiint \mathbf{H}_m \cdot \mathbf{H}_m \, dV \tag{13.5.3}$$

which establishes the dimensions of $\mathbf{E}_m$ and $\mathbf{H}_m$ mentioned earlier. We can also prove that the Slater mode functions are orthogonal by using the traditional route to prove Poynting's theorem, but the exercise is quite boring and takes us far afield. Suffice it to say that we can easily show that

$$(k_{m'}^2 - k_m^2) \iiint (E_{m'} \cdot \mathbf{E}_m) \, dV = 0 = (k_{m'}^2 - k_m^2) \iiint (\mathbf{H}_m \cdot \mathbf{H}_{m'}) \, dV \tag{13.5.4}$$

and we chose a normalization such that

$$\iiint \mathbf{E}_m \cdot \mathbf{E}_{m'} \, dV = \delta_{mm'} = \iiint \mathbf{H}_m \cdot \mathbf{H}_{m'} \, dV \tag{13.5.5}$$

If $k_m \neq k_{m'}$, the orthogonality is proven. If $k_m = k_{m'}$, the modes are said to be degenerate and the above proof fails. In such a case, we can find a linear combination such that (13.5.4) does apply and thus establishes an orthonormal set.

13.5.1 Slater Modes of a Lossless Cavity

For a lossless and source free cavity, the actual fields obey Maxwell's equations.

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t} \tag{13.5.6a}$$

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \tag{13.5.6b}$$

The quantities $\mathbf{E}$ and $\mathbf{H}$ are fields (volts/meter, amps/meter), and there is no subscript. Expand these fields into a "Fourier series" of the Slater modes with time-dependent

amplitudes, $p_m(t)$ and $q_m(t)$ according to

$$\mathbf{E} = -\sum_m \frac{1}{\sqrt{\epsilon}} p_m(t)\mathbf{E}_m \tag{13.5.7a}$$

$$\mathbf{H} = \sum_m \frac{\omega_m}{\sqrt{\mu}} q_m(t)\mathbf{H}_m \tag{13.5.7b}$$

We might suggest using a mnemonic letter symbol, say, $e(t)$ or $h(t)$ for the amplitudes rather than p and q as was done. As we shall see, the present notation leads to a remarkable resemblance to the mechanical model of a simple harmonic oscillator with momentum p_m and displacement q_m, the traditional symbols for such variables, and this fact accounts for the choice of the symbol for the field amplitudes.

The "equations of motion" for p_m and q_m are found by substituting (13.5.7) into (13.5.6) and by using the relationships between $\mathbf{E}_m$, $\mathbf{H}_m$ (13.5.1) and the orthogonality of these modes.

$$\nabla \times \mathbf{E} = -\sum_m \frac{p_m(t)}{\sqrt{\epsilon}} \nabla \times \mathbf{E}_m = -\mu \frac{\partial \mathbf{H}}{\partial t} = -\mu \sum_m \frac{\omega_m}{\sqrt{\mu}} \dot{q}_m \mathbf{H}_m \tag{13.5.8}$$

But

$$\nabla \times \mathbf{E}_m \triangleq k_m \mathbf{H}_m \triangleq \omega_m \sqrt{\mu\epsilon}\, \mathbf{H}_m$$

Thus,
$$-\sum_m p_m(t)\omega_m \sqrt{\mu}\mathbf{H}_m = -\sum_m \omega_m \sqrt{\mu}\dot{q}_m(t)\mathbf{H}_m \tag{13.5.9}$$

Scalar multiply by $\mathbf{H}_{m'}$, integrate over the volume, and use the orthogonality of $(\mathbf{H}_m, \mathbf{H}_{m'})$ to eliminate the summation.

$$\boxed{p_m = \dot{q}_m} \tag{13.5.10}$$

Now substitute (13.5.7) into the other Maxwell equation (13.5.6a).

$$\nabla \times \mathbf{H} = \sum_m \frac{\omega_m}{\sqrt{\mu}} q_m(t) \nabla \times \mathbf{H}_m = \epsilon \frac{\partial \mathbf{E}}{\partial t} = -\epsilon \sum_m \frac{\dot{p}_m}{\sqrt{\epsilon}} \mathbf{E}_m$$

Use
$$\nabla \times \mathbf{H}_m = k_m \mathbf{E}_m = \omega_m \sqrt{\mu\epsilon}\mathbf{E}_m$$

to find
$$\sum_m \omega_m^2 \sqrt{\epsilon} q_m(t)\mathbf{E}_m = -\Sigma\sqrt{\epsilon}\, \dot{p}_m \mathbf{E}_m \tag{13.5.11}$$

Scalar multiply by $\mathbf{E}_{m'}$, integrate over the volume of the cavity, and use the orthogonality of $(\mathbf{E}_m, \mathbf{E}_{m'})$ as before

$$\boxed{\dot{p}_m = -\omega_m^2 q_m} \tag{13.5.12}$$

Equations (13.5.10) and (13.5.12) are of the form of Hamilton's "equations of motion" for a simple harmonic oscillator. If we identify q to be the generalized displacement and p

to be the generalized momentum, then the total energy in the field can be identified with the Hamiltonian:

$$H = \frac{1}{2}\iiint (\mathbf{B}\cdot\mathbf{H}+\mathbf{D}\cdot\mathbf{E})\,dV = \frac{1}{2}\mu\iiint \mathbf{H}\cdot\mathbf{H}\,dV + \frac{1}{2}\epsilon\iiint \mathbf{E}\cdot\mathbf{E}\,dV \quad (13.5.13)$$

Inserting the expansion given by (13.5.7) into (13.5.13) yields

$$H = \frac{\mu}{2}\iiint \left(\sum_m \frac{\omega_m}{\sqrt{\mu}} q_m \mathbf{H}_m\right)\cdot\left(\sum_{m'} \frac{\omega_{m'}}{\sqrt{\mu}} q_{m'}\mathbf{H}_{m'}\right) dV$$

$$+ \frac{\epsilon}{2}\iiint \left(\sum_m \frac{p_m}{\sqrt{\epsilon}}\mathbf{E}_m\right)\cdot\left(\sum_{m'} \frac{p_{m'}}{\sqrt{\epsilon}}\mathbf{E}_{m'}\right) dV$$

All integrals vanish except when $m' = m$.

$$\boxed{H = \sum_m \frac{1}{2}[(\omega_m q_m)^2 + p_m^2]} \quad (13.5.14)$$

In other words, the total electromagnetic energy looks like the sum of the energies of simple harmonic oscillators (modes) with their different characteristic frequencies of natural oscillation. Since the total energy is a constant independent of time (it has no place to go in a lossless cavity) and the modes are orthogonal, (13.5.14) states that the sum of the electric and magnetic energy in each mode (m) is also a constant, a fact obvious to anyone who has analyzed a simple LC resonant circuit. Many books use this stage as the point of departure to *quantize* the electromagnetic fields in terms of the number of photons (or energy) per mode. This results in an added degree of mathematical complexity that is not warranted for the cases encountered in this book and in much of quantum electronics. We will keep the field as a classical variable and only consider the atom as quantized, an approach which is described by the words *semi-classical quantum theory.*

13.5.2 Lossy Cavity With a Source

Now we address the more interesting question of the fields within a cavity with a polarization current present. The fields obey

$$\nabla\times\mathbf{H} = \sigma\mathbf{E} + \epsilon\frac{\partial\mathbf{E}}{\partial t} + \frac{\partial\mathbf{P}_a}{\partial t} \quad (13.5.15)$$

where σ is a quantity introduced to account for passive cavity losses, either in the medium itself or by coupling through imperfect mirrors to the outside world.

We again use the Slater mode representation for the spatial part of the field and assign a time dependent amplitude to each mode.

$$\mathbf{E} = -\sum_m \frac{p_m(t)}{\sqrt{\epsilon}}\mathbf{E}_m$$

and

$$\mathbf{H} = \sum_m \frac{\omega_m}{\sqrt{\mu}} q_m(t)\mathbf{H}_m \tag{13.5.16}$$

Manipulate (13.5.6b) with the same procedures used for the lossless cavity.

$$\nabla \times \mathbf{E} = -\sum_m \frac{p_m(t)}{\sqrt{\epsilon}} \nabla \times \mathbf{E}_m = -\mu \sum_m \frac{\omega_m}{\sqrt{\mu}} \dot{q}_m(t)\mathbf{H}_m$$

Substitute $[(k_m = \omega_m\sqrt{\mu\epsilon})\mathbf{H}_m = \nabla \times \mathbf{E}_m]$, multiply by another member of the orthonormal set of modes $\mathbf{H}_{m'}$, integrate over the volume, and recognize that the summation disappears because of orthogonality.

$$\boxed{p_m(t) = \dot{q}_m(t)} \tag{13.5.17}$$

which is the same as before [cf., (13.5.10)].

Equation 13.5.15 is manipulated in a similar fashion but now yields a result different from (13.5.12) because of the source and the losses.

$$\nabla \times \mathbf{H} = \sum_m \frac{\omega_m}{\sqrt{\mu}} q_m(t) \nabla \times \mathbf{H}_m = \sigma\mathbf{E} + \epsilon \frac{\partial \mathbf{E}}{\partial t} + \frac{\partial \mathbf{P}_a}{\partial t}$$

$$\sum_m \frac{\omega_m}{\sqrt{\mu}} q_m(t) \nabla \times \mathbf{H}_m = -\sigma \sum_m \frac{p_m(t)}{\sqrt{\epsilon}} \mathbf{E}_m - \epsilon \sum_m \frac{\dot{p}_m(t)}{\sqrt{\epsilon}} \mathbf{E}_m + \frac{\partial \mathbf{P}_a}{\partial t}$$

Substitute $(k_m = \omega_m\sqrt{\mu\epsilon})\mathbf{E}_m = \nabla \times \mathbf{H}_m$, multiply by $\mathbf{E}_{m'}$, integrate over the volume, and use as much of the orthogonality as the mathematics permits.

$$\sum_m \omega_m^2 \sqrt{\epsilon} q_m(t) \iiint \mathbf{E}_m \cdot \mathbf{E}_{m'}\, dV = -\sigma \sum_m \frac{p_m(t)}{\sqrt{\epsilon}} \iiint \mathbf{E}_m \cdot \mathbf{E}_{m'}\, dV$$
$$- \sum_m \sqrt{\epsilon}\, \dot{p}_m(t) \iiint \mathbf{E}_m \cdot \mathbf{E}_{m'}\, dV$$
$$+ \iiint \frac{\partial \mathbf{P}_a}{\partial t} \cdot \mathbf{E}_{m'}\, dV'$$

All the integrals, except the last one, vanish except when $m' = m$. The prime on dV' in the last integral is a reminder that the integration is only over that part of the cavity filled by the atoms.

$$\dot{p}(t) + \frac{\sigma}{\epsilon} p_m(t) + \omega_m^2 q_m(t) = \frac{1}{\sqrt{\epsilon}} \iiint \frac{\partial P_a}{\partial t} \cdot \mathbf{E}_{m'}\, dV' \tag{13.5.18}$$

Now differentiate (13.5.18) with respect to time and use (13.5.17) for $\dot{q}_m$.

$$\boxed{\ddot{p}_m + \frac{\sigma}{\epsilon} \dot{p}_m + \omega_m^2 p_m = \frac{1}{\sqrt{\epsilon}} \iiint \left(\frac{\partial^2 \mathbf{P}_a}{\partial t^2} \cdot \mathbf{E}_{m'}\, dV' \right)} \tag{13.5.19}$$

Let us assume a relationship between $\mathbf{P}_a$ and $\mathbf{E}$ of the form indicated by (13.2.9):

$$\mathbf{P}_a = \epsilon_0 \chi_a \mathbf{E} = \epsilon_0(\chi_a' - j\chi_a'')\mathbf{E}$$

where the susceptibility of χ_a may be a nonlinear function of the magnitude of the total field but is assumed to be slowing varying on a time scale of an optical cycle. Thus, the second time derivative of $\mathbf{P}_a(t)$ becomes

$$\frac{\partial^2 \mathbf{P}_a}{\partial t^2} = -\epsilon_0 \chi_a \sum_m \frac{\ddot{p}_m(t)}{\sqrt{\epsilon}} \mathbf{E}_m \tag{13.5.20}$$

and the overlap integral becomes (with the assumption of $d\chi_a/dt \sim 0$)

$$\frac{1}{\sqrt{\epsilon}} \iiint \frac{\partial^2 \mathbf{P}_a}{\partial t} \cdot \mathbf{E}_{m'} \, dV' = -\frac{\epsilon_0}{\epsilon} \sum_m \chi_a \ddot{p}_m \iiint \mathbf{E}_m \cdot \mathbf{E}_{m'} \, dV' \tag{13.5.21}$$

We can no longer eliminate the summation because the integration is not over the complete volume of the cavity. This fact couples modes m to m', and the strength of that coupling can be partially evaluated by this integral. More important is the fact that the atomic susceptibility χ_a is a nonlinear function of the *total* electric field, and this fact introduces additional coupling between the modes. Equation (13.5.20) has already avoided this problem through the assumption that $d\chi_a/dt = 0$.

Let us ignore the coupling issue but keep the idea that χ_a is an implicit function of the amplitude of the total field and approximate the summation by a fill-factor f:

$$\frac{1}{\sqrt{\epsilon}} \iiint \frac{\partial^2 \mathbf{P}_a}{\partial t^2} \cdot \mathbf{E}_{m'} \, dV' \approx -\frac{\epsilon_0 f}{\epsilon} \chi_a \ddot{p}_m = -f \frac{\chi_a}{n^2} \ddot{p}_m \tag{13.5.22}$$

where

$$f = \iiint_m \mathbf{E}_m \cdot \mathbf{E}_{m'} \, dV' < 1 \tag{13.5.23}$$

The fill factor f is related to the optical confinement factor Γ used in the semiconductor literature. Thus, our equation of motion for the electric field amplitude of the m mode obeys

$$\boxed{\ddot{p}_m(t) + \left(\frac{\sigma}{\epsilon}\right) \dot{p}_m(t) + \omega_m^2 p_m(t) = -f \frac{\chi_a}{n^2} \ddot{p}_m(t)} \tag{13.5.24}$$

Let us examine this equation under a few special circumstances to appreciate the implications and complications. This will allow us to identify factors in terms of more familiar quantities. At the very least you should now have the appreciation of the fact that the atom susceptibility is the driving term for the field.

13.6 DYNAMICS OF THE FIELDS

13.6.1 Excitation Clamped to Zero

Let us assume that the atomic driving function $\mathbf{P}_a$ is suddenly clamped to zero after some initial value of electric field $p_m(t = 0) = p_0$ has been established. Then, we need a solution

to (13.5.25) with the right side equal to zero.

$$\ddot{p}_m + \left(\frac{\sigma}{\epsilon}\right)\dot{p}_m + \omega_m^2 p_m = 0$$

$$\therefore\ p_m(t) = p_0 \exp[-(\sigma/2\epsilon)t]\cos\omega_m t \tag{13.6.1a}$$

and

$$q_m(t) = \frac{1}{\omega_m} p_0 \exp[-(\sigma/2\epsilon)t]\sin\omega_m t \tag{13.6.1b}$$

where $\sigma/2\epsilon$ is assumed to be much less than ω_m. Since the total energy in the field of the m^{th} mode (i.e., the Hamiltonian) is $[p_m^2(t) + \omega_m^2 q_m^2(t)]/2$ we find that

$$H_m(t) = \frac{1}{2} p_0^2 \exp[-(\sigma/\epsilon)t](\cos^2\omega_m t + \sin^2\omega_m t) = H_m(0)\exp(-t/\tau_p) \tag{13.6.2}$$

where $H_m(0)$ is initial energy in the m^{th} mode of the cavity at $t = 0$. Equation (13.6.2) shows that, if all energy sources are removed, the total energy of the m mode would decay with the classical photon lifetime τ_p given by

$$\tau_p := \frac{\epsilon}{\sigma} \triangleq \frac{Q_m}{\omega_m} \tag{13.6.3}$$

where Q_m is the quality factor for the field in the m mode with a resonant frequency ω_m. This equation defines the fictional conductivity σ in terms of the cavity specifications.

Now let us return to (13.5.24) and use the above ideas to rewrite it.

$$\boxed{\ddot{p}_m(t) + \frac{1}{\tau_p}\dot{p}_m(t) + \omega_m^2 p_m(t) = -f\frac{\chi_a}{n^2}\ddot{p}_m(t)} \tag{13.6.4}$$

Note that (13.6.4) states that the inhomogeneous term (the right-hand side with the atoms χ_a) is the *driving function* for the response, the electric field whose dynamics is governed by the differential equation on the left-hand side.

13.6.2 Time Evolution of the Field

The previous subsection taught us that there will be two factors associated with the electric field: an *envelope* function, which, in the previous case, was $p_0 \exp[-t/2\tau_p]$, and the harmonic term of the form $Re\{\exp[j\omega t]\}$, where ω is on the order of but not necessarily equal to ω_m, the resonant frequency of the passive cavity. We also note that if (13.6.4) is multiplied by $-\epsilon^{1/2}$, then all terms involving p_m are proportional to the electric field. Thus, it will save us some future work if we assume that fact at the beginning and choose

$$p_m(t) = E_m(t)\exp[j\omega t] \tag{13.6.5}$$

and assume that the envelope function has slow enough time variation such that

$$\ddot{E}_m(t) \ll \omega\dot{E}_m(t) \tag{13.6.6}$$

an assumption that avoids unnecessary complications and needless headaches. We now use the mnemonic symbol $E_m(t)$ for the envelope of the electric field to remind ourselves of the quantity being followed in time. Equation (13.6.6) is not a serious assumption if we consider the numerics of a typical case:

> The angular frequency ω is on the order of 10^{14} $\sec^{-1}$, whereas a reasonably fast envelope for the field has a rise time of $10^{-9} - 10^{-12}$ sec. Hence, (13.6.6) is well obeyed.

Please note also that the angular frequency ω is *not* equal to the natural resonant frequency ω_m of the Slater mode, although we can anticipate that the two frequencies are close. The detuning $\omega - \omega_m$ remains to be determined.

The various terms in (13.6.4) become

$$p_m(t) = E_m(t) \exp[j\omega t] \tag{13.6.7a}$$

$$\dot{p}_m(t) = [j\omega E_m(t) + \dot{E}_m(t)] \exp[j\omega t] \tag{13.6.7b}$$

$$\ddot{p}_m(t) = [-\omega^2 E_m(t) + 2j\omega \dot{E}_m(t) + \ddot{E}_m(t)] \exp[j\omega t] \tag{13.6.7c}$$

$$\simeq [-\omega^2 E_m(t) + 2j\omega \dot{E}_m(t)] \exp[j\omega t] \tag{13.6.7d}$$

where (13.6.6) has been used to neglect the last term in (13.6.7c). Insert those terms into (13.6.4) and collect factors with common order of derivatives

$$\left\{ \left[(\omega_m^2 - \omega^2) + j \frac{\omega}{\tau_p} \right] E_m(t) + \left[2j\omega + \frac{1}{\tau_p} \right] \dot{E}_m(t) \right\} \exp[j\omega t]$$

$$= -\frac{1}{n^2} (\chi_a' - j\chi_a'') f \frac{\partial^2}{\partial t^2} \left\{ E_m(t) \exp[j\omega t] \right\} \tag{13.6.8}$$

The inhomogeneous term [i.e., right side in (13.6.8)] is kept separate from the response (i.e., left side) to emphasize the role played by each. We have also assumed that the susceptibility can be broken up into real and imaginary parts, both of which are dependent on the total electric field in the cavity. To aid in the interpretation of the various terms, we become very specific and consider the simplest case of a CW laser.

Example 1: A CW or Steady State Laser. For a steady state laser, the envelope function is a constant and thus $\dot{E}_m = \dot{E}_m = 0$ and $E_m(t) = E_0$ (a constant).

Thus

$$\frac{\partial^2}{\partial t^2} [E_m(t) \exp(j\omega t)] = -\omega^2 E_0 \exp[j\omega t]$$

and (13.6.8) simplifies to

$$\left[(\omega_m^2 - \omega^2) + j \frac{\omega}{\tau_p} \right] E_0 \exp[j\omega t] = \frac{\omega^2}{n^2} f(\chi_a' - j\chi_a'') E_0 \exp[j\omega t] \tag{13.6.9}$$

Equate real and imaginary parts of (13.6.9) and cancel $\exp[j\omega t]$ to obtain

$$\text{real part} \Rightarrow \qquad \left[(\omega_m^2 - \omega^2) - f\omega^2 \frac{\chi_a'}{n^2}\right] \cdot E_0 = 0 \tag{13.6.10a}$$

$$\text{imaginary part} \Rightarrow \qquad \left[\frac{\omega}{\tau_p} + f\omega^2 \cdot \frac{\chi_a''}{n^2}\right] \cdot E_0 = 0 \tag{13.6.10b}$$

Obviously, the field amplitude $E_0 \neq 0$ and thus (13.6.10) specifies a relationship between the angular frequency of oscillation ω the natural mode frequency ω_m, and the Q or quality factor of the cavity to the real and imaginary parts of the susceptibility of the atoms (which may be implicitly dependent upon the total field inside the cavity).

Equation (13.6.10a) yields the detuning of oscillation frequency ω from the passive resonance ω_m.

$$\omega^2 = \frac{\omega_m^2}{1 + f(\chi_a'/n^2)} \tag{13.6.11}$$

Thus, depending upon the sign of χ_a', the oscillation frequency deviates from the characteristic Slater mode frequency. As we have seen earlier, the quantity χ_a'/n^2 is small compared to 1, is negative for $\omega_m < \omega_0$ and positive for $\omega_m > \omega_0$, and thus all oscillation frequencies are pulled *toward* ω_0, the center of the atomic transition as was found in Sec. 13.3. The answer is the same as found previously but now it arises in a "natural" fashion.

Equation (13.6.10b) states that χ_a'' must be related to passive cavity photon lifetime or Q by

$$-\frac{f\chi_a''}{n^2} = \frac{1}{\omega\tau_p} = \frac{1}{Q} \tag{13.6.12}$$

This equation is equivalent to the usual one of requiring the saturated gain (or χ_a'') per pass to make up for the passive losses, but now it arises in a natural manner for the field formulation. To show this, we recall from (13.3.5)

$$\frac{\chi_a''}{n^2} = -\frac{\gamma}{k}$$

with $k \approx \omega n/c$ and γ equal to the saturated gain coefficient for intensity. Hence, (13.6.12) becomes

$$-f\left(\frac{-\gamma c}{\omega n}\right) = \frac{1}{\omega\tau_p} \qquad \text{or} \qquad f\gamma = \frac{1}{(c/n)} \cdot \frac{1}{\tau_p} \tag{13.6.13}$$

The photon lifetime of a passive cavity is defined by the following statement in words

$$[\text{photon lifetime} = \tau_p] = \frac{\text{time for a round trip}}{\text{fractional loss per round trip}} \tag{6.42}$$

For simple cavity of length d, this statement becomes

$$\tau_p = \frac{2nd/c}{\text{fractional loss per round trip}}$$

Hence (13.6.13) becomes

$$f\gamma = \frac{\text{fractional loss per round trip}}{2\,d} \tag{13.6.14}$$

The fill factor f can be approximated by ratio of the gain length l_g to the total cavity length d, and thus (13.6.14) becomes

$$\boxed{\gamma l_g = \text{fractional loss per pass}}$$

Notice that there is no subscript 0 on the gain coefficient γ, implying that we must use the saturated value of the gain coefficient. This equation also restates the fact that the line integrated gain must be exactly equal to the fractional loss for a CW laser. At this stage of our development we can only use the value derived previously; that is,

$$\gamma = \frac{\gamma_0}{1 + I/I_s}$$

and obtain the same answer as in Chapter 9.

Example 2: Dynamics of an Active Cavity: General Considerations. The first case ignored the approach to equilibrium by assuming CW operation. Let us now return to (13.6.8), repeated for convenience, and follow its time history.

$$\left\{\left[(\omega_m^2 - \omega^2) + j\frac{\omega}{\tau_p}\right]E_m(t) + \left[j2\omega + \frac{1}{\tau_p}\right]\dot{E}_m(t)\right\}\exp[j\omega t]$$

$$= -\frac{1}{n^2}(\chi_a' - j\chi_a'')f\frac{\partial^2}{\partial t^2}[E_m(t)\exp[j\omega t]] \tag{13.6.8}$$

After neglecting the second derivative $E_m(t)$ in comparison to $\omega\dot{E}_m(t)$ [(cf., 13.6.6)], the second derivative on the right side becomes

$$\frac{\partial^2}{\partial t^2}[E_m(t)\exp[j\omega t]] \approx [-\omega^2 E_m(t) + 2j\omega\dot{E}_m(t)]\exp(j\omega t) \tag{13.6.15}$$

Now collect real and imaginary parts of the right side of (13.6.8)

$$\text{RHS of (13.6.8)} = \left\{ + \omega^2 \cdot \frac{f\chi_a'}{n^2} \cdot E_m(t) - 2j\omega\frac{f\chi_a'}{n^2} \cdot \dot{E}_m(t) \right.$$

$$\left. - j\omega^2 \cdot \frac{f\chi_a''}{n^2} \cdot E_m(t) - 2\omega \cdot \frac{f\chi_a''}{n^2} \cdot \dot{E}_m(t)\right\} \cdot \exp[j\omega t]$$

Arrange terms according to the ascending order of the derivatives of $E_m(t)$

$$\left\{\omega_m^2 - \omega^2\left(1 + f\frac{\chi_a'}{n^2}\right) + j\omega\left[\frac{1}{\tau_p} + \omega f\frac{\chi_a''}{n^2}\right]\right\}E_m(t)$$

$$+ \left\{2j\omega\left[1 + f\frac{\chi_a'}{n^2}\right] + \left[\frac{1}{\tau_p} + 2\omega f\frac{\chi_a''}{n^2}\right]\right\}\dot{E}_m(t) = 0 \tag{13.6.16}$$

Two reasonable approximations are in order here to reduce the volume of different factors.

1. Assume

$$\omega_m^2 = \omega^2\left(1 + f\frac{\chi_a'}{n^2}\right)$$

which is equivalent to ignoring the dynamics of the pulling of the oscillation frequency toward line center. The fact that it does occur leads to a "chirp" in the laser frequency, but, while it is a measurable effect, it is small.

2. Neglect $f\chi_a'/n^2$ as compared to 1 in the second factor. (This is just laziness on our part but it also is small.)

Equation (13.6.16) becomes a bit more manageable.

$$\left[2j\omega + \frac{1}{\tau_p} + 2\omega\frac{f\chi_a''}{n^2}\right]\dot{E}_m(t) + j\omega\left[\frac{1}{\tau_p} + \omega\frac{f\chi_a''}{n^2}\right]E_m(t) = 0 \qquad (13.6.17)$$

The quantity $2j\omega$ is much larger than the second and third terms of the first bracket, and thus (13.6.17) becomes even simpler.

$$\dot{E}_m(t) + \frac{1}{2}\left[\frac{1}{\tau_p} + \omega f \cdot \frac{\chi_a''}{n^2}\right]E_m(t) = 0$$

Now we again insert the fact that $\chi_a''/n^2 = -\gamma/k$ and $k \approx \omega n/c$ (13.3.13) and rewrite

$$\boxed{\dot{E}_m(t) + \frac{1}{2}\left(\frac{1}{\tau_p} - \frac{fc\gamma}{n}\right)E_m(t) = 0} \qquad (13.6.18)$$

Obviously, if the envelope of the field is to grow in time, then the factor multiplying $E_m(t)$ must be negative—equilibrium (or CW operation) is established when it is equal to zero.

Example 3: Time Scale for Growth of a Pulsed Laser. It is instructive to estimate the time scale for growth of the laser amplitude from small value provided by spontaneous emission (which has been neglected) Let

$\gamma = 0.05\ \text{cm}^{-1}$ (i.e., 5% gain in intensity per cm)
$n = 1$ (a gas)
$f = 1$ (a completely filled cavity)
$\therefore fc\gamma = 1.5 \times 10^9\ \text{sec}^{-1}$

If the photon lifetime of the passive cavity $\tau_p = 1$ ns, then the growth or e-folding rate is $0.25 \times 10^9\ \text{sec}^{-1}$, which translates to a growth time constant of 4 ns. If we accept the conventional wisdom that the intensity has to grow by e^{30} (from spontaneous seeding), then the field must grow by 15 e-folding times or at least 60 ns to approach its CW value for the

above numbers. This is the same result obtained from the rate equation analysis in Chapter 9.

Example 4: A Bistable Cavity. Equation (13.6.18) can also be used for the case of a cavity filled with a saturable absorber (or a subthreshold gain) driven by an external source as depicted in Fig. 13.3.

The modification is to include an external driving field E_x, and we write (13.6.18) in terms of the internal electrical field

$$\dot{E} + \frac{1}{2}\left\{\frac{1}{\tau_p} + \frac{\alpha_0(c/n)}{1 + (E/E_s)^2}\right\} \cdot E = \kappa_1 E_x \tag{13.6.19}$$

where κ_1 represents the coupling of the external field to the internal one by transmission through M_1, a homogeneous saturation law is assumed for the absorption coefficient α whose small signal value is α_0, and E_s is defined by

$$\frac{1}{2}\frac{E_s^2}{\eta_0} = I_s \tag{13.6.20}$$

Equation (13.6.19) has some very interesting characteristics of practical significance. While it is impossible to solve analytically for the transient case, there are some trends easily discernable for a steady state. Normalize all fields to E_s and let

$$y = E/E_s \qquad x = E_x/E_s$$

then (13.6.19) becomes

$$\dot{y} + \frac{1}{2\tau_p}\left(1 + \frac{\alpha_0(c/n)\tau_p}{1 + y^2}\right) y = \kappa_1 x$$

For $y = E/E_s \ll 1$ and $\dot{y} = 0$, we have a simple relation between y and x

$$y = \frac{2\kappa_1\tau_p}{1 + \alpha_0(c/n)\tau_p} x$$

If the external input is large so that $y = E/E_s \gg 1$ and $\dot{y} = 0$, the relation between y (and thus the output) and x has a much larger slope.

$$y = (2\kappa_1\tau_p)x$$

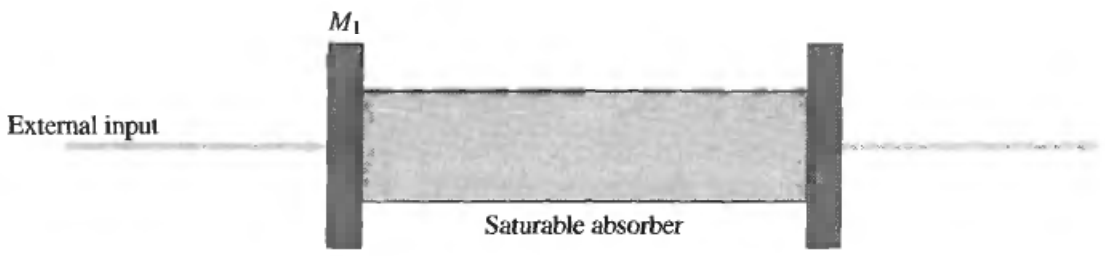


FIGURE 13.3. A bistable optical cavity.

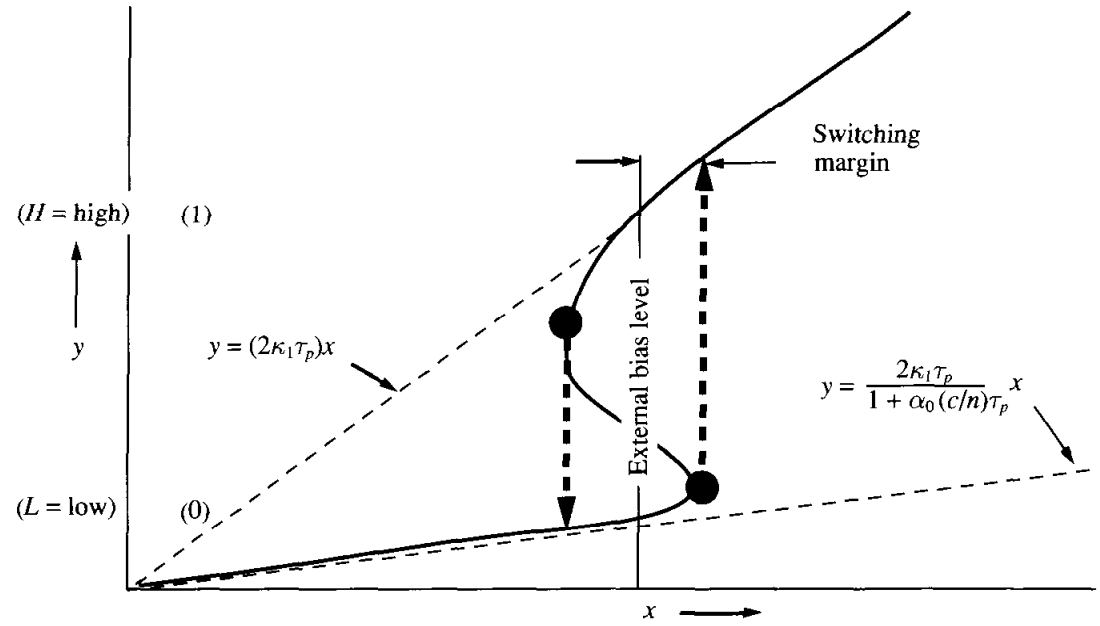


FIGURE 13.4. The relative output (y) as a function of the input for the bistable cavity.

If $\alpha_0(c/n)\tau_p > 8$, the output/insert transfer characteristics have the graphical format of Fig. 13.4 and it is easily shown that there are two values of x for which $dx/dy = 0$.

If the cavity is biased at the point indicated, there are two stable operating points (L, H), and thus the output can be either low or high. If the external bias maintains the cavity at the low (or 0 state) and then is increased by the switching margin indicated, there is an enormous jump in the internal field y and thus the output through the other side. After this temporary increase is over, the cavity will return to the high (or 1 state) and remain there. If another Δx (positive) is applied by the external source, the resultant transmission increase is small. If Δx is negative and beyond the switching margin, the cavity returns to the 0 state.

Obviously, this can be used for an optical flip-flop, an essential element for optical computing.

13.7 SUMMARY

We now have the tools to utilize a quantum calculation of the complex susceptibility and thus complete the semiclassical description of the laser. There are some key equations that bear repeating without the fog of the mathematical detail leading to them. They are

$$\boxed{\nabla^2\mathbf{E} - \frac{n^2}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} = \mu_0\frac{\partial^2\mathbf{P}_a}{\partial t^2}} \tag{13.2.4}$$

and a constitutive equation relating $\mathbf{P}_a$ to $\mathbf{E}$

$$\boxed{\mathbf{P}_a(t) \triangleq \epsilon_0(\chi_a' - j\chi_a'')\mathbf{E}_0 \exp[j\omega t] = \langle N_a \boldsymbol{\mu} \rangle} \tag{13.2.9}$$

The real and the imaginary parts of the complex susceptibility are related to the gain coefficient and the phase constant by

$$\boxed{\frac{(k - \omega n/c)l_g}{\gamma l_g} = -\frac{\chi_a'}{2\chi_a''} = -\left(\frac{\nu_{21} - \nu}{\Delta\nu}\right)} \tag{13.3.13}$$

The laser field can be expanded in terms of the orthonormal Slater modes defined by

$$\boxed{\mathbf{E} = -\sum_m \frac{p_m(t)}{\sqrt{\epsilon}}\mathbf{E}_m \quad \text{and} \quad \mathbf{H} = \sum_m \frac{\omega_m}{\sqrt{\mu}} q_m(t)\mathbf{H}_m} \tag{13.5.16}$$

leading to an equation of motion for the field amplitudes

$$\boxed{p_m(t) = \dot{q}_m(t)} \tag{13.5.17}$$

and

$$\boxed{\ddot{p}_m(t) + \frac{1}{\tau_p}\dot{p}_m(t) + \omega_m^2 p_m(t) = -f\frac{\chi_a}{n^2}\ddot{p}_m} \tag{13.6.4}$$

While all of these have been derived by using classical concepts, the formalism is still valid when the atom is quantized, and thus the next chapter is devoted to the computation of χ_a under those circumstances.

PROBLEMS

13.1. If a bound electron undergoes simple harmonic motion at a frequency ω_0, it is being accelerated and thus radiates power at the frequency ω_0. Therefore the amplitude of the simple harmonic motion must be damped because of that radiation, since the total energy of system (bound electron + radiation) must be conserved.

(a) Use such arguments to show that the equation of motion for the bound electron is

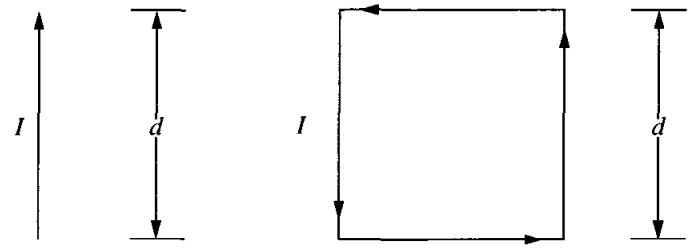
$$\ddot{x} + \omega_0^2 x - \frac{1}{6\pi\epsilon_0}\frac{e^2}{mc^3}\dddot{x} = 0$$

(b) Assume a displacement of the form $x = d\exp[(1/2\tau) + j\omega_0]t$. Neglect all terms involving powers of $1/\tau$ greater than 1, and show that the radiation damping is given by

$$\frac{1}{\tau} = \frac{1}{6\pi\epsilon_0}\frac{e^2\omega_0^2}{mc^3}$$

13.2. The power radiated by a classic magnetic dipole, $\mathbf{m} = \mathrm{IA}$, is

$$P_{\text{rad}} = \frac{1}{2\pi} \cdot \frac{\omega^4}{c^4} \cdot \sqrt{\frac{\mu_0}{\epsilon_0}} \cdot \mathbf{m}^2$$



Electric dipole Magnetic dipole

(a) Find an expression for the classical A_m coefficient for magnetic dipole radiation.

(b) Evaluate the ratio of the A coefficient derived in (a) to that of an electric dipole assuming the above geometry for each. Assume $d = 1$ Å, $\lambda_0 = 10,000$ Å, and the same current for each.

13.3. Use the data provided in Chapter 10 concerning the ruby system to estimate the amount of pulling on a mode that would have been located 1 GHz away from line center. Assume an inversion sufficient to sustain a gain coefficient of 0.05 cm^{-1} (at line center), oscillation on R_1 line at 6943 Å with $E \perp$ to c axis, and a temperature of 300 K. Approximate the line shape by a Lorentzian.

13.4. Compute the effect of atomic dispersion on the group velocity of an optical pulse propagating in an atomic medium with the frequency center of the pulse coinciding with the atomic resonance for both a normal and an inverted population. Express the group velocity as a function of the peak loss (or gain) coefficient for the case of a Lorentzian line. Ignore hole burning, and assume that the spectrum of the pulse is narrow compared to the width of the line.

13.5. The frequency spacing between modes of a low gain laser, such as the 6328 Å laser, is given by $\nu_{0,0,q+1} - \nu_{0,0,q} = c/2d$. For high gain lasers, the dispersion caused by the inverted population modifies this formula. (See 13.3.) Consider the He:Ne laser transition at 3.39 μm ($3s_2 - 3p_4$) with a small-signal gain coefficient of 30 dB/m, line width of $\sim$ 300 MHz, and $A_{21} = 2.87 \times 10^6$ sec^{-1} with $g_2/g_1 = 3/5$ (see Table 10.3).

(a) Evaluate the stimulated emission cross section.

(b) Express 30 dB/m in terms of a fraction per centimeter.

(c) What must be the inversion $N_2 - (g_2/g_1)N_1$ to obtain the 30 dB/m gain coefficient?

(d) Derive and evaluate the mode spacing for the two modes "close" to ν_0.

13.6. The simplified form of the real and imaginary parts of the susceptibility as expressed by (13.2.10) and (13.3.8) do not obey the Kramer-Kronig relations, whereas that expressed by (13.2.8) does. What has changed?

13.7. **(a)** Find the orthonormal Slater mode functions, $\mathbf{E}_m$ and $\mathbf{H}_m$, *uniform* in the $(x, y$ plane (but of limited extent over a diameter $2a$) for the parallel-plane Fabry Perot cavity shown below.



(b) Suppose we consider a cavity driven by an external wave that can be repre sented by electric and magnetic current sources at the entrance ($z = 0$) plan of the cavity. Maxwell's equations become

$$\nabla \times \mathbf{H} = \mathbf{J}_t \delta(z) + \sigma \mathbf{E} + \epsilon \frac{\partial \mathbf{E}}{dt} + \frac{\partial \mathbf{P}_a}{\partial t}$$

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} - \frac{\partial \mathbf{M}_t}{\partial t} \delta(z)$$

Expand the field (**E**, **H**) in the usual fashion to find an equation of motion fo the electric field amplitude $p_m(t)$ of the m mode involving the source term $\mathbf{M}_t$ and $\mathbf{J}_t$ (t is transverse) and the polarization term

$$\frac{d^2 p_m(t)}{dt} + \frac{1}{\tau_p} \frac{dp_m(t)}{dt} + \omega_m^2 p_m(t) = ?$$

13.8. Derive the inhomogeneous wave equation for the optical electric field without assuming $\nabla \cdot \mathbf{E} = 0$.

(a) Present an argument justifying the neglect of $\nabla(\nabla \cdot \mathbf{P}_a)$ in comparison to the other sources; that is, $\mu(\partial^2/\partial t^2)\mathbf{P}_a$. Assume σ and the free charge density are zero.

(b) What is the corresponding equation for optical magnetic field **H**.

13.9. Consider the case of a cavity excited by an external field with $\Delta\omega = \omega_m - \omega \neq 0$.

(a) Add the external drive as was done in Example 4, and make the usual assumptions: $\omega_m + \omega(1 + f\chi_a')^{1/2} \approx 2\omega$, $(1 + f\chi_a''/n^2)^{1/2} \approx 1 + f\chi_a''/2n^2$, $|2j\omega| \gg [1/\tau_p + 2\omega f\chi_a''/n^2]$, and use the relations expressed by (13.2.10) and (13.3.4) to simplify (13.6.16) to

$$\dot{E}_m + \left\{ \left[\frac{1}{2\tau_p} - \frac{fc\gamma}{n} \right] - j \left[(\omega_m - \omega) + \frac{fc\gamma}{n} \left(\frac{\omega_{21} - \omega}{\Delta\omega_a} \right) \right] \right\} E_m = \kappa E_x$$

where ω_{21} is atomic transition frequency, ω_m is cavity resonance of the m mode, $\Delta\omega_a$ is atomic line width, τ_p is photon lifetime of the passive cavity with $\gamma_0 = 0$, and assume the homogeneous saturation law

$$\gamma = \frac{\gamma_0}{1 + (E/E_s)^2}$$

γ_0 is small signal value (and can be negative).

(b) Express this equation in terms of a normalized time $T = t/\tau_p$, the passive cavity Q, the cavity linewidth $\Delta\omega_c$, and the fractional detuning from cavity resonance $\delta = (\omega_m - \omega)/\omega$ to show

$$\dot{E}_m + \frac{1}{2}\left\{\left[1 - \frac{2fc\gamma}{n}\frac{1}{\Delta\omega_c}\right] - j2\delta Q\left[1 + \frac{fc\gamma}{n}\frac{1}{\Delta\omega_a}\right]\right\} E_m = \kappa E_x$$

(c) Discuss the relative importance of saturation on the real and imaginary parts of this differential equation.

13.10. Start with Maxwell's equations and include the polarization and magnetization terms

$$\nabla \times \mathbf{H} = \mathbf{i} + \frac{\partial}{\partial t}(\epsilon_0\mathbf{E} + \mathbf{P}) \text{ and } \nabla \times \mathbf{E} = -\frac{\partial}{\partial t}[\mu_0(\mathbf{H} + \mathbf{M})]$$

Derive Poynting's theorem in the form

$$-\int_V \nabla \cdot (\mathbf{E} \times \mathbf{H})\, dV = -\iint_S \mathbf{E} \times \mathbf{H} \cdot d\mathbf{A}$$

$$= \int_V \left[\mathbf{E} \cdot \mathbf{i} + \frac{\partial}{\partial t}\left(\frac{\epsilon_0\mathbf{E} \cdot \mathbf{E}}{2}\right) + \frac{\partial}{\partial t}\left(\frac{\mu_0\mathbf{H} \cdot \mathbf{H}}{2}\right) + \mathbf{E} \cdot \frac{\partial \mathbf{P}}{\partial t} + \mu_0\mathbf{H} \cdot \frac{\partial \mathbf{M}}{\partial t}\right] dV$$

13.11. Consider a cavity with a saturable loss driven by an external source as was done in Sec. 13.6.

(a) Prove that $f\alpha_0(c/n)\tau_p$ must be greater than 8 to have optical bistability.

(b) Label the transfer characteristic of Fig. 13.4 showing the "bias" value of the external field, the values of a 1 and a 0, and the change in the external field required to switch the state of the cavity by using $\alpha_0 f(c/n)\tau_p = 9$.

(c) If the cavity is in the 0 state, then we must increase the external signal by *more* than the switching margin for a sufficient time interval to cause a transition to state 1. We do not need to keep this excess field on until the 1 state has been reached, but we do need to encourage y to increase sufficiently until the instability can drive it the rest of the way. Discuss the interplay between product of the excess injected field times the time interval needed to cause the system to switch.

(d) We might be tempted to write dynamical equation for the internal intensity as

$$\frac{d}{dt}\left[\frac{I}{I_s}\right] + \frac{1}{\tau_p}\left\{1 + \frac{f\alpha_0(c/n)\tau_p}{1 + (I/I_s)}\right\}\left[\frac{I}{I_s}\right] = \kappa\frac{I_x}{I_s}$$

to take the place of (13.6.19), where I_x is the injected external intensity. Show that this equation does *not* predict bistability. At first glance, this equation might appear as a simple multiplication of (13.6.19) by E and a relabeling of

the variables, or it might appeal to those wanting to follow "photons" rather than fields. What has gone wrong? (NOTE: We can follow photons, but not using the simple minded reasoning given above.)

13.12. Assume a passive cavity (i.e., $\gamma = 0$) with a nonlinear index $n = n_0 + n_2|E|^2$ such that the resonant frequency $\omega_m = 2\pi[c/2nd]$ is dependent upon the amplitude of the field. If the frequency of the external signal ω is chosen correctly with respect to the small signal value, $\omega_{m0} = 2\pi[c/2n_0d]$, then the resonant enhancement of the field pulls the cavity resonance toward the external frequency, which enhances the field more.

(a) Adapt (13.6.16) to fit this situation and show that

$$\dot{y} + \frac{1}{2}\left\{1 - j2Q\frac{(\delta - |y|^2)}{1 + |y|^2}\right\} y = \frac{u}{2}$$

where $y = E/E_s$, $\tau = t/\tau_p$, $E_s^2 = (2\eta_0 I_s)$, $u = \kappa\tau_p E_x/E_s$, $Q = \omega\tau_p$, $\delta = \omega_{m0}/\omega - 1$, and $\dot{y} = \partial/\partial\tau$.

(b) Explain why $\delta > 0$ (i.e., $\omega < \omega_{m0}$) to obtain bistability.

(c) Discuss the interplay between Q, δ, and the external field u to obtain bistability.

(d) Construct a graph similar to Fig. 13.4 for the $Q\delta = 2$ times a threshold value.

13.13. Show that the Hermite-Gaussian beam modes are suitable candidates for the Slater modes of the stable cavity shown in Fig. 5.1. What is an expression for the mode function $\mathbf{E}_{m,p,q}$ obeying

$(-\mathbf{a}_z) \times \mathbf{E}_{m,p,q}(z = 0) = (\mathbf{a}_z) \times \mathbf{E}_{m,p,q}(z = d) = 0$ and

$\iiint \mathbf{E}^*_{m,p,q} \cdot \mathbf{E}_{m',p',q'}\, dV = [\delta_{m,m'}][\delta_{p,p'}][\delta_{q,q'}]$

(HINT: Perform the integration over (dx, dy) first and use the orthogonality of the Hermite polynomials. See *Mathematical Handbook*, pp. 151–152. Use the facts that most optical cavities are many wavelengths long, $w(z)$ is a "slow" function of z, and $kz - (1 + m + p)\tan^{-1}(z/z_0) \sim kz$ to obtain an approximate evaluation of the integral along z.)

13.14. The purpose of this problem is to describe the time evolution of a CW laser as it builds up from that contributed by spontaneous emission. Assume a cavity in which the inversion was created instantaneously, the gain coefficient is above threshold by a factor of r, and it saturates according to the homogenous law.

(a) Use (13.6.18) to show that evolution of the normalized laser field, $y(t) = E(t)/E_s$, can be expressed by

$$\ln\left\{\left(\frac{y}{y_f}\right)^2\left(\frac{y_f}{y_0}\right)^2\left[\frac{1 - (y_0/y_f)^2}{1 - (y/y_f)^2}\right]^r\right\} = (r - 1)\frac{t}{\tau_p}$$

where E_s is saturation field defined by $E_s^2/2\eta_0 = I_s$, y_0 is initial value of the normalized field contributed by spontaneous emission, y_f is final or CW value of field ($y_f^2 = r - 1$), and $r = f\gamma_0(c/n)\tau_p$. (HINT: Separate the variables, use a partial fraction expansion, and integrate between the initial value y_0 and y.)

(b) Suppose $r = 3$ (i.e. the gain is three times threshold) and $y_0^2 = 10^{-6}$. How long does it take the laser to build up to 90% of its final value? (Ans.: $t/\tau_p = 10.2$.)

REFERENCES AND SUGGESTED READINGS

1. G. Herzberg, *Atomic Spectra and Atomic Structure*, 2nd ed. (New York: Dover, 1944).
2. A.C.G. Mitchell and M.W. Zemansky, *Resonance Radiation and Excited Atoms* (New York: Cambridge University Press, 1971), especially Chap. 3.
3. R.M. Eisberg, *Fundamentals of Modern Physics* (New York: John Wiley & Sons, 1961), Chap. 9.
4. E. Merzbacker, *Quantum Mechanics* (New York: John Wiley & Sons, 1971), Chaps. 19 and 20.
5. A. Yariv, *Quantum Electronics*, 3rd ed. (New York: John Wiley & Sons, 1989).
6. W.S. Chang, *Principles of Quantum Electronics* (Reading, Mass.: Addison-Wesley, 1969), Chaps. 5 and 6.
7. A. Maitland and M.H. Dunn, *Laser Physics* (Amsterdam: North-Holland, 1969), Chaps. 2 and 3.
8. M.O. Scully and M. Sargent, III, "The Concept of the Photon," *Phys. Today*, 38–47, Mar. 1972.
9. A. Matveyev, *Principles of Electrodynamics* (New York: Reinhold, 1966), Chap. 7.
10. G. Herzberg, *Spectra of Diatomic Molecules* (Princeton, N.J.: D. Van Nostrand, 1950).
11. A. van der Ziel, *Solid State Physical Electronics*, 2nd ed. (Englewood Cliffs, N.J.: Prentice Hall, 1968), Chap. 2.
12. R.L. White, *Basic Quantum Mechanics* (New York: McGraw-Hill, 1966), Chap. 11.
13. Willis E. Lamb, "Theory of Optical Maser," *Phys. Rev. A134*, A1429–1450, June 15, 1964.
14. See also some of the collected papers in *Laser Theory*, Ed. Frank S. Barnes (New York: IEEE Press, 1972). Part I, Historical Papers, is especially recommended.
15. M. Sargent, III, M. Scully, and W. Lamb, Jr., *Laser Physics* (Reading, Mass.: Addison-Wesley, 1974).
16. R.H. Pantell and H.E. Puthoff, *Fundamentals of Quantum Electronics* (New York: John Wiley & Sons, 1969).
17. R.P. Feynman, R.B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Vol. III (Reading, Mass.: Addison-Wesley, 1965).
18. U-Fano, "Description of States in Quantum Mechanics by Density Matrix and Operator Techniques," *Rev. Mod. Phys. 20*, 74–93, 1957.
19. R. Loudon, *The Quantum Theory of Light* (Oxford, England: Clarendon Press, 1983).
20. M. Born and E. Wolf, *Principles of Optics*, (Princeton, N.J.: Van Nostrand, 1950).
21. J.C. Slater, *Microwave Electronics* (Princeton, N.J.: Van Nostrand, 1950).
22. Treatments similar to Sec. 13.5 can be found in E. Fermi, *Rev. Mod. Phys. 4*, 131, 1932; and W. Heitler, *The Quantum Theory of Radiation* (New York: Oxford University Press, 1944).
23. D. Marcuse, *Engineering Quantum Electrodynamics* (New York: Harcourt, Brace, 1970).
24. Robert E. Collin, *Field Theory of Guided Waves* (New York: McGraw-Hill, 1960).
25. J.A. Stratton, *Electromagnetic Theory* (New York: McGraw-Hill, 1941).
26. L.D. Landau, E.M. Lifshitz, and L.P. Petaevskii, *Electrodynamics of Continuous Media*, (New York: Pergamon, 1984).



Quantum Theory of the Field-Atom Interaction

14.1 INTRODUCTION

All of the preceding chapters have relied on the Einstein A and B coefficients to describe the radiation-atom interaction along with a bit of sophistication introduced in the last chapter that focused on the field-atom interaction. Even there we had to resort to a bit of chicanery to relate the complex susceptibility back to the A coefficient. The goal of this chapter is to bypass such shenanigans and address the problem from first principles – quantum theory.

Unfortunately, this will cost us a significant investment in mathematics just to outline the *procedure* we must follow. Most of the time, we will not be able to complete some of the essential steps, and, under such circumstances, we will retreat to experiment (or the Einstein A coefficient) to identify the answer, say, for the dipole moment. If we accept this fact for a moment, then a natural inclination is to say, "Why bother?"

The reason for the ensuing onslaught is that the true dynamical behavior of the atom-field interaction can be much different from that predicted by the simple rate equations, but, for the majority of the common cases, the two approaches yield precisely the same answer. Hence, one goal will be to identify the circumstances under which the two approaches converge or diverge.

This begets the next reaction: "If the semiclassical quantum theory will handle most cases, why not use it to the exclusion of the rate equations?" The reason lies in the mathematical forest generated by the quantum approach. We can easily get lost among the trees

and forget the reason why we ventured into the forest. Besides, there are many important issues – such as the creation of the inversion – which are described to very good accuracy by the rate equations. Indeed some results are so accurate that we use them to describe the starting point of the quantum calculation. The bottom line is that both approaches need be addressed.

In this chapter, we will build up our level of sophistication:

1. We first address the calculation of the Einstein coefficients from Schrödinger's equation and find, much to our dismay, that we can compute the B coefficients, B_{12}, and B_{21}, but not A_{21}. However, since all coefficients are related [see (7.3.10)], we can still obtain an expression for A_{21}. (This is the one penalty we pay for not quantizing the field, but if, we did so, we would obtain the A coefficient directly.)
2. Then we will address the dynamics of the state occupation probability of an isolated atom. Here we discover a bit of deviation from the rate equation analysis of the same problem.
3. Then we will fold in the statistics of a collection of atoms interacting with the field; that is, the density matrix. By making reasonable approximations, we will find precisely the same answer as predicted by the rated equations. However, we can easily see how a higher-order analysis might yield results that cannot be predicted by the procedures of the prior chapters.
4. Finally, we will address examples of quantum phenomena that cannot be described by a rate equation approach (except in a phenomenological fashion). Some will show a dramatic difference between the semiclassical and rate equation approach.

14.2 SCHRÖDINGER DESCRIPTION

Originally, the Einstein coefficients were introduced in prior chapters to describe the dynamics of the populations N_2 and N_1 in response to a broadband spectrum energy density given by $\rho(\nu)$ [recall (7.3.4)]

$$\left.\frac{dN_2}{dt}\right|_{\text{rad}} = -A_{21}N_2 - B_{21}N_2\rho(\nu) + B_{12}N_1\rho(\nu)$$

and, since it must apply under *all* circumstances – even when $\rho(\nu)$ is the Planck blackbody value (7.2.10), the coefficients are related by

$$g_2B_{21} = g_1B_1 \qquad \text{and} \qquad \frac{A_{21}}{B_{21}} = \frac{8\pi n^2 n_g \nu^2}{c^3}h\nu \tag{7.3.10}$$

Thus, our task is to prepare our mathematical tools to compute one of the coefficients, and then the other two are supplied by (7.3.10). Since the coefficients are characteristic of the atom, under any and all circumstances, their value must be found from details of the atomic forces and the configuration of the electrons.

The allowed energy levels E_m and the wave functions Ψ_m of a state m are determined by a solution to Schrödinger's equation

$$\boldsymbol{H}_0\Psi = j\hbar\frac{\partial\Psi}{\partial t} \tag{14.2.1a}$$

where $\boldsymbol{H}_0$ is the operator associated with the total energy of the isolated atom, kinetic plus potential, and (14.2.1a) is usually written as

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V\right)\Psi = j\hbar\frac{\partial\Psi}{\partial t} \tag{14.2.1b}$$

Unfortunately, we cannot solve this equation for an arbitrary atom (except in the case of the hydrogen atom, not a molecule), but nature has no problem in doing so.* We fervently hope that there is a set of functions that do solve it and is suitability normalized and described by

$$\{\Psi_m(r, t)\} = \{u_m(r)\exp[-j(E_m/\hbar)t]\} \tag{14.2.2a}$$

where

$$\int \Psi_m^*\Psi_{m'}\, d^3x = \delta_{mm'} \tag{14.2.2b}$$

and

$$E_m = \text{energy of the } m \text{ state}$$

Obviously, it is the transitions occurring between two of the states, m and m', which are of interest producing or absorbing a photon of energy $\hbar\omega_{m'm} = E_{m'} - E_m$. Let us assume that someone else has fought the battle with (14.2.1) and (14.2.2) and provided us with the wavefunctions.

The Einstein B coefficient describes the response of the atom when there is an external field in the vicinity of the atom. Hence, we must return to the nearly impossible task of solving (14.2.1) and make it harder by adding the external field–electron(s) interaction.

$$[\boldsymbol{H}_0 + \boldsymbol{h}']\Psi = j\hbar\frac{\partial\Psi}{\partial t} \tag{14.2.3a}$$

where $\boldsymbol{h}'$ is the interaction energy between the active electron and the field. If we restrict our attention to only electric dipole interactions, then (14.2.3a) becomes

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V + v\right)\Psi = j\hbar\frac{\partial\Psi}{\partial t} \tag{14.2.3b}$$

where

$$v = e\phi \qquad \text{and} \qquad \phi = -\int_0^{\mathbf{x}} \mathbf{E}\cdot d\mathbf{l} \tag{14.2.3c}$$

The small letter for the perturbation is traditional and is intended to imply that $v \ll V$, a statement that was surely true before the invention of the laser, but a bit of caution is in order as the following example illustrates.

Consider the electric field experienced by an electron in the lowest orbit of the hydrogen atom with $r = a_0 = 0.526\text{Å}$. The field pulling the electron toward the nucleus is $e/4\pi\epsilon_0 a_0^2$ or

*We can get a good solution if we are willing to use all of the wizardry of orthogonal functions, use a large computer to invert a very large matrix, and to force fit the energy levels to experimental data.

5.2×10^{11} V/m. An electromagnetic wave having a peak field of (1/10) of that value would have an intensity of $E^2/2\eta_0 = 3.59 \times 10^{+18}\text{W}/m^2 = 3.59 \times 10^{+14}\text{W}/cm^2 = 359\,\text{TW}/cm^2$. Some current lasers can be focused to exceed that intensity, so a bit of caution is needed, but, for many cases, the perturbation assumption is quite safe.

We keep the words of caution in mind and proceed by expanding the wavefunction required for (14.2.3) in terms of the orthonormal set for the isolated atom wavefunctions (provided by someone else), a Fourier series as was done with the Slater mode expansion for the fields in a cavity.*

$$\Psi = \sum_m c_m(t) u_m(r) \exp[-j(E_m/\hbar)t \tag{14.2.4}$$

Thus, we hope to catch a snapshot of the atom as it makes a transition between a set of states with different quantum numbers m (a short-hand representation for combination of the principal quantum number n, angular l, magnetic m_s, and spin s). The quantities $c_m(t)$ are assumed to be time dependent and are related to occupation probability of state m.

$$c_m^*(t) \cdot c_m(t) = \text{probability of an atom being in state } m \tag{14.2.5}$$

For our simple rate equations involving a large number of atoms, we would use (14.2.5), multiplied by the total density of atoms, to describe the density in state m (i.e., N_m).

Now we substitute (14.2.4) into (14.2.3) and group terms in a very special fashion.

$$\sum_m \left\{ \left[\frac{-\hbar^2}{2m} \nabla^2 + V - E_m \right] u_m \right\} c_m(t) \exp[-j(E_m/\hbar)t]$$

$$+ \sum_m c_m(t) \left\{ \exp[-j(E_m/\hbar)t] \right\} [v(x,t)] u_m(x)$$

$$= j\hbar \sum_m \dot{c}_m(t) u_m(x) \exp[-j(E_m/\hbar)t] \tag{14.2.6}$$

The first brace in (14.2.6) is equal to zero because the wavefunctions $\{u_m\}$ are solutions to the time-independent Schrödinger's equation without the perturbation. Now we use the standard procedure of the Fourier series evaluation: multiply by the complex conjugate of a member of the assumed set, $u_{m'}^* \exp[+j(E_{m'}/\hbar)t]$, integrate over the domain of orthogonality, and use the fact that

$$\int u_{m'}^* u_m \, d^3x = \delta_{mm'}$$

to obtain:

$$\sum_m c_m(t) \exp[j\omega_{m',m}t] \int_{-\infty}^{+\infty} u_{m'}^*(x) v(x,t) u_m(x) \, d^3x = j\hbar \dot{c}_m \tag{14.2.7}$$

where

$$\hbar\omega_{m'm} = E_{m'} - E_m$$

*It is debatable which procedure came first, by whom, and in what context.

The summation on the right side of (14.2.6) disappeared because the integral is zero if $m' \neq m$ and 1 otherwise. Now we become very specific and restrict our attention to only electric field interactions to evaluate $v(x, t)$ from (14.2.3) with an assumed field $\mathbf{E} = E_{0x} \cos \omega t \, \mathbf{a}_x$.

$$\phi = -\int_0^{\mathbf{x}} \mathbf{E}_{0x} \cos \omega t \cdot d\mathbf{x} = -\frac{\mathbf{x} \cdot \mathbf{E}_{0x}}{2} [\exp(j\omega t) + \exp(-j\omega t)]$$

$$\therefore v(x, t) = e\phi = -\frac{(e\mathbf{x} \cdot \mathbf{E}_{0x})}{2} [\exp(j\omega t) + \exp(-j\omega t)] \tag{14.2.8}$$

The angular frequency ω of the electromagnetic wave need not be equal to a particular transition frequency although we would anticipate that it should be close for life to be interesting. We can now rewrite (14.2.7) by extracting the time dependence from the integral and abbreviate the spatial overlap integral by a matrix element $v_{m',m}$.

$$v_{m'm} \cos \omega t = [\int_{-\infty}^{+\infty} u_{m'}^*(x)(-e\mathbf{x})u_m(x) \, d^3x] \cdot \mathbf{E}_{0x} \cos \omega t \triangleq -\mu_{m'mx} E_{0x} \cos \omega t \tag{14.2.9a}$$

where $\mu_{m'mx}$ is the electric dipole moment for the transition between states m' and m.* We also assume $\mu_{mm} = 0$; that is, there is no permanent dipole moment (i.e., a spatial separation of the charge within an atom) associated with any state. While we have carefully used the complex conjugation symbol, the spatial part of the wave function is real, and thus we also see that $\mu_{m'm}$ is also real and therefore

$$\mu_{mm'} = \mu_{m'm} \tag{14.2.9b}$$

We break up the time harmonic cosine term into positive and negative frequencies, reverse the order of (14.2.7), and use the definition of the matrix element given in (14.2.9) to obtain

$$\dot{c}_{m'}(t) = -\left\{\sum_m \left[\frac{\mu_{m'm} E_{0x}}{2j\hbar} c_m(t)\right] \exp[j(\omega_{m'm} + \omega)t] + \left[\frac{\mu_{m'm} E_{0x}}{2j\hbar} c_m(t)\right] \exp[j(\omega_{m'm} - \omega)t]\right\} \tag{14.2.10}$$

Equation (14.2.10) is exact, but useless as it stands. We have to make some sort of an approximation to decouple this infinite set, and now we fold in a bit of physical insight. Suppose $\omega_{m'm}$ is positive corresponding to state m' higher in energy than m. Then the first term on the right side varies rapidly as $\exp[j(\omega_{m'm} + \omega)t$, whereas the other side varies as $\exp[+j(\omega_{m'm} - \omega)t]$. Only the term almost synchronized with the natural transition frequency will be slow enough to appreciably affect the left side or the population in state m'. Thus, we discard the antiresonant term but keep the one where $\omega_{m'm} - \omega$ is small. This type of an approximation has been used before in Chapters 12 and 13 and will be encountered again (and again) and is called the rotating wave approximation (RWA). By

*Unless explicitly stated, all dipole moments should have an x as a subscript to indicate a displacement along the direction of the inducing field. Most of the time, it will not be explicit so as to save in notation. We will also drop the vector relations between μ and $\mathbf{E}$.

this same line of reasoning, we need not consider all states but just those obeying (roughly) the Planck condition. Hence, we now focus on two states $m' = 2$ and 1, and our equations become (after remembering that μ_{mm} is assumed to be equal to 0):

$$\dot{c}_2(t) = j\Omega\, c_1(t) \exp[+j(\omega_{21} - \omega)t \tag{14.2.11a}$$

$$\dot{c}_1(t) = j\Omega\, c_2(t) \exp[-j(\omega_{21} - \omega)t] \tag{14.2.11b}$$

where

$$\boxed{\Omega \triangleq \frac{\mu_{21x} E_{0x}}{2\hbar} = \text{Rabi "flopping frequency"}} \tag{14.2.12}$$

The Rabi frequency is a very important parameter that plays a pivotal role in many laser considerations. If we multiply (14.2.12) by $\hbar$ and divide by the electronic charge e, a quantity with dimensions of energy (in volts) associated with the applied external field E_{0x} results.

$$\frac{\hbar\Omega}{e} = \frac{\mu_{21x} E_{0x}}{2e} = \frac{\langle x \rangle\, E_{0x}}{2} \text{ (volts)} \tag{14.2.13}$$

This form of the definition is a quick way of estimating the strength of the interaction in familiar units.

> If $\langle x \rangle \sim 0.1$ Å $= 10^{-9}$ cm and $E_0 = 10^4$ V/cm, then $\hbar\Omega/e = 5 \times 10^{-6}$ V or 5 μV, a value quite minuscule compared to the transition itself ~ 1 eV and is small compared to that contributed by the Doppler effect $\Delta\bar{\nu} \sim 0.1$ cm^{-1} $= 12.4\mu$V, but the 5μV is larger than that contributed by lifetime or natural broadening. As we will see, Ω will play the role of "broadening" the bandwidth over which the field can interact with the atom, and the format of (14.2.13) quickly estimates this broadening.

Equations (14.2.11a) and (14.2.11b) are the starting points for the next two sections.

14.3 DERIVATION OF THE EINSTEIN COEFFICIENTS

Let us examine (14.2.11a) and (14.2.11b) under some very restrictive but simplifying conditions. We presume that we have other information that suggests that the atom is in state 1 and thus

$$c_1(t = 0) = c_1(0) \sim 1 \qquad \text{and} \qquad c_2(t = 0) \sim 0 \tag{14.3.1}$$

at which time the electric field of the previous section is applied. For this section, we restrict our attention to time intervals short enough so that $c_1(t)$ can be replaced by 1 in (14.2.11a) yielding a simple integral for $c_2(t)$.

$$c_2(t) = \int_0^t \dot{c}_2(t')\, dt' = +j\Omega[c_1(0) \sim 1] \int_0^t \exp[j(\omega_{21} - \omega)t']\, dt'$$

where the initial condition $c_2(t = 0) = 0$ has been used for the lower limit on the left.

$$c_2(t) = +\, j\Omega c_1(0) \frac{\exp[j(\omega_{21} - \omega)t] - 1}{j(\omega_{21} - \omega)}$$

$$c_2(t) = j\Omega c_1(0)\left\{\exp\left[+j\frac{(\omega_{21}-\omega)t}{2}\right]\right\} \times \frac{\left\{\dfrac{\exp[+j(\omega-\omega_{21})t/2]-\exp[-j(\omega-\omega_{21})t/2]}{2j}\right\}}{\left(\dfrac{\omega_{21}-\omega}{2}\right)}$$

or

$$|c_2(t)|^2 = |c_1(0)|^2\Omega^2\left\{\frac{\sin \Delta\omega t/2}{\Delta\omega t/2}\right\}^2 t^2 \tag{14.3.2}$$

where

$$\Delta\omega = \omega_{21} - \omega$$

This result has some dismaying features and it is worthwhile to take a moment to identify the problem. It should be clear that $\Delta\omega$ has to be small; otherwise the sinc factor in the braces oscillates violently. If $\Delta\omega \to 0$, then the brace $\to 1$ and we find that the time dependence of the occupation probability of state 2 is proportional to

$$\boxed{|c_2(t)|^2 \sim |c_1(0)|^2 \cdot \Omega^2 \cdot t^2}$$

The first factor is quite logical and conforms to our intuition: If the electric field (with $\omega \sim \omega_{21}$) is to increase the population of state 2, there had to be population in state 1 to begin with and that population is just $|c_1(0)|^2$.

The second term, $\Omega^2 = [\mu_{21}E_{0x}/2\hbar]^2$ could be anticipated in hindsight (which is always perfect). There is always going to be some atomic parameter–here it is $(\mu_{21}/2\hbar)^2$–which is characteristic of the atom and the transition and which affects the rate at which state 2 is populated. The square of the electric field could be anticipated since the energy in the electromagnetic wave varies as E^2.

The problem arises with the t^2 variation.

Let us work the same problem with the rate equation and the Einstein coefficients assuming $N_1(t) \approx N_{10}$, $N_2(t) \sim 0$, $\omega \sim \omega_{21}$, and consider times short enough to justify neglecting all other processes exactly as was done for the above. From the definition of the Einstein coefficients (7.3.4), we would have

$$\frac{dN_2}{dt} \sim B_{12} \cdot N_1 \cdot g(\nu) \cdot [\rho_\nu \sim E_{0x}^2]$$

or

$$N_2 \sim N_1 \cdot B_{12} \cdot g(\nu) \cdot [\rho_\nu \sim E_{0x}^2] \cdot t \tag{14.3.3}$$

Notice that (14.3.3) predicts a linear increase with time in the population of state 2, whereas (14.3.2) predicts a quadratic variation. Is this the first instance of the rate equation yielding an incorrect answer? No. The rate equation result is the correct one, and (14.3.2) is an

example of paying too much attention to the mathematical formalities without checking whether the physics of the boundary conditions can hold.

We have assumed a perfectly sharp transition with $\omega \sim \omega_{21}$, whereas there is always some distribution of center frequencies associated with any collection of atoms. We know that the center frequency has to be somewhere, and thus we now re-encounter one of our old friends – the line shape – which describes the probability of the line center being at ν'_{21} along with the normalization condition

$$\int_0^\infty g(\nu'_{21})\, d\nu'_{21} = 1 \tag{14.3.4}$$

Thus, we return to (14.3.2), multiply it by the relative probability of ν'_{21} being the center frequency of the transition, and integrate over all possible choices, with the specified external frequency ν being a constant with respect to the variable of integration.

$$\therefore \qquad \left\langle |c_2(t)|^2 \right\rangle = |c_1(0)|^2 \Omega^2 \int_0^\infty \left[\frac{\sin \pi(\nu'_{21} - \nu)t}{\pi(\nu'_{21} - \nu)t} \right]^2 t^2 g(\nu'_{21})\, d\nu'_{21} \tag{14.3.5}$$

The $[\text{sinc}]^2$ function is such a rapidly varying function that the only contributing interval of the integrand is $\nu \sim \nu'_{21}$. Hence, we evaluate $g(\nu'_{21} = \nu)$, pull it out of the integral, and rewrite the remaining debris as

$$\left\langle |c_2(t)|^2 \right\rangle = |c_1(0)|^2 \Omega^2 g(\nu) t \left[\int_0^\infty \frac{1}{\pi} \left(\frac{\sin x}{x} \right)^2 dx \right] \tag{14.3.6}$$

where $x = \pi(\nu'_{21} - \nu)t$. The last bracket integrates to 1, and we obtain

$$\left\langle |c_2(t)|^2 \right\rangle = |c_1(0)|^2 \cdot \Omega^2 \cdot g(\nu) \cdot t \tag{14.3.7}$$

Thus, we see perfect agreement (so far) in functional form with the Einstein approach. If we let $N_1 = |c_1(0)|^2$ in (14.3.3), we now have an implicit formula for the Einstein coefficient.

$$|c_1(0)|^2 \cdot B_{12} \cdot g(\nu) \cdot t \cdot \rho_\nu = \frac{g_2}{g_1} |c_1(0)|^2 \cdot \Omega^2 \cdot g(\nu) \cdot t$$

$$(14.3.3) \qquad \Leftrightarrow \qquad (14.3.7)$$

or

$$B_{12} = \frac{g_2}{g_1} \frac{\Omega^2}{\rho_\nu} \tag{14.3.8}$$

The only task remaining is to convert (14.3.8) into one that is more easily interpreted. The energy density of an electromagnetic wave with a peak field of E_{0x} is

$$\rho_\nu = \frac{1}{2} \epsilon_r \cdot \epsilon_0 E_{0x}^2 \tag{14.3.9}$$

Now, $\Omega = (\mu_{21x} \cdot E_{0x})/2\hbar$ and μ_{21x}^2 can be related to the square of the total dipole moment describing the response of the atom to isotropic and unpolarized light.*

$$\mu_{21x}^2 = \frac{(\mu_{21})^2}{3} \tag{14.3.10}$$

*Now the subscript on the dipole moment is important and it is explicitly noted.

Thus
$$\boxed{B_{12} = \frac{2}{3}\frac{\pi^2}{\epsilon_r \cdot \epsilon_0}\frac{g_2}{g_1}\frac{\mu_{21}^2}{h^2} = \frac{g_2}{g_1}B_{21}} \tag{14.3.11}$$

where (g_2, g_1) are the degeneracies of the two states 2 and 1. The A coefficient is found from (7.3.10)

$$\frac{A_{21}}{B_{21}} = \frac{8\pi n^3 h\nu^3}{c^3}$$

leading to

$$\boxed{A_{21} = \frac{64\pi^4}{3\epsilon_r}\left(\frac{1}{4\pi\epsilon_0}\right)\frac{\nu^3}{(c/n)^3}\frac{\mu_{21}^2}{h}} \tag{14.3.12}$$

which is exactly four times the classical value found in (13.4.6).

14.4 DYNAMICS OF AN ISOLATED ATOM

Let us solve (14.2.11a) and (14.2.11b) simultaneously without making the assumption of a constant population in state 1. The two equations are repeated for convenience.

$$\dot{c}_2(t) = +j\Omega c_1(t)\exp[+j\Delta\omega t] \tag{14.4.1}$$

$$\dot{c}_1(t) = +j\Omega c_2(t)\exp[-j\Delta\omega t] \tag{14.4.2}$$

where
$$\Delta\omega = \omega_{21} - \omega \tag{14.4.3}$$

Differentiate (14.4.1) with respect to time, substitute (14.4.2) for $\dot{c}_1$, and obtain a single equation for $c_2(t)$.

$$\begin{aligned}\ddot{c}_2 &= +j\Omega\dot{c}_1\exp[+j\Delta\omega t] - \Omega\Delta\omega c_1\exp[+j\Delta\omega t] \\ &= -\Omega^2 c_2 - \Omega\Delta\omega\exp[j\Delta\omega t]\left(\frac{\dot{c}_2}{+j\Omega}\exp[-j\Delta\omega t]\right)\end{aligned}$$

$\therefore$
$$\ddot{c}_2 - j\Delta\omega\dot{c}_2 + \Omega^2 c_2 = 0 \tag{14.4.4}$$

Assume a time harmonic variation of the form $\exp[j\alpha t]$ and find that α must satisfy

$$-\alpha^2 + \Delta\omega\alpha + \Omega^2 = 0$$

or

$$\alpha_{2,1} = (\Delta\omega/2) \pm [(\Delta\omega/2)^2 + \Omega^2]^{1/2} \tag{14.4.5}$$

with solution 2 being associated with the positive sign and 1 with the negative choice.

The solution for $c_2(t)$ involves two arbitrary constants A_1 and A_2 to be determined from the initial conditions.

$$c_2(t) = A_2\exp[j\alpha_2 t] + A_1\exp[j\alpha_1 t] \tag{14.4.6}$$

At $t = 0$, let us choose $c_2 \sim 0$ – as we did in Sec. 14.3 – and find that $A_1 = -A_2$. We also use (14.4.1) to evaluate $\dot{c}_2$ at $t = 0$:

$$\dot{c}_2(t = 0) = +j\Omega[c_1(0) \sim 1] \Big|_{(14.4.1)} = j(\alpha_2 - \alpha_1)A_2$$

$$\therefore \qquad A_2 = \frac{\Omega}{\alpha_2 - \alpha_1}[c_1(0) \sim 1] \tag{14.4.7}$$

Thus,

$$c_2(t) = c_1(0) \cdot \frac{\Omega}{\alpha_2 - \alpha_1}\left\{\exp[j\alpha_2 t] - \exp[j\alpha_1 t]\right\} \tag{14.4.8a}$$

It is worthwhile to massage (14.4.8a) into a better format to aid in the interpretation of $|c_2(t)|^2$. Pull out a "common factor" of $\exp[j(\alpha_2 + \alpha_1)t/2]$ from the exponential terms and atone for this by including terms of the form $\exp[\pm j(\alpha_2 - \alpha_1)t/2]$ in the residue. We also include an extra factor of 2 and j inserted at a strategic point in the denominator.

$$\frac{c_2(t)}{c_1(0)} = j\Omega \exp[j(\alpha_2 + \alpha_1)t/2 \cdot \left\{ \frac{\exp[(j(\alpha_2 - \alpha_1)t/2] - \exp[-j(\alpha_2 - \alpha_1)t/2])}{\frac{2j}{(\alpha_2 - \alpha_1)}{2}} \right\} \tag{14.4.8b}$$

We insert the values of $\alpha_{2,1}$ from (14.4.5) and find the occupation probability of state 2 to be

$$|c_2(t)|^2 = |c_1(0)|^2 \cdot \frac{\Omega^2}{(\Delta\omega/2)^2 + \Omega^2} \cdot \sin^2\left\{[(\Delta\omega/2)^2 + \Omega^2]^{1/2}t\right\} \tag{14.4.9}$$

If we restrict our attention to short times, we recover (14.3.2) and re-encounter the t^2 problem, which can be avoided as we did before. Let us concentrate for now on some of the other issues contained in (14.4.9).

If, for instance, the detuning between the transition frequency ω_{21} and the external ω is zero, then the population in state 2 flops back and forth between 0 (our starting value) and 1 (indicating that the atom is completely in state 2) and then back to zero and so forth. It does so at a frequency Ω determined by the strength of the external field (and the other factors associated with Ω). If $\Delta\omega \neq 0$ and Ω is small, then there is an incomplete flop, with the maximum of $|c_2|^2$ equal to $\Omega^2/[(\Delta\omega/2)^2 + \Omega^2]$. These features are sketched in Fig. 14.1.

Equation (14.4.9) indicates phenomena not predicted by the rate equations:

1. If the external field is large enough such that $\Omega^2 \gg (\Delta\omega/2)^2$ (or the detuning is zero), then the population can be *completely inverted* (i.e., flipped), whereas the rate equations would indicate that the limiting value of $|c_2|^2$ (interpret as N_2) would be one half of the initial population in state 1.
2. The occupation probabilities continue to flop back and forth between the two states. While the flopping frequency is rather rapid, as the examples of the next paragraph show, there is no counterpart in the rate equations.

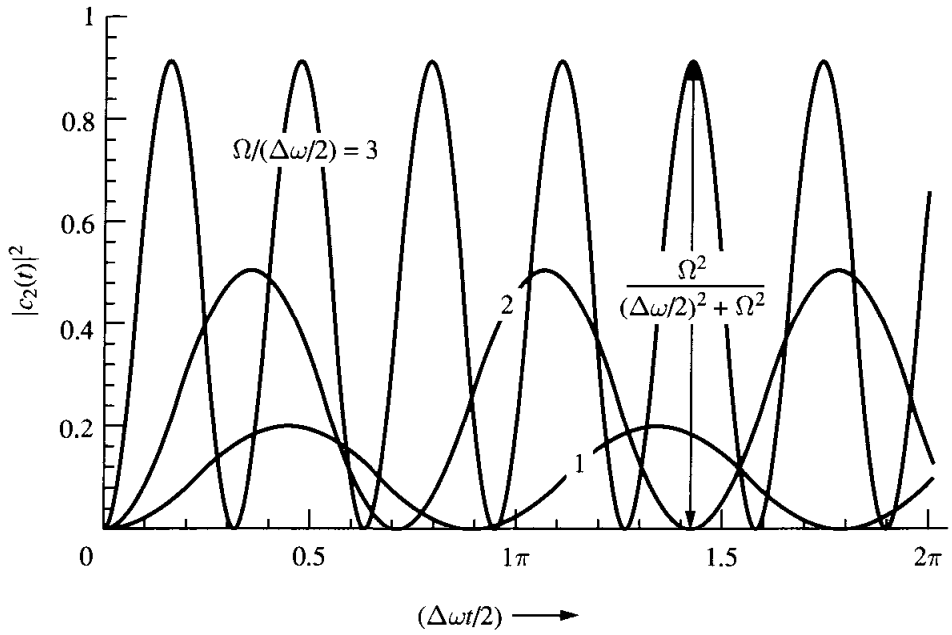


FIGURE 14.1. The occupation probability of state 2 as a function of time for a constant detuning and with increasing field strength.

3. If the field is strong enough such that $\Omega^2 \gg (\Delta\omega/2)^2$, then the complete time harmonic interchange takes place even though the photon energy $h\nu \neq E_2 - E_1$.

As we will see, the rate equations presume a time scale long compared to the flopping periods and/or elastic phase-interrupting collisions of the active atoms with something else. To handle that issue requires the density matrix of the next section.

Let us illustrate the fact that we can sample the transition over a bandwidth that is comparable if not larger than the width of the characteristic line shape $g(\nu'_{21})$, an effect referred to as power broadening.

Let us take a typical case to obtain a feeling for the extent of this power broadening. Let $\mu_{21x} = e\langle x\rangle = 1$ Debye$= 3.33 \times 10^{-30}$ C–m, and $\therefore \langle x\rangle = 2.08 \times 10^{-11}$ m $= 0.208$ Å, which is a big displacement of an electron from its equilibrium value. Let us pick a modest intensity of 100 W/cm^2 $= E_{0x}^2/2\eta_0$

$$\therefore E_{0x} = 275 \text{ V/cm} = 27.5 \text{ kV/m}$$

Hence

$$\frac{\Omega}{2\pi} = \frac{\mu_{21x} E_{0x}}{2h} = 69.1 \text{ MHz}$$

In other words, the field can cause transitions over a portion of the spectrum that is comparable to the homogeneous linewidth due to collisions. For this example, the "power broadening" is large compared to the "natural" broadening process (i.e., from spontaneous emission) but still small compared to a Doppler width for visible wavelengths for the numerical case considered. If we choose an intensity of 10 kW/cm^2, the power broadened

sampling FWHM width would be 1.38 GHz, a value comparable to the Doppler width of many transitions. It becomes very significant if we consider the intensity in the focal volume of a high peak power laser.

> Consider a 1 MW laser beam focused to a 1.0 μm spot size (area $= 10^{-8}$ cm^2), so that $E_0 = 2.75 \times 10^{10}$ V/m and $\Omega/2\pi = 69$ THz. If we compute an energy spread associated with this power broadening, $\hbar\Omega/e = \Delta E/e = \langle x \rangle E_{0x}/2 = 0.286$ eV. Thus, what started out to be a nice sharp transition to be interrogated by only those frequencies close to ν_{21} is now affected by radiation detuned from $\nu_{21} = \nu$ by as much as at 0.286 eV.

We should not be surprised, therefore, to find a host of "nonlinear" optical phenomena occurring with these intense fields. We need a related but different formalism to handle this case, the density matrix description is discussed in the next section.

If we include the decay of the populations to other states (by whatever process), then (14.4.1) and (14.4.2) become

$$\dot{c}_2(t) + \frac{1}{2\tau_2} c_2(t) = + j\Omega\, c_1(t) \exp[+j\Delta\omega t] \tag{14.4.10}$$

$$\dot{c}_1(t) + \frac{1}{2\tau_1} c_1(t) = + j\Omega\, c_2(t) \exp[-j\Delta\omega t] \tag{14.4.11}$$

A general solution is tedious, but if $\tau_1 = \tau_2 = \tau$, then a simple substitution

$$c_{2,1}(t) = c'_{2,1}(t) \exp[-(t/2\tau)] \tag{14.4.12}$$

reduces (14.4.10) and (14.4.11) to the same format as before. Hence the solution is just a slight modification of (14.4.9) to account for the decay.

$$|c_2(t)|^2 = |c_1(0)|^2 \cdot [\exp(-t/\tau)] \cdot \frac{\Omega^2}{(\Delta\omega/2)^2 + \Omega^2} \cdot \sin^2 \left\{ [(\Delta\omega/2)^2 + \Omega^2]^{1/2} t \right\} \tag{14.4.13}$$

14.5 DENSITY MATRIX APPROACH

14.5.1 Introduction

The previous two sections have hinted at a limitation of the rate equation approach to laser theory, but those limits have not been clearly identified. It is the purpose of this section to establish a formal procedure for coupling Maxwell's equations with quantum mechanics so as to provide a more complete and correct description of the laser.

We will show that our prior analysis of the earlier chapters is correct provided the field (or intensity) is not too large and we will be able to establish the limits. The following analysis will also point the way toward problems that cannot be addressed by the rate equations except in a rather phenomenological fashion. Unfortunately, this generalization will come at the expense of greatly expanded mathematical detail.

14.5.2 Definition

For the vast majority of problems, Maxwell's equations with classical fields will describe the laser-atom interaction quite adequately provided we compute the polarization current contributed by the atoms which in turn is induced by the field. Thus, the problem is to compute P_a and, as opposed to the classical approach of Chapter 13, we now use Schrödinger's equation for the dynamical behavior.

We again assume that the wave function of the atoms in the presence of the field can be expressed as a time dependent combination of the unperturbed eigenfunctions of Schrödinger's behavior.

$$\Psi(r, t) = \sum_n c_n(t) u_n(r) \tag{14.5.1a}$$

where $u_n(r)$ are the solutions to

$$\left[-\frac{\hbar}{2m}\nabla^2 + \mathrm{V}\right] u_n = E_n u_n \tag{14.5.1b}$$

Quite often we will need the expectation value of some operator $\mathbf{A}$ (which for our needs will be identified with the electric dipole moment), and it is computed by the following prescription

$$\langle \mathbf{A} \rangle = (\Psi, A\Psi) \triangleq \int \Psi^* \mathbf{A} \Psi \, d^3x \tag{14.5.2a}$$

a compact expression for a rather messy expression when written out in detail.

$$\begin{aligned} \langle \mathbf{A} \rangle &= \int \left\{ \sum_m c_m^*(t) u_m^*(r) \mathbf{A} \left[\sum_n c_n(t) u_n(r) \right] \right\} d^3x \\ &= \sum_m \sum_n c_m^*(t) \left[\int u_m^*(r) \mathbf{A} u_n(r) \, d^3x \right] \cdot c_n(t) \end{aligned} \tag{14.5.2b}$$

The quantity in the brackets is the matrix element A_{mn} of the operator $\mathbf{A}$, and we can compress (14.5.2) to a more manageable form:

$$\langle \mathbf{A} \rangle = \sum_m \sum_n c_m^*(t) A_{mn} c_n(t) \tag{14.5.2c}$$

We need to keep focused on the goal of computing the polarization of the atoms and for that goal $\mathbf{P}_a = N \langle \mu \rangle$ with $\mathbf{A} = \mu$, the electric dipole moment operator $\langle \mu \rangle = -e \langle \mathbf{x} \rangle$, but there are other possibilities of interest also. Because of these other cases, electric quadrupole, magnetic dipole, etc., we will continue to use the generic $\mathbf{A}$, until it is necessary to become specific.

The last form of (14.5.2) indicates that the bilinear products $c_m^*(t) \cdot c_n(t)$ control the time dependence of the operator $\mathbf{A}$. We also recall from Sec. 14.3 and 14.4 that there is always something different about each atom in our collection: It may be a host of different items, and thus we have to insert a probability for the occurrence of a particular case and

average over these possibilities

$$\overline{\langle A \rangle_s} = \sum_s \mathcal{P}_s \langle \mathbf{A} \rangle_s \tag{14.5.3a}$$

where the subscript denotes the class s to which the group belongs, and the bar indicates the average over the distribution.

This break up of the collection of atoms into various classes is an important step. As was mentioned, the probability of class s may be a classical variable, such as the velocity distribution, or it may be a detailed quantum statistic such as the phase of the wavefunction of one state compared to that of another state or atom. In this last instance, it is highly unlikely that sufficient information is available to specify these values, and it is virtually impossible to follow the dynamics of each atom. Hence, a statistical average is a necessity. We need not be specific as to the functional form of $\mathcal{P}_s$ at this stage, but we must recognize its existence.

Now we "tag" all atoms and prior results of this section with a pre-superscript (s), rewrite (14.5.2c) and (14.5.3a) and interchange the order of summations to obtain

$$\langle \mathbf{A} \rangle_s = \sum_m \sum_n \left\{ \sum_s \mathcal{P}_s [{}^s c_m^*(t)][{}^s c_n(t)] \right\} A_{mn} \tag{14.5.3b}$$

The quantity inside the braces is defined to be the density matrix element ρ_{nm} (note the reversal of the order of the subscripts).

$$\boxed{\rho_{nm} \triangleq \sum_s \mathcal{P}_s [{}^s c_m^*(t)][{}^s c_n(t)] = \left\langle [{}^s c_m^*(t)] \cdot [{}^s c_n(t)] \right\rangle} \tag{14.5.4}$$

We now have a compact notation for the expectation value of the operator $\mathbf{A}$ (and we will drop superfluous superscripts and the bar in the interest of simplicity for the future).

$$\langle \mathbf{A} \rangle = \sum_n \sum_m \rho_{nm} A_{mn} \tag{14.5.5}$$

This can be compacted even further by utilizing the rules of multiplication of two matrices C and D to obtain a particular element, say, the (j, k) one.

$$(CD)_{jk} = \sum_m C_{jm} D_{mk}$$

Thus, (14.5.5) becomes

$$\langle \mathbf{A} \rangle = \sum_n (\rho A)_{nn} \triangleq tr(\rho A) = \text{trace } (\rho A) \tag{14.5.6}$$

(i.e., the sum of the diagonal elements).

If there are N atoms in our collection, then $\rho_{nn} N$ is the number of atoms in state n. Since atoms have to be in some state, we also know that

$$\boxed{\sum_n \rho_{nn} = tr(\rho) \equiv 1} \tag{14.5.7}$$

The density matrix is Hermitian as is easily proven from its definition:

$$\rho_{nm} = \langle c_m^*(t) c_n(t) \rangle$$

that is, (14.5.4)

$$\therefore \quad \boxed{\rho_{mn} = \langle c_n^*(t) c_m(t) \rangle = \rho_{nm}^*} \tag{14.5.8}$$

So far we have merely introduced notation for a general operator **A**. Let us now become specific and focus on the electric dipole moment for a $2 \rightarrow 1$ transition according to the density matrix prescription.

$$\langle \mu \rangle = \sum_n \sum_m \rho_{nm} \mu_{mn}$$

$$= \rho_{11}\mu_{11} + \rho_{12}\mu_{21} + \rho_{21}\mu_{12} + \rho_{22}\mu_{22} \tag{14.5.9}$$

The first and last terms are zero since we are only considering those atoms without a permanent dipole moment. Furthermore, the spatial part of the wavefunctions used to compute μ_{21} and μ_{12} are real and thus so also are those quantities. Hence we have

$$\langle \mu \rangle = \mu_{21}(\rho_{12} + \rho_{21}) = \mu_{21}(\rho_{21} + \rho_{21}^*) \tag{14.5.10}$$

The last of (14.5.10) emphasizes the fact that the net dipole moment is real since μ_{21} is real and $(\rho_{21} + \rho_{21}^*)$ is always real whatever it might be. This should be comforting since we need a real quantity to insert into Maxwell's equations.

$$\boxed{P_a = N \langle \mu \rangle = N\mu_{21}(\rho_{21} + \rho_{21}^*)} \tag{14.5.11}$$

This completes the definition of the density matrix as we will need it for the next section, where "equations of motion" for ρ_{ij} are developed. However, before we leave these general matters and get buried in specifics, it is worthwhile to examine the issue of the distribution function of phases between states in greater detail to illustrate a very important point needed for the next section. We will consider two extreme cases so as to point out the contrast.

Example 1

Suppose we could prepare all atoms in the same fashion so that each state had a common phase origin. Thus, $\mathcal{P}_s$ can be expressed as a δ function around the phase angle θ_0

$$\mathcal{P}_s(\theta_s) = \delta(\theta_s - \theta_0) \qquad \text{and} \qquad \therefore \int_0^{2\pi} \mathcal{P}_s(\theta_s)\, d\theta_s = 1$$

For a single atom, the wavefunction given by $\Psi_m = u_m(r) \exp[-j(E_m/\hbar)t] \exp[-j\theta_m]$ is a perfectly acceptable solution to the time-dependent Schrödinger equation if the one without a phase factor solves it. Furthermore, there is nothing to say that the phase of state m must be the same as that of n. Rather than change all of our prior work, we include this possibility by including the extra phase angle factor as a multiplier times our original $c_m(t)$.

The summation required by (14.5.4) becomes an integral and is easily evaluated

$$\rho_{nm} = \int \delta(\theta_s - \theta_0) c_m^*(t) \exp[+j\theta_m] c_n(t) \exp[-j\theta_n]\, d\theta_s$$

$$= c_m^*(t)c_n(t) \int \delta(\theta_s - \theta_0) \exp[j(\theta_m - \theta_n)]\, d\theta_s$$

$$= c_m^*(t)c_n(t) \exp[j\theta_0] \tag{14.5.12}$$

(The phase angle θ_0 is actually a red herring since we can always choose the reference phase to be zero.) The diagonal terms ρ_{nn} are unaffected in any phase distribution since all θ's cancel and the integral of the δ function equals 1. This is logical since $N\rho_{nn} = N_n$, the density in state n, and densities (L^{-3}) do not have phases!

$$\rho_{nn} = |c_n(t)|^2 \tag{14.5.13}$$

There are really no surprises in this example.

Example 2

Now suppose that the phases of states of the atoms are uniformly distributed over 2π and hence

$$\mathcal{P}_s(\theta_s) = \frac{1}{2\pi} \qquad 0 < \theta_s \leq 2\pi$$

If we follow the prescription for diagonal elements, we again obtain (14.5.13), as expected, but a different result for the off-diagonal terms

$$\rho_{nm} = \frac{1}{2\pi} \int_0^{2\pi} c_m^*(t) \exp[+j\theta_m] c_n(t) \exp[-j\theta_n]\, d\theta_s$$

$$= c_m^*(t)c_n(t) \cdot \left\{ \frac{1}{2\pi} \int_0^{2\pi} \exp[j(\theta_m - \theta_n)]\, d\theta_s = 0 \text{ (i.e., nothing!)} \right\}$$

Thus, if the phases were distributed uniformly over 2π, the off-diagonal matrix elements are *zero* – always – and thus there is *no polarization* and *no* interaction with the electromagnetic wave. These two examples are extreme cases but they demonstrate an important fact

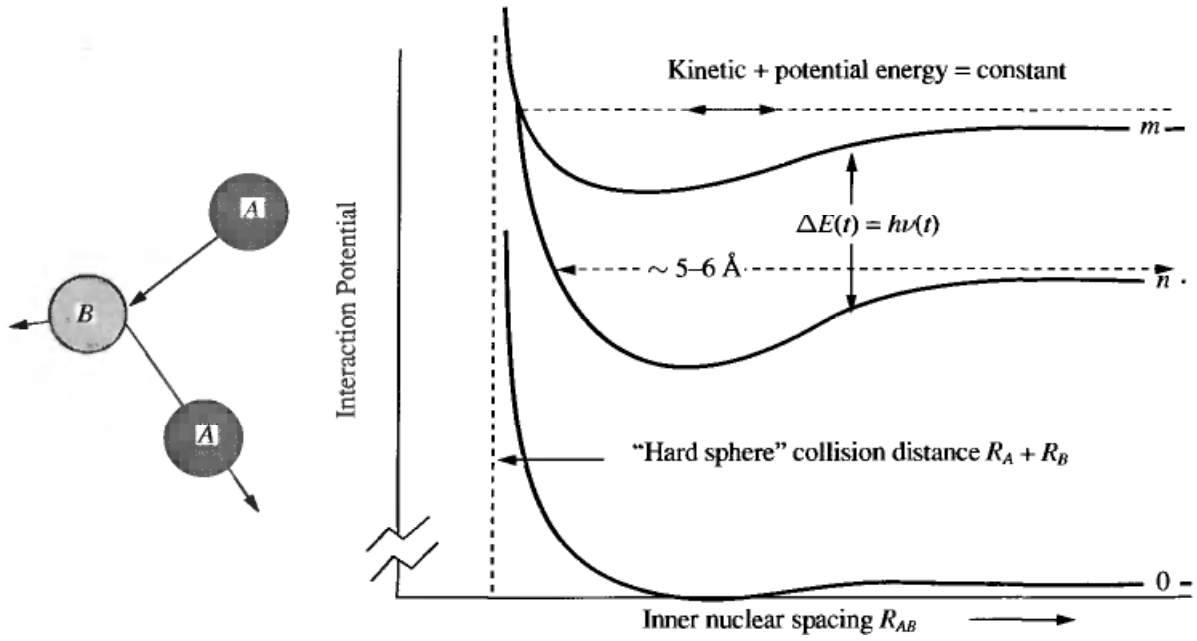
> Anything that randomizes the phase of state m with respect to n reduces ρ_{nm} and thus reduces the polarization.

This is something completely new and is not contained explicitly in the rate equations.

What is an example of this process? One example is an elastic collision between an atom in state (m or n) with another atom of the same kind or even of a different type. Usually we think of such a collision as being two billiard balls colliding. While that is acceptable in some cases, we should also recognize that they can interact over distances larger than the atomic size. Hence, there can be various forces between the two atoms represented by potential energy diagram along the trajectory of the collision as shown in Fig. 14.2.

The *elastic* collision process is indicated in Fig. 14.2 for case of an excited atom – say, A in state m – colliding with another one (of the same type or of a different variety) – call it B, going through all the trial and tribulations of the collision, and emerging with atom A still in state m.* We presume that atom A is excited to state m and can radiate to n but

*The collision sequence takes a very short time interval, say, ~ 10 Å divided by the relative velocity of 200 m/sec, or 5 ps. This is far shorter than the mean time it takes an atom to radiate, but it represents the time scale for the randomization of the phases between the two states.



(a) The collision (b) The interaction potential

FIGURE 14.2. (a) Two atoms colliding and preserving total energy and momentum. (b) The interaction potential for a head-on collision. For glancing collisions, add the centripetal potential to ensure conservation of angular momentum.

that the interaction potential between A and B is dependent upon the quantum numbers m and n (because of the way the excited electrons are arranged). The details of the collision are extremely complicated, but some general statements can be made and have been used to construct Fig. 14.2:

1. If the separation between A and B is large – say > 5 Å – then the two atoms ignore each other, and the normal energy level diagrams apply.
2. If the distance between the centers of A and B is sum of the radii of the two atoms, then the interaction potential becomes very large to prevent nuclear penetration.
3. In between those extremes, there may be some wondrous potential variations that are, in general, dependent on the excited state. Thus the energy difference between state m and state n is dependent upon the separation, and it, in turn, is dependent on time in the collision event. (Figure 14.2 was drawn assuming a significant binding energy between A and B in the excited states but a mostly repulsive potential for the 0 or ground state. This is similar to the excimer situation discussed in Chapter 10.)

To account for such collisions properly would take us far afield, but it surely should be palatable to accept the fact that the phase of m with respect to n after a collision might be anything whatsoever. Hence, only those atoms that have not collided can contribute to the polarization.

The effects of such collisions play a significant role in the dynamic behavior of the density matrix as discussed in the next section. As might be anticipated, it is virtually impossible to predict the collision rate. Rather, we must depend upon an experiment to provide the actual numbers.

14.6 EQUATION OF MOTION FOR THE DENSITY MATRIX

It is just a matter of patience with arithmetic and careful attention to the definitions to obtain an equation of motion for the density matrix. We recall that (14.5.4)

$$\rho_{nm} \triangleq \langle c_m^*(t)c_n(t)\rangle \tag{14.6.1}$$

Differentiate with respect to time, multiply by $j\hbar$, and drop the averaging symbol to save a bit on notation

$$j\hbar\frac{\partial\rho_{nm}}{\partial t} = c_m^*[+j\hbar\frac{\partial c_n}{\partial t}] - [-j\hbar\frac{\partial c_m^*}{\partial t}]c_n \tag{14.6.2}$$

To evaluate these partial derivatives, we recall the definition of the time dependent amplitudes

$$\Psi = \sum_k c_k(t)u_k(r) \tag{14.5.1}$$

which is to be a solution to the time dependent Schrödinger's equation.

$$j\hbar\frac{\partial\Psi}{\partial t} = \boldsymbol{H}\Psi \tag{14.6.3}$$

where $\boldsymbol{H}$ is the Hamiltonian operator and contains both the internal interactions (leading to $u_k(r)$ and E_k) and the external applied fields. Substitute (14.5.1) into (14.6.3) to find

$$j\hbar\sum_k\frac{\partial c_k(t)}{\partial t}u_k(r) = \boldsymbol{H}\sum_k c_k(t)u_k(r) = \sum_k(\boldsymbol{H}c_k)u_k + \sum_k c_k(\boldsymbol{H}u_k) \tag{14.6.4}$$

Now we employ the standard Fourier series technique: multiply by u_m^*, a member of the set used in (14.5.1), use the orthogonality of the set $\{u_m\}$ and also the definition of a matrix element of an operator (14.5.2) which is the Hamiltonian for the present case, and integrate.

$$j\hbar\sum_k\frac{\partial c_k(t)}{\partial t}\int u_m^*u_k\,d^3x = \sum_k[\boldsymbol{H}c_k(t)]\int u_m^*u_k\,d^3x + \sum_k c_k(t)(\int u_m^*\boldsymbol{H}u_k\,d^3x)$$

or

$$j\hbar\frac{\partial c_m(t)}{\partial t} = [\boldsymbol{H}c_m(t)] + \sum_k c_k(t)H_{mk} \tag{14.6.5}$$

where H_{mk} is the matrix element of the total energy operator and the square bracket represents a quantity that will disappear in a moment. Now, let $m = n$ so that we have the first

differential term in (14.6.2).

$$j\hbar \frac{\partial c_n(t)}{\partial t} = [\boldsymbol{H} c_n(t)] + \sum_k c_k(t) H_{nk} \tag{14.6.6a}$$

Now take the complex conjugate of (14.6.5), reverse the order of operator multiplication in response to the conjugation, and use the fact that $H^*_{mk} = H_{km}$ to obtain

$$-j\hbar \frac{\partial c^*_m(t)}{\partial t} = [c^*_m(t)\boldsymbol{H}] + \sum_k H_{km} c^*_k(t) \tag{14.6.6b}$$

which is the second bracket in (14.6.2). Substitute (14.6.6a) and (14.6.6b) for the square brackets in (14.6.2) and pay careful attention to the position of the multipliers:

$$j\hbar \frac{\partial \rho_{nm}}{\partial t} = c^*_m(t)\left\{[\boldsymbol{H} c_n(t)] + \sum_k c_k(t) H_{nk}\right\} - \left\{[c^*_m(t)\boldsymbol{H}] + \sum_k H_{km} c^*_k(t)\right\} c_n(t)$$

The square brackets involving the operator cancels leaving

$$\frac{\partial \rho_{nm}}{\partial t} = \frac{1}{j\hbar} \sum_k c^*_m c_k H_{nk} - c^*_k c_n H_{km} = \frac{1}{j\hbar} \sum_k (\rho_{km} H_{nk} - \rho_{nk} H_{km}) \tag{14.6.7}$$

where the definition of the density matrix (14.6.1) and the fact that it is Hermitian has been used. Some prefer to express (14.6.7) in commutator notation.

$$\frac{\partial \rho_{nm}}{\partial t} = \frac{1}{j\hbar} [\rho, \boldsymbol{H}]_{nm} \tag{14.6.8}$$

which is a convenient form for generalized manipulations, but we will stick with (14.6.7) for our analysis. It is the starting point for a formal description of the field-atom interaction, but we need to fold in a connection with a real life environment.

For instance, we know that the population $N_n = N\rho_{nn}$, in the n state will decay at a rate τ_n^{-1} if left unpumped, and we know that those states will be maintained at some equilibrium level, ρ^0_{nn}, by a pump in the absence of stimulated emission [which is hidden on the right side of (14.6.7)]. Hence, to include real-life effects in the analysis, we modify (14.6.7) for the diagonal elements ρ_{nn} to read

$$\boxed{\frac{\partial \rho_{nn}}{\partial t} + \frac{\rho_{nn}}{\tau_n} = + \frac{\rho^0_{nn}}{\tau_n} + \frac{1}{j\hbar} \sum_k (\rho_{kn} H_{nk} - \rho_{nk} H_{kn})} \tag{14.6.9}$$

which has the intuitive appeal of the correct limits: the correct steady state value of ρ^0_{nn} in the presence of pumping and absence of stimulated emission (the summation on the right side) and an exponential decay with time if all sources are removed. If the population of a higher level (m) decays into state n with a branching ratio ϕ_{mn}, then we would add another source on the right side of (14.6.9).

The off-diagonal elements of the density matrix are affected primarily by *elastic* collisions as discussed in the latter part of Section 14.6, although τ_n also contributes to a

dephasing of the wavefunctions.* We therefore include a different type of relaxation for the ρ_{nm} equation and use a different type of symbol for the rate of decay, T_{mn}^{-1}, which includes both effects.

$$\boxed{\frac{\partial \rho_{nm}}{\partial t} + \frac{\rho_{nm}}{T_{nm}} = \frac{1}{j\hbar} \sum_k (\rho_{km} H_{nk} - \rho_{nk} H_{km})} \tag{14.6.10}$$

Quire often T_{nm} is shortened to T_2. The next section will indicate the logic behind (14.6.10) in a very simple fashion.

14.7 TWO-LEVEL SYSTEM

Let us focus on the simplest case of an external time harmonic electric field interacting with our collection of atoms with the frequency ω being close to the transition frequency ω_{21} between two states, 2 and 1, and thus we ignore all other states. Let us rediscover the logic for the relaxation terms by starting with (14.6.7). The Hamiltonian consists of $\boldsymbol{H}_0$, representing the internal forces on the electron, and a perturbation, $\boldsymbol{h}' = -\mu \cdot \mathbf{E}(t)$, where $\boldsymbol{H}_0 \gg \boldsymbol{h}'$.

Now we carefully expand (14.6.7) and examine each term. For $n = 2, m = 1$, we have

$$\frac{\partial \rho_{21}}{\partial t} = \frac{1}{j\hbar} \left\{ (\rho_{11} H_{21} - \rho_{21} H_{11}) + (\rho_{21} H_{22} - \rho_{22} H_{21}) \right\} \tag{14.7.1}$$

Now let us examine the various terms. Recall that H_{11} would be given by

$$H_{11} = \int u_1^* (\boldsymbol{H}_0 + \boldsymbol{h}') u_1 \, d^3x \approx \int u_1^* \boldsymbol{H}_0 u_1 \, d^3x \tag{14.7.2}$$

if the induced energy $\boldsymbol{h}'$ is small compared to the internal value or if $\boldsymbol{h}'$ is such that its integral vanishes. But we also know that the set of eigenfunctions $\{u_n\}$ satisfy the time independent Schrödinger's equation

$$\boldsymbol{H}_0 u_n = E_n u_n \tag{14.7.3a}$$

Hence,

$$H_{11} = E_1 \quad \text{the energy of state 1} \tag{14.7.3b}$$

and

$$H_{22} = E_2 \quad \text{the energy of state 2} \tag{14.7.3c}$$

Finally, we address the quantity H_{21}

$$H_{21} = \int u_2^* (\boldsymbol{H}_0 + \boldsymbol{h}') u_1 \, d^3x = \int u_2^* [\boldsymbol{H}_0 u_1 = E_1 u_1] \, d^3x + \int u_2^* [\boldsymbol{h}' u_1] \, d^3x$$

$$= E_1 \int u_2^* u_1 \, d^3x - e\mathbf{E}(t) \cdot \int u_2^* \, \mathbf{x} \, u_1 \, d^3x$$

*Since T_2 collisions are elastic, they conserve energy and momentum and thus cannot affect the populations in the various levels.

The first integral is zero since the set $\{u_n\}$ is orthogonal. The last integral is the induced dipole moment matrix element μ_{21}, and hence,

$$H_{21} = -\mu_{21} \cdot \mathbf{E}(t) = H_{12} \tag{14.7.4}$$

where the last equality uses the same steps as the first one. Thus, (14.7.1) becomes more meaningful when expressed in terms of the energy levels and the dipole moment. Using $\hbar\omega_{21} = E_2 - E_1$, we have

$$\boxed{\frac{\partial \rho_{21}}{\partial t} = -j\frac{\mu_{21x}E_x(t)}{\hbar}(\rho_{22} - \rho_{11}) - j\omega_{21}\rho_{21}} \tag{14.7.5}$$

Now, let $n = 2, m = 2$ in (14.6.7)

$$\frac{\partial \rho_{22}}{\partial t} = \frac{1}{j\hbar}\underbrace{[(\rho_{12}H_{21} - \rho_{21}H_{12})}_{k=1} + \underbrace{(\rho_{22}H_{22} - \rho_{22}H_{22})]}_{k=2} \tag{14.7.6}$$

The last term is identically equal to zero. We have noted many times that $\boldsymbol{H}$ is Hermitian and the matrix elements are real. Hence, $H_{12} = H_{21} = -\mu_{21x}E_x(t)$ and so also is fact that $\rho_{mn} = \rho_{nm}^*$ [see (14.5.8)]. Hence, (14.7.6) becomes

$$\frac{\partial \rho_{22}}{\partial t} = -j\frac{\mu_{21x}E_x(t)}{\hbar}(\rho_{21} - \rho_{21}^*) \tag{14.7.7}$$

For $n = 1, m = 1$ in Eq. (14.6.7) we obtain (after paying careful attention to signs)

$$\frac{\partial \rho_{11}}{\partial t} = +j\frac{\mu_{21x}E_x(t)}{\hbar}(\rho_{21} - \rho_{21}^*) \tag{14.7.8}$$

A separate equation for ρ_{12} is not necessary since $\rho_{12} = \rho_{21}^*$.

It is worthwhile to examine these last equations in greater detail to make sure that the functional form makes sense from a physical point of view. Notice that a term $(\rho_{21} - \rho_{21}^*)$ appears in (14.7.7) and (14.7.8) and is obviously an imaginary quantity. But it is also multiplied by j, which makes the right side real, a most comforting fact since $N\rho_{nn} = N_n$, the density in the n state which is surely a real quantity.

It is worthwhile repeating our goal in case the preceding mathematics has obscured the reason for the exercise. We need to compute the polarization P_a to be inserted into Maxwell's equation (14.5.11)

$$P_{ax} = N\mu_{21x}(\rho_{21} + \rho_{21}^*) \qquad (14.5.11) \rightarrow (14.7.9)$$

where N, the density of active atoms is real, as is μ_{21x} and so also is the quantity $(\rho_{21} + \rho_{21}^*)$ and thus the product yields a real polarization.

Equations (14.7.5), (14.7.7) and (14.7.8) need to be modified to account for real-life effects as was mentioned in the previous section. The states are pumped by some external mechanism, and the densities in states 2 and 1 do decay, usually with a characteristic lifetime $\tau_{2,1}$ (unless state 1 is the ground state), and we modify (14.7.7) and (14.7.8) to account for that fact.

$$\frac{\partial \rho_{22}}{\partial t} + \frac{1}{\tau_2}\rho_{22} = -j\frac{\mu_{21x}E_x(t)}{\hbar}(\rho_{21} - \rho_{21}^*) + r_2 \tag{14.7.10a}$$

and

$$\frac{\partial \rho_{11}}{\partial t} + \frac{1}{\tau_1}\rho_{11} = +j\frac{\mu_{21x}E_x(t)}{\hbar}(\rho_{21} - \rho_{21}^*) + r_1 \tag{14.7.10b}$$

where $Nr_{2,1} = R_{2,1}$ is the pumping rate for state (2,1), the familiar friend of Chapter 8. Usually, we hide those terms by defining a steady state and small signal value of the diagonal elements (denoted by a superscript ^{0}) and thus $r_2 = \rho_{22}^0/\tau_2$ with identical manipulations applied to ρ_{11}.

Quite often, we subtract these two equations, do a large amount of arm waving (to be justified in a problem), and define a composite lifetime τ for the deviation of the inversion from its equilibrium value $(\rho_{22} - \rho_{11})^0$.

$$\boxed{\frac{\partial(\rho_{22} - \rho_{11})}{\partial t} + \frac{(\rho_{22} - \rho_{11}) - (\rho_{22} - \rho_{11})^0}{\tau} = -j\frac{2\mu_{21x}E_x(t)}{\hbar}(\rho_{21} - \rho_{21}^*)} \tag{14.7.11}$$

where $\Delta N^0 = N(\rho_{22} - \rho_{11})^0$ is the inversion density in the absence of any electric field. Such a quantity is computed from the elementary rate equations, and the definition of τ in terms of τ_2 and τ_1 is saved for a problem. (Notice that if $E_x = 0$, the inversion recovers to its equilibrium value with an initial time constant of τ, a helpful hint in establishing the value of τ.)

Finally we rewrite and modify (14.7.5) to account for these phase interrupting collisions discussed in Sec. 14.5. We *know* (or intuitively feel) that if the field is turned off, the induced polarization of the atoms will eventually become zero because of the randomization of the phases by these collisions; hence, we modify (14.7.5) to account for this "known" effect and assign a time constant T_2 to the decay of ρ_{21}.

$$\boxed{\frac{\partial \rho_{21}}{\partial t} + \left(\frac{1}{T_2} + j\omega_{21}\right)\rho_{21} = -j\frac{\mu_{21x}E_x(t)}{\hbar}(\rho_{22} - \rho_{11})} \tag{14.7.12}$$

The time constant T_2 should not be confused with τ_2 nor should it be associated with only state 2. It is the time constant for the randomization of the phase of the wave function of state 2 as defined by state 1 (or conversely) of the collection of atoms. It is, by far, the shortest time "constant", and is exceedingly difficult to compute. For a gas phase system, T_2 is roughly the mean time between collisions of the active atoms with the background atoms or molecules. Its presence can be inferred from experimental data, and thus we need only to acknowledge its existence.

We note that if $E_x(t)$ is suddenly clamped to zero in (14.7.12), the ρ_{21} decays with a time constant T_2.

$$\rho_{21}(t) = \rho_{21}(0)\exp[-t/T_2]\exp[-j\omega_{21}t] \tag{14.7.13}$$

and thus the polarization will also decay

$$P_{ax}(t) = e^{-t/T_2}\cdot[\rho_{21}(0)\exp[-j\omega_{21}t] + \rho_{21}^*(0)\exp[+j\omega_{21}t]) \tag{14.7.14}$$

Thus, after $t \gg T_2$, there is no polarization left to insert into Maxwell's equations in accordance with our intuition.

The functional form of (14.7.14) suggests that there will always be two time scales of interest: one describing the envelope of $\rho_{21}(t)$ and the other varying at the optical frequency rate. Hence, we will use that observation to simplify matters in the next section.

We can obtain a deceptively simple differential equation for the polarization by adding and subtracting the complex conjugate of (14.7.12) to itself, differentiating the sum, using the difference to express $(\rho_{21} - \rho_{21}^*)$ in terms of the sum, and identifying $P_{ax} = N\mu_{21x}(\rho_{21} + \rho_{21}^*)$ and $\Delta N = N(\rho_{22} - \rho_{11})$:

$$\frac{\partial^2 P_a}{\partial t^2} + \frac{2}{T_2} \cdot \frac{\partial P_a}{\partial t} + \left[\frac{1}{T_2^2} + \omega_{21}^2\right] P_a = -2\omega_{21} \frac{\mu_{21x}^2 E_x(t)}{\hbar} \Delta N \tag{14.7.15}$$

This should be a comforting equation since it has a firm quantum basis and appears to compute the polarization in terms of the electric field with a simple differential equation (similar to that of a series RLC circuit) thereby providing the information required by Maxwell's equations. Unfortunately, ΔN is also dependent upon $E_x(t)$ and $P_{ax}(t)$, and hence (14.7.15) does not completely specify P_a in terms of E. (It does for a "small" signal so that ΔN can be replaced by ΔN^0.)

Final closure is achieved by substituting $(\rho_{21} - \rho_{21}^*)$ into (14.7.11), multiplying by N, and again identifying the inversion densities $\Delta N = N(\rho_{22} - \rho_{11})$ and $\Delta N^0 = N(\rho_{22} - \rho_{11})^0$ with the following result:

$$\frac{\partial(\Delta N - \Delta N^0)}{\partial t} + \frac{(\Delta N - \Delta N^0)}{\tau} = \frac{2\mathbf{E}(t)}{\hbar\omega_{21}} \cdot \left[\frac{\partial \mathbf{P}_a(t)}{\partial t} + \frac{\mathbf{P}_a(t)}{T_2}\right] \tag{14.7.16}$$

where ΔN^0 was inserted inside the first time derivative (i.e., adding zero) to make the equation appear more symmetric. Equation (14.7.16) verifies a very important principle by using only a conservation of energy argument. The energy is either in the inversion or in the optical field. Hence if ΔN decreases below ΔN^0, the left and thus the right side of (14.7.16) must be negative, or

$$\boxed{\left\langle \mathbf{E}(t) \cdot \left\{ \frac{\partial \mathbf{P}_a(t)}{\partial t} + \frac{\mathbf{P}_a(t)}{T_2} \right\} \right\rangle \text{ must be } < 0 \text{ for amplification to take place.}} \tag{14.7.17}$$

The time average (angle bracket) is over a few optical cycles, which is much shorter than the time scale for changes in the inversion.

There are key trends expressed by the combination of (14.7.15) and (14.7.16) that are deserving of discussion.

1. If we start with the reasonable assumption of a simple electric field of the form, $Re[E_{0x}e^{j\omega t}]$, then the polarization will be complex: $P_{ax}(t) = Re[P_{0x}e^{j\omega t}]$ provided that the inversion ΔN is not a function of time.
2. However, (14.7.16) indicates that ΔN *is* a function of time, being driven by the product $\mathbf{E} \cdot \dot{\mathbf{P}}_\mathbf{a}(t)$. If (1) were correct, then this product contains terms at 0 frequency (i.e., DC) and 2ω, thereby proving that (1) is in error.
3. Once such nonlinearities appear, phasor notation is of limited utility, and we must ensure that E_x, P_{ax}, and ΔN are real functions of time.

4. The product of $E_x(t)$, as used in (1), and $\Delta N(t)$, as found from (2), will generate terms varying at ω and 3ω on the right side of (14.7.15). Thus the polarization will also vary at ω and 3ω as will E_x.
5. Once (4) is acknowledged, then (14.7.16) generates fluctuations at even harmonics in the density and odd ones for the polarization.

It should be clear from this "DO" loop that we should start with the following expansion:

$$E_x(t) = \frac{E_1 e^{j\omega t}}{2} + \frac{E_3 e^{j3\omega t}}{2} + [\text{odd harmonics}] \cdots + [\text{cc}]$$

where [cc] denotes the complex conjugate.

$$P_a(t) = \frac{P_1 e^{j\omega t}}{2} + \frac{P_3 e^{j3\omega t}}{2} + [\text{odd harmonics}] \cdots + [\text{cc}]$$

$$\Delta N(t) = \Delta N_0 + \frac{\Delta N_2 e^{j2\omega t}}{2} + \frac{\Delta N_4 e^{j4\omega t}}{2} + [\text{even harmonics}] \cdots + [\text{cc}] \quad (14.7.18)$$

where all coefficients are to be related in a self-consistent manner by the use of (14.7.15) and (14.7.16).

These higher-order harmonic terms are responsible for high intensity effects such as the AC Stark effect (the shift in the resonant frequency owing to the optical electric field), and three-photon absorption. While the assumption of a simple two-state system greatly restricts the utility of (14.7.15) and (14.7.16), the trends and logic path followed are quite general.

14.8 STEADY STATE POLARIZATION CURRENT

Let us now apply (14.7.11) and (14.7.12) to a two-level system exposed to a constant amplitude electromagnetic wave of frequency ω close to but not necessarily equal to the transition frequency ω_{21}.

$$E_x(t) = E_{0x} \cos \omega t = E_{0x} \frac{\exp[j\omega t] + \exp[-j\omega t]}{2} \quad (14.8.1)$$

If we substitute (14.8.1) into those equations, we see a degree of mathematical complexity that has not been encountered previously. For instance, the unknown time variation of $(\rho_{21} - \rho_{21}^*)$ multiplies $E_x(t)$ in (14.7.11) to yield $(\rho_{22} - \rho_{11})$ and its unknown time variation, which in turn multiplies $E_x(t)$ in (14.7.12). As mentioned at the end of section 14.7, the population difference equation will have a slow component (varying on the time scale of τ) plus terms varying at $\pm 2\omega$, $\pm 4\omega$, (i.e., even harmonics), whereas ρ_{21} will vary harmonically at odd multiples of ω (i.e., $\pm 3\omega$, $\pm 5\omega$, etc.). Thus, we employ the technique of taking a harmonic balance of these equations and, for now, keeping only the lowest-order terms.

The functional form of (14.7.14) indicates that there will be two time scales of interest: the time scale of the envelope of ρ_{21} and a time harmonic variation at the frequency of the

driving term, the electric field. Hence, we assume that format at the start by using a new variable σ_{21}:

$$\rho_{21}(t) = \sigma_{21}(t)\exp[-j\omega t] \tag{14.8.2}$$

$$\rho_{21}^*(t) = \sigma_{21}^*(t)\exp[+j\omega t]$$

We also discard the part of the electromagnetic wave that does not lead to a resonant behavior by invoking the rotating wave approximation (RWA) of the prior sections. Our equations of motion become

$$\left\{\frac{\partial\sigma_{21}}{\partial t} + \left[\frac{1}{T_2} + j(\omega_{21} - \omega)\right]\sigma_{21}\right\}\exp[-j\omega t] = -j\frac{\mu_{21x}E_{0x}\exp[-j\omega t]}{2\hbar}(\rho_{22} - \rho_{11}) \tag{14.8.3}$$

Cancel common $\exp[-j\omega t]$ terms and identify the Rabi flopping frequency, $\Omega = \mu_{21x}E_{0x}/2\hbar$ to obtain

$$\frac{\partial\sigma_{21}}{\partial t} + \left[\frac{1}{T_2} + j(\omega_{21} - \omega)\right]\sigma_{21} = -j\Omega(\rho_{22} - \rho_{11}) \tag{14.8.4}$$

The slowly varying component of the population difference equation is

$$\frac{\partial(\rho_{22} - \rho_{11})}{\partial t} + \frac{(\rho_{22} - \rho_{11}) - (\rho_{22} - \rho_{11})^0}{\tau} = -j2\Omega(\sigma_{21} - \sigma_{21}^*) \tag{14.8.5}$$

For a steady state field, the envelope σ_{21} is a constant, and the time derivative of σ_{21} is zero. From (14.8.4), we obtain

$$\sigma_{21} = \frac{-j\Omega(\rho_{22} - \rho_{11})}{(1/T_2) + j(\omega_{21} - \omega)} \tag{14.8.6a}$$

and

$$\sigma_{21}^* = \frac{+j\Omega(\rho_{22} - \rho_{11})}{(1/T_2) - j(\omega_{21} - \omega)} \tag{14.8.6b}$$

Thus, the difference $(\sigma_{21} - \sigma_{21}^*)$ is given by

$$(\sigma_{21} - \sigma_{21}^*) = -j2\Omega\frac{(\rho_{22} - \rho_{11})}{T_2}\left\{\frac{1}{(\omega_{21} - \omega)^2 + (1/T_2)^2}\right\} \tag{14.8.7}$$

We substitute this result into (14.8.5) (with $\partial/\partial t = 0$) to solve for the inversion in the presence of the field in terms of the equilibrium (or small signal) inversion.

$$(\rho_{22} - \rho_{11}) - (\rho_{22} - \rho_{11})^0 = -j2\Omega\tau(\sigma_{21} - \sigma_{21}^*)$$

$$= \frac{-4\Omega^2\tau}{T_2}(\rho_{22} - \rho_{11})\left[\frac{1}{(\omega_{21} - \omega)^2 + (1/T_2)^2}\right]$$

Solving for $(\rho_{22} - \rho_{11})$ yields

$$(\rho_{22} - \rho_{11}) = \frac{(\rho_{22} - \rho_{11})^0}{1 + 4\Omega^2\tau T_2 \left[\dfrac{(1/T_2)^2}{(\omega_{21} - \omega)^2 + (1/T_2)^2}\right]} \tag{14.8.8}$$

Before we insert this result back into our prior equations to obtain the polarization current, it is worthwhile to examine (14.8.8), recognize familiar friends, and make some abbreviations. If we properly define some of the variables and convert to frequency units, the quantity in the square bracket has the same functional form as a line shape $\bar{g}(\nu)$ normalized to be unity when $\omega = \omega_{21}$: Define

$$\boxed{\Delta\nu = \frac{1}{\pi T_2}} \tag{14.8.9a}$$

and thus

$$\boxed{\bar{g}(\nu) = \frac{(\Delta\nu/2)^2}{(\nu_{21} - \nu)^2 + (\Delta\nu/2)^2}} \tag{14.8.9b}$$

Then (14.8.8) can be written as

$$(\rho_{22} - \rho_{11}) = \frac{(\rho_{22} - \rho_{11})^0}{1 + 4\Omega^2\tau T_2\bar{g}(\nu)} \tag{14.8.10}$$

If we remember that $\Omega^2 \sim E_{0x}^2 \sim$ intensity, then the inversion varies as

$$\frac{1}{[1 + (I/I_s)\bar{g}(\nu)]}$$

just as was found from the rate equations in Chapter 8. We find the saturation intensity by a strategic insertion of η_0/η_0 into an expanded version of $4\Omega^2\tau T_2$, extracting the intensity $I = E_{0x}^2/2\eta_0$, and identify the residue to be the inverse of the saturation intensity, with the details saved for a problem.

Reinsert (14.8.10) into (14.8.6), include the time harmonic terms multiplying σ_{21} and σ_{21}^* to obtain ρ_{21} and ρ_{21}^*, and compute the polarization current P_{ax}.

$$\begin{aligned} P_{ax} &\triangleq N\mu_{21x}(\rho_{21} + \rho_{21}^*) = N\mu_{21x}(\sigma_{21}\exp[-j\omega t] + \sigma_{21}^*\exp[+j\omega t]) \\ &= N\mu_{21x}\left\{[2Re(\sigma_{21})]\cos\omega t + [2Im(\sigma_{21})]\sin\omega t\right\} \end{aligned} \tag{14.8.11}$$

which shows that an electric field $E_{0x}\cos\omega t$, generates both an in-phase and out-of-phase component to the polarization. If we use the phasor notation

$$P_{ax} = \epsilon_0(\chi_a' - j\chi_a'')E_{0x}\left\{\exp[j\omega t] = \cos\omega t + j\sin\omega t\right\} \tag{14.8.12}$$

and agree to take the real part, then what follow is a straightforward identification of the terms, χ_{ax}' and χ_{ax}'', from (14.8.11).

$$\epsilon_0\chi_{ax}'E_{0x} = N\mu_{21x}(2Re\sigma_{21}) \tag{14.8.13a}$$

$$\epsilon_0 \chi''_{ax} E_{0x} = N\mu_{21x}(2Im\sigma_{21}) \tag{14.8.13b}$$

After the electric field (hidden in Ω) is canceled, we obtain

$$\boxed{\chi'_{ax} = -\frac{\mu_{21x}^2 T_2}{\epsilon_0 \hbar} \frac{(\omega_{21} - \omega)T_2 \cdot \Delta N^0}{1 + 4\Omega^2 \tau T_2 + (\omega_{21} - \omega)^2 T_2^2}} \tag{14.8.14}$$

and

$$\boxed{\chi''_{ax} = -\frac{\mu_{21x}^2 T_2}{\epsilon_0 \hbar} \frac{\Delta N^0}{1 + 4\Omega^2 \tau T_2 + (\omega_{21} - \omega)^2 T_2^2}} \tag{14.8.15}$$

where $\Delta N^0 = (N_2 - N_1)^0$ in the absence of stimulated emission.

Now if we use the total dipole moment (for isotropic radiation) and substitute $\mu_{21x}^2 = (\mu_{21}^2)/3$ and use a fundamental electromagnetic expression relating the gain coefficient to χ''_a [recall (13.3.5)].

$$\frac{\gamma(\omega)}{k} = -\frac{\chi''_a}{n^2}$$

then

$$\boxed{\gamma(\omega) = \frac{\omega}{c}\frac{1}{n}\frac{\mu_{21}^2 T_2}{3\hbar\epsilon_0}\left[\frac{(N_2 - N_1)^0}{1 + 4\Omega^2\tau T_2 + (\omega_{21} - \omega)^2 T_2^2}\right]} \tag{14.8.16}$$

If we use the normalized line shape function (14.8.9b), then it appears even more familiar.

$$\boxed{\gamma(\nu) = \frac{\gamma_0(\nu_{21})\bar{g}(\nu)}{1 + (I/I_s)\cdot \bar{g}(\nu)}} \tag{14.8.17}$$

where $\gamma_0(\nu_{21})$ is the small signal gain coefficient at line center $\nu = \nu_{21}$, the normalized line shape function is given by (14.8.9b), and the saturation intensity is a collection of constants involving τ, T_2, μ_{21x}, and $\hbar$ (with the exact relation being reserved for a problem). This is precisely the same functional form found in Chapter 8 and used throughout the book, but now it is firmly backed by quantum theory.

Unfortunately, many of the coefficients appearing in the prior equations cannot be computed so we have to depend upon experiment to provide the values. We will find that all of our prior results from earlier chapters are precisely correct provided we interpret the line width $\Delta\nu = 1/(\pi T_2)$ (which is the experimental method of evaluating T_2) and also use the prior expression for the A coefficient (14.3.12).*

Inasmuch as the density matrix utilized considerably more arithmetic to arrive at an identical answer, we can legitimately ask the question of whether this trip was necessary. The answer is yes for a number of reasons:

*This assumes that the homogeneous linewidth is dominated by collisions or that the natural or lifetime broadening are small, a very safe assumption for most of laser phenomena.

1. It should give you confidence in the simple analysis of the earlier chapters.
2. The rate equation analysis has problems with "fast" phenomena as will be shown, but the above has shown it will predict the laser phenomena provided the temporal behavior is slow compared to T_2 or to the inverse of the Rabi frequency $[\Omega/2\pi]^{-1}$.
3. The density matrix formulation is the toolbox needed to address much more complicated problems.

The following sections will address a few problems that the rate equations cannot handle, but for now we should be happy with the agreement between the two approaches.

14.9 MULTILEVEL OR MULTIPHOTON PHENOMENA

The last section has demonstrated that the rate equation approach using the Einstein A and B coefficients yields precisely the same expression for the gain coefficient as does the full semiclassical quantum theory with the proviso that the time scale of the envelope of the electric field is long compared to T_2 and that line width $\Delta\nu_h$ is correctly interpreted. Even under this umbrella of conditions there are instances of important optical phenomena that the A and B coefficient approach will not address from first principles except in an after-the-fact fashion: define a coefficient describing the process and then use experiment to determine the numerical value of the coefficient. We will always obtain the right answer (by definition), but predicting it in advance and in the absence of an experiment is another matter.

A small sampling of three cases are shown in Fig. 14.3, and all involve more than two levels and/or multiphoton effects. The density matrix approach will address these problems from first principles at an expenditure of considerable mathematics.

Figure 14.3(a) is the so-called two-photon absorption. If the one-photon absorption between 2 and 1 is permitted (i.e., $\mu_{21} \neq 0$), then the two-photon process is *forbidden* if the states 1 and 2 have definite parity as is the case for an isolated atom. If we consider

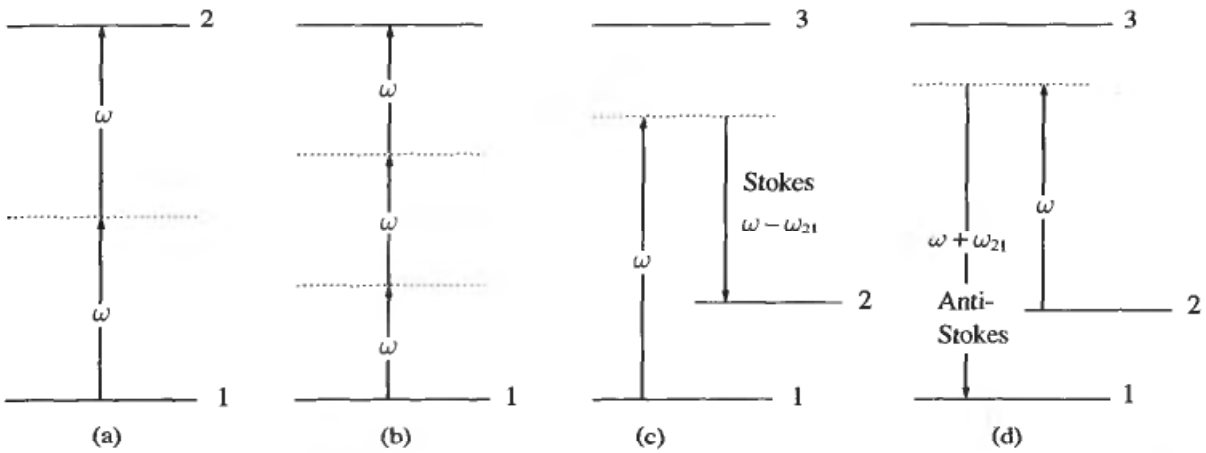


FIGURE 14.3. Various phenomena not handled by the A and B coefficients.

a higher-order interactions (magnetic dipole, electric quadrupole, etc.) such a process is allowed but is very weak.

Figure 14.3(b) involves the simultaneous absorption of three photons and exhibits a characteristic resonance when $3\omega \sim \omega_{21}$. It is the simplest case and is presented to illustrate the mathematical gyrations necessary to predict such phenomena accurately.

Figures 14.3(c) and 14.3(d) represent the Raman effect. We can show that if we apply a strong signal at ω in Fig. 14.3(c), we can generate a new frequency – either as a scattered signal or a stimulated one – at $(\omega - \omega_{21})$ with ω_{21} being the difference between the levels 2 and 1. The new frequency appears at the so-called "Stokes" shifted value $(\omega - \omega_{21})$. This is a very important technique for the study of molecular vibrations (corresponding to ω_{21}), especially if transitions between 2 and 1 are forbidden. If state 2 is reasonably populated (say, by thermal processes), then we can generate another frequency at $\omega + \omega_{21}$ (an "anti-Stokes" shift) by process shown in Fig. 14.3(d). Both Figs. 14.3(c) and 14.3(d) involve an absorption of one photon from a real state to a virtual one (shown dotted) followed by an emission back to a different real state. The Raman effect will be covered in the next section.

The major mathematical stumbling block appearing in the use of the density matrix is that equations are coupled with $E_x(t)$ multiplying ρ_{ij}, and thus harmonics and sum and difference frequency terms appear in ρ_{ij} and ρ_{jj} even though the driving term $E_x(t)$ may consist of only a simple sinusoid at one frequency. As will become apparent, the amplitude of these harmonics depend quite drastically on the amplitude of the electric field, and thus the harmonic balance by using the rotating wave approximation requires considerable mathematical insight to avoid multiple false starts. The three-photon case is addressed here because the complexity is tolerable and indicates some of the pitfalls to be avoided in the analysis of a multilevel system.

Example

Before we get too far into heavy mathematics, consider a one-photon process in a hypothetical gas with an allowed transition at $h\nu_{21}/e = 5$ eV or $\lambda_{21} = 2480$ Å, $g_2 = g_1$, with $T_2 = 10^{-10}$ sec, and thus $\Delta\nu_h = 1/\pi T_2 = 3.18$ GHz, and with $\mu_{21x} = 2$ Debye $= 6.6 \times 10^{-30}$ cm.* Hence the total dipole moment for isotropic light is $\mu_{21}^2 = 3\mu_{21x}^2$ and the A_{21} coefficient is given by

$$A_{21} = \frac{64\pi^4}{3}\left(\frac{1}{4\pi\epsilon_0}\right)\frac{\nu_{21}^3}{c^3}\frac{\mu_{21}^2}{h} = 2.42 \times 10^{+8}\ \text{sec}^{-1} \qquad \text{recall (14.3.11)}$$

Thus the one-photon absorption cross section at $\nu = \nu_{21}$ is

$$\sigma(\nu_{21}) = A_{21}\frac{\lambda^2}{8\pi}\left[\frac{2}{\pi\Delta\nu_h}\right] = 1.9 \times 10^{-12}\ \text{cm}^2$$

which is typical of strongly allowed transitions. At $\nu \sim \nu_{21}/3$, the one-photon absorption cross section is much smaller:

$$\sigma(\nu \sim \nu_{21}/3) = \left[A_{21} \cdot \frac{\lambda^2}{8\pi} \cdot \frac{2}{\pi\Delta\nu_h}\right] \cdot \frac{(\Delta\nu_h/2)^2}{(\nu_{21} - \nu_{21}/3)^2 + (\Delta\nu_h/2)^2} \qquad (14.9.1)$$

*The parameters are chosen to be close to that of mercury, but a real atom always has complicating features that obscures the point being addressed. Hence, consider the case discussed as an academic example.

$$\sigma(\nu \sim \nu_{21}/3) = \sigma(\nu_{21})\left[\frac{3\Delta\nu_h}{4\nu_{21}}\right]^2 = 4.63 \times 10^{-24}\ \text{cm}^2$$

which appears to be negligible insofar as attenuation on the signal at $\nu \sim \nu_{21}/3$ even with solid-like densities. However, if the photon beams are intense so that multiphoton processes are to be considered, then we should also compute the one-photon saturation energy/intensity at $\nu \sim \nu_{21}/3$, which indicates the value at which the populations $N_{1,2}$ are affected by the optical beam. For the values chosen

$$w_s = \frac{h(\nu \sim \nu_{21}/3)}{2\sigma(\nu \sim \nu_{21}/3)} = 2.88 \times 10^4\ \text{J/cm}^2$$

and

$$I_s = \frac{h\nu_{21}/3}{2\tau_{21}\sigma(\nu \sim \nu_{21}/3)} = 6.97\ \text{TW/cm}^2$$

Those values appear large, but if the optical beam were focused to a $(10\mu\text{m})^2$ area, then the total energy is only 28.8 mJ, a value well within the capabilities of table-top lasers, and we have yet to even consider the three-photon process that enhances the net absorption cross section by orders of magnitude.

When we allow the three-photon process, it takes much less energy or power at $\nu \sim \nu_{21}/3$ to deplete the ground state and populate state 2 because of the three-photon resonance. Once state 2 is populated to any degree, successive photons from the same beam can push the excitation farther up the quantum ladder toward the ionization level since the density of states always increases with energy. Thus it is possible to create significant ionization requiring 10–20 eV in a single step with large numbers of relatively modest photon energies of 1–2 eV.

This fact implies that a simple two-level model is limited to the intensity/energy range in which there is negligible change in the populations. The alternative is to use a multilevel model that is extremely complicated. The former is chosen here.

Consider Fig. 14.3(b) where the frequency of the electromagnetic wave is $\omega \sim \omega_{21}/3$, and thus it requires three absorbed photons to have an atom make a transition from 1 to 2. As discussed, the *absorption* cross section for a single photon process at ω is exceedingly small. If, however, the optical intensity or field strength is large enough (and how large comes later), we can absorb three photons simultaneously and extract much more energy from the electromagnetic wave then predicted by (14.9.1). As we will find, the intensity will obey the following generic equation:

$$\frac{dI}{dz} = -N_1[\sigma_1 I + \sigma_3 I^3] = -N_1\sigma_1 I\left\{1 + \left(\frac{\sigma_3}{\sigma_1}\right) I^2\right\} \tag{14.9.2}$$

where σ_1 is given by the above and $(\sigma_3 I^3)$ is the contribution owing to the three-photon process and N_1 is the density of atoms in state 1. The first form of (14.9.2) indicates, by the exponent on the intensity, the process being considered. The second form indicates the method used to report experimental measurements of the three-photon "cross section" with the dimensions of σ_3 (not to be confused with the amplitude of ρ at the third harmonic) being $\text{cm}^6\ \text{watts}^{-2}$. Our goal is to predict σ_3 to the same level of rigor used to predict σ_1, with the proviso that saturation does not occur.

We assume a two-level atom, which is a significant approximation since other states may become involved at high intensities as discussed earlier. The equations for the diagonal elements of the density matrix become [recall (14.7.10a) and (14.7.10b)

$$\frac{\partial \rho_{22}}{\partial t} + \frac{\rho_{22}}{\tau_2} = -j\frac{\mu_{21x}E_x}{\hbar}(\rho_{21} - \rho_{21}^*) \tag{14.9.3a}$$

$$\frac{\partial \rho_{11}}{\partial t} - \frac{\rho_{22}}{\tau_2} = +j\frac{\mu_{21x}E_x}{\hbar}(\rho_{21} - \rho_{21}^*) \tag{14.9.3b}$$

where it is assumed that all of the "natural" decay of 2 is back to 1, which, in turn, has no place to go since it is the ground state and its lifetime is infinite. Two equations are really not necessary since the sum of (14.9.3a) and (14.9.3b) indicates the obvious (but important) statement:

$$\left(\frac{\partial \rho_{22}}{\partial t} + \frac{\rho_{22}}{\tau_2} = -j\frac{\mu_{21x}E_x}{\hbar}(\rho_{21} - \rho_{21}^*)\right) + \left(\frac{\partial \rho_{11}}{\partial t} - \frac{\rho_{22}}{\tau_2} = +j\frac{\mu_{21x}E_x}{\hbar}(\rho_{21} - \rho_{21}^*)\right)$$

$$\rightarrow \frac{\partial}{\partial t}(\rho_{22} + \rho_{11}) \equiv 0$$

that is, $\rho_{22} + \rho_{11} = 1$ by (14.5.7). The dynamics of ρ_{21} is described by (14.7.12)

$$\frac{\partial \rho_{21}}{\partial t} + \left(\frac{1}{T_2} + j\omega_{21}\right)\rho_{21} = -j\frac{\mu_{21x}E_x}{\hbar}(\rho_{22} - \rho_{11}) \tag{14.9.4}$$

As noted in the previous section, (14.9.3) and (14.9.4) are coupled with products of harmonic components of ρ multiplying the field. We will use even harmonics for the diagonal elements and odd values for ρ_{21}, using the reasoning of Sec. 14.8. To make the arithmetic bearable, we assume steady state and a Fourier series in the following format.

$$\rho_{22} = \rho_b + \frac{\rho_2 e^{-j2\omega t} + \rho_2^* e^{+j2\omega t}}{2} + \frac{\rho_4 e^{-j4\omega t} + \rho_4^* e^{+j4\omega t}}{2} + \cdots \tag{14.9.5a}$$

$$\rho_{11} = \rho_a - \frac{\rho_2 e^{-j2\omega t} + \rho_2^* e^{+j2\omega t}}{2} - \frac{\rho_4 e^{-j4\omega t} + \rho_4^* e^{+j4\omega t}}{2} - \cdots \tag{14.9.5b}$$

where the facts that the populations must be real and the sum has to be constant has been utilized in formulating (14.9.5a) and (14.9.5b). The expansion for ρ_{21} is given by:*

$$\rho_{21} = \sigma_{11}e^{-j\omega t} + \sigma_{31}e^{-j3\omega t} + \sigma_{51}e^{-j5\omega t} + \cdots \tag{14.9.5c}$$

All of the Fourier components are "driven" by the electric field:

$$E_x = \frac{E_{1x}}{2}(e^{j\omega t} + e^{-j\omega t}) \tag{14.9.5d}$$

Implicit is the fervent hope that the two series converge, which in turn requires that the higher-order coefficients, ρ_{2n} and $\sigma_{(2n+1)1}$, should decrease if n is made large enough. We do not require the amplitudes to decrease uniformly. A particular resonant value (here σ_{31})

*The σ's in (14.9.5) are amplitude coefficients of ρ_{21}, not cross sections. To avoid such a possible misinterpretation, a double subscript will be employed with the first number being the harmonic order.

may be much larger than σ_{11}, but is required that $\sigma_{71} < \sigma_{51} < \sigma_{31}$. Substitute (14.9.5) into (14.9.4) and group according to $e^{-j\omega t}$, $e^{-j3\omega t}$, etc.

$$\begin{bmatrix} \left[\frac{1}{T_2} + j(\omega_{21} - \omega)\right]\sigma_{11}e^{-j\omega t} \\ +\left[\frac{1}{T_2} + j(\omega_{21} - 3\omega)\right]\sigma_{31}e^{-j3\omega t} \\ +\left[\frac{1}{T_2} + j(\omega_{21} - 5\omega)\right]\sigma_{51}e^{-j5\omega t} \\ +\text{etc.} \end{bmatrix} = -j\Omega\left\{e^{j\omega t} + e^{-j\omega t}\right\} \cdot \begin{bmatrix} \Delta\rho_0 = \rho_b - \rho_a \\ \rho_2 e^{-j2\omega t} + \rho_2^* e^{+j2\omega t} \\ +\rho_4 e^{-j4\omega t} + \rho_4^* e^{+j4\omega t} \\ +\text{etc.} \end{bmatrix} \tag{14.9.6}$$

where $\Omega = \mu_{21x}E_{1x}/2\hbar$ as usual.

We equate terms with common time variation by using the standard Fourier ploy of multiplying by the conjugate of one of the terms – say, by $e^{+jn\omega t}$ – and averaging over a few cycles. The only terms that survive are those containing the chosen harmonic $e^{-jn\omega t}$ (which is a mathematical justification of the rotating wave approximation). This ploy leads to

$$\left[\frac{1}{T_2} + j(\omega_{21} - \omega)\right]\sigma_{11} = -j\Omega[\Delta\rho_0 + \rho_2] \tag{14.9.7a}$$

where

$$\Delta\rho_0 = \rho_b - \rho_a$$

$$\left[\frac{1}{T_2} + j(\omega_{21} - 3\omega)\right]\sigma_{31} = -j\Omega(\rho_2 + \rho_4) \tag{14.9.7b}$$

$$\left[\frac{1}{T_2} + j(\omega_{21} - 5\omega)\right]\sigma_{51} = -j\Omega(\rho_4 + \rho_6) \tag{14.9.7c}$$

Now we substitute (14.9.5a) to (14.9.5d) into (14.9.3a) and go through the same harmonic balancing act:

$$\begin{bmatrix} \frac{\rho_b}{\tau_2} \\ +\frac{\rho_2}{2}\left[\frac{1}{\tau_2} - j2\omega\right]e^{-j2\omega t} \\ +\frac{\rho_4}{2}\left[\frac{1}{\tau_2} - j4\omega\right]e^{-j4\omega t} \\ +\text{etc.} \end{bmatrix} = -j\Omega\left\{e^{j\omega t} + e^{-j\omega t}\right\} \cdot \begin{bmatrix} \sigma_{11}e^{-j\omega t} - \sigma_{11}^* e^{j\omega t} \\ \sigma_{31}e^{-j3\omega t} - \sigma_{31}^* e^{j3\omega t} \\ \sigma_{51}e^{-j5\omega t} - \sigma_{51}^* e^{j5\omega t} \\ +\text{etc.} \end{bmatrix}$$

to obtain

$$e^{-j0t} \quad \rightarrow \quad \frac{\rho_b}{\tau_2} = -j\Omega(\sigma_{11} - \sigma_{11}^*) \tag{14.9.8a}$$

$$e^{-j2\omega t} \quad \rightarrow \quad \left[\frac{1}{\tau_2} - j2\omega\right]\frac{\rho_2}{2} = -j\Omega(\sigma_{11} + \sigma_{31}) \tag{14.9.8b}$$

$$e^{-j4\omega t} \rightarrow \left[\frac{1}{\tau_2} - j4\omega\right]\frac{\rho_4}{2} = -j\Omega(\sigma_{31} + \sigma_{51}) \tag{14.9.8c}$$

$$e^{-j6\omega t} \rightarrow \left[\frac{1}{\tau_2} - j6\omega\right]\frac{\rho_6}{2} = -j\Omega(\sigma_{51} + \sigma_{71}) \tag{14.9.8d}$$

Now surely the optical frequency terms $2\omega, 4\omega, \cdots$ are much greater than the rate $1/\tau_2$ for the population difference. Hence, (14.9.8b), (14.9.8c), and (14.9.8d) can be simplified.

$$\rho_b = -j\Omega\tau_2(\sigma_{11} - \sigma_{11}^*) \tag{14.9.9a}$$

$$\rho_2 = \frac{\Omega}{\omega}(\sigma_{11} + \sigma_{31}) \tag{14.9.9b}$$

$$\rho_4 = \frac{\Omega}{2\omega}(\sigma_{31} + \sigma_{51}) \tag{14.9.9c}$$

$$\rho_6 = \frac{\Omega}{3\omega}(\sigma_{51} + \sigma_{71}) \tag{14.9.9d}$$

Let us check on the convergence by assuming $\rho_6 \ll \rho_4$ and evaluating σ_{51}/σ_{31}. Substitute (14.9.9c) into (14.9.7c) and combine terms involving σ_{51}.

$$\left[\frac{1}{T_2} + j\left(\omega_{21} - 5\omega + \frac{\Omega^2}{2\omega}\right)\right]\sigma_{51} = -j\frac{\Omega^2}{2\omega}\sigma_{31} \tag{14.9.10a}$$

For a three-photon absorption, $3\omega \sim \omega_{21}$. If we use the example discussed earlier and choose a fairly large electric field such that $\hbar\Omega/e = 0.1$ V, $\hbar\omega_{21}/e = 5$ V, and $1/T_2 = 10^{10}$ sec$^{-1} \ll \omega_{21}$, then for these typical (if not extreme) values, the multiplier of σ_{51} can be approximated by $-j2\omega_{21}/3$ and we obtain

$$\sigma_{51} = \frac{9}{4}\left(\frac{\Omega}{\omega_{21}}\right)^2\sigma_{31} \sim \frac{9}{10{,}000}\sigma_{31} \tag{14.9.10b}$$

for the numerical values chosen. Hence, it is safe to neglect σ_{51} in comparison with σ_{31}.

Now we work backward through the equations. Substitute $\rho_4 = (\Omega/4\omega)\sigma_{31}$ into (14.9.7b) and collect terms involving σ_{31}.

$$\left[\frac{1}{T_2} + j\left(\omega_{21} - 3\omega + \frac{\Omega^2}{2\omega}\right)\right]\sigma_{31} = -j\Omega\rho_2 \tag{14.9.11a}$$

and use (14.9.9b) for ρ_2 to obtain

$$\left[\frac{1}{T_2} + j\left(\omega_{21} - 3\omega + \frac{3\Omega^2}{2\omega}\right)\right]\sigma_{31} = -j\frac{\Omega^2}{2\omega}\sigma_{11} \tag{14.9.11b}$$

Here we note a distinct possibility of a resonance, not exactly at the $3\omega = \omega_{21}$ value, but detuned slightly by a value proportional to the intensity (Ω^2). This is the Stark shift

mentioned in Sec. 14.8. When the imaginary part vanishes on the left side of (14.9.11b), σ_{31} will be a maximum and be related to σ_{11} by

$$\sigma_{31}(\max) = -j\frac{\Omega^2 T_2}{\omega}\sigma_{11} \rightarrow -j911.6\,\sigma_{11}$$

for the values used earlier. Thus, the amplitude of ρ_{21} at 3ω is nearly 1000 times larger than that at ω.

Now we substitute (14.9.9b) for ρ_2 into (14.9.7a).

$$\left[\frac{1}{T_2} + j(\omega_{21} - \omega)\right]\sigma_{11} = -j\Omega\left\{\Delta\rho_0 + \left[\rho_2 = \frac{\Omega}{\omega}(\sigma_{11} + \sigma_{31}\right]\right\}$$

or

$$\left[\frac{1}{T_2} + j\left(\omega_{21} - \omega + \frac{\Omega^2}{\omega}\right)\right]\sigma_{11} + j\frac{\Omega^2}{\omega}\sigma_{31} = -j\Omega\Delta\rho_0 \tag{14.9.11c}$$

The shear volume of symbols in our equations is becoming unbearable so make some logical abbreviations and reasonable approximations: Let $3\omega = \omega_{21}$, except where a small difference between the two arises, and assume that Ω^2/ω is small compared to ω_{21}. Define

$$\delta\omega = 3\omega - \omega_{21} \tag{14.9.12a}$$

a real frequency detuning from the three-photon resonance

$$\Delta\omega = \frac{3}{2}\frac{\Omega^2}{\omega} \approx \frac{9}{2}\frac{\Omega^2}{\omega_{21}} \tag{14.9.12b}$$

an intensity dependent frequency

$$D_1 = \left[\frac{1}{T_2} + j\left(\omega_{21} - \omega + \frac{\Omega^2}{\omega}\right)\right] \approx \left[\frac{1}{T_2} + j\frac{2}{3}\omega_{21}\right] \tag{14.9.12c}$$

$$D_3 = \left[\frac{1}{T_2} + j\left(\omega_{21} - 3\omega + \frac{3\Omega^2}{2\omega}\right)\right] \approx \left[\frac{1}{T_2} + j(\Delta\omega - \delta\omega)\right] \tag{14.9.12d}$$

We will recognize the D's as being the denominators of complex Lorentzians.

Use those definitions and (14.9.11b) for σ_{31} in (14.9.11c):

$$D_1\sigma_{11} + j\frac{2}{3}\Delta\omega\left[\sigma_{31} = -j\frac{2}{3}\frac{\Delta\omega}{D_3}\sigma_{11}\right] = -j\Omega\Delta\rho_0$$

or

$$\sigma_{11} = -j\Omega\Delta\rho_0\frac{D_3}{D_1 D_3 + (4/9)\Delta\omega^2} \tag{14.9.13}$$

Now, we should form the difference $\sigma_{11} - \sigma_{11}^*$ and insert that quantity into (14.9.9a) and solve for $\Delta\rho_0 = \rho_b - \rho_a$. The arithmetic mess is terrible, most difficult to interpret in its

complete form, and applicability of the result is questionable since other states higher than 2 have been neglected. Rather than do so, assume that $\rho_b \approx 0$ and $\rho_a \approx 1$, which amounts to neglecting any *saturation* of the population difference. This limits the applicability of the following to "small" intensities, with the upper limit on I to be determined.

To compute the polarization P_a, we reinsert the time harmonic variation and return to the expression for ρ_{21}

$$\begin{aligned} P_{ax} &= N\mu_{21x}(\rho_{21} + \rho_{21}^*) \\ &= N\mu_{21x}(\sigma_{11}e^{-j\omega t} + \sigma_{11}^* e^{+j\omega t} + \sigma_{31}e^{-j3\omega t} + \sigma_{31}^* e^{+j3\omega t} + \cdots) \end{aligned} \tag{14.9.14}$$

A bit of surprise appears. The atomic polarization has a real variation at 3ω (and higher odd harmonics) that might generate E_x at 3ω, even though we started with a field at ω. Thus, we have made a (slight) error in our formulation by not including that possibility and expressing the electric field as

$$E_x(t) = \frac{E_{1x}}{2}(e^{j\omega t} + e^{-j\omega t}) + \frac{E_{3x}}{2}(e^{j3\omega t} + e^{-j3\omega t}) + \cdots \tag{14.9.15}$$

As long as our attention is restricted to the *absorption* of the electromagnetic wave (at ω) and to the case where the field E_{3x} is very small compared to E_{1x}, then our formulation is accurate.*

Equation (14.8.12b) taught us that we will need to compute $2Im\sigma_{11}$ to obtain the absorption coefficient of the wave at the frequency ω (14.8.12b)

$$\epsilon_0 \chi''_{ax} E_{1x} = N\mu_{21x}(2Im\sigma_{11}) \qquad (14.8.3) \rightarrow (14.9.16)$$

where σ_{11} is given by (14.9.13). Within the context of the last paragraph, the 3ω terms in (14.9.4) do not affect the absorption of the field (at ω). The electric field E_{1x} on the left will be canceled by E_{1x} appearing in the Ω multiplier of (14.9.13). Thus, our current theory should yield both the one-photon (unsaturated) absorption coefficient of Sec. 14.8 as well as the three-photon contribution with the major limitation being the assumption of $\Delta\rho_0 = \rho_b - \rho_a \sim -1$.

$$\begin{aligned} 2Im\sigma_{11} &= -j(\sigma_{11} - \sigma_{11}^*) = -\Omega\Delta\rho_0 \\ &\times \left\{ \frac{|D_3|^2(D_1 + D_1^*) + (4/9)\Delta\omega^2(D_3 + D_3^*)}{|D_1|^2|D_3|^2 + (4/9)(\Delta\omega)^2[(D_1D_3 + D_1^*D_3^*) + (4/9)(\Delta\omega)^2]} \right\} \end{aligned} \tag{14.9.17}$$

We rewrite (14.9.17) using the fact that $D_1 + D_1^* = D_3 + D_3^* = 2/T_2$ from (14.9.12)

$$2Im\sigma_{11} = \frac{-2\Omega\Delta\rho_0}{|D_1|^2 T_2} \left\{ \frac{1 + \frac{4}{9}\frac{\Delta\omega^2}{|D_3|^2}}{1 + \frac{1}{|D_1|^2}\left[\frac{4}{9}\frac{\Delta\omega^2}{|D_3|^2}\right][D_1D_3 + D_1^*D_3^* + (4\Delta\omega^2/9)]} \right\} \tag{14.9.18}$$

*This does suggest that we could "triple" the frequency of a laser by making $\omega \sim \omega_{21}/3$. It works (with difficulty), but including that effect here would take us far from our current path.

If the intensity is small, then $\Delta\omega$, the intensity dependent frequency, is also very small and the brace $\rightarrow$ 1. Hence, the premultiplier contains the essential terms required to compute the one-photon absorption cross section as was done in Sec. 14.8 and yields the same answer [compare (14.9.18) with (14.8.7)] and thus need not be repeated here. The last brace in (14.9.18) is precisely the correction as expressed by the second form of (14.9.2), and we focus on that term after a bit more massaging of the expression.

Notice that the factor added to 1 in the numerator of (14.9.18) also appears as a bracket in the denominator, but it is also divided by $|D_1|^2 \sim 4\omega_{21}^2/9$ (a very large number), and thus the denominator in the brace of (14.9.18) is close to 1 except at extreme intensities where the use of a two-level model is questionable. Thus, we have a relatively simple expression for effective absorption cross section including σ_3, the three-photon process.

$$\sigma_{\text{eff}} I = \sigma_1 I \left\{1 + \frac{\sigma_3}{\sigma_1} I^2\right\} = \sigma_1 I \left\{1 + \left[\frac{4}{9} \cdot \frac{(\Delta\omega T_2)^2}{1 + (\Delta\omega T_2 - \delta\omega T_2)^2}\right]\right\} \tag{14.9.19}$$

Notice that the three-photon cross section is a product of the one-photon value (σ_1 evaluated at $\omega = \omega_{21}/3$) times a correction factor with a resonant rise or Lorentzian behavior around $\delta\omega = \Delta\omega(\sim E_x^2)$. The line width of this resonance is the same homogeneous value $\Delta\nu_h = 1/\pi T_2$. For any reasonable value of the intensity, the second term or the three-photon resonance dominates. At the peak, the cross section will vary as $\Delta\omega^2$ (i.e., I^2). The detuning from the exact three-photon value, $3\omega = \omega_{21}$, is real (a quadratic Stark shift in the resonance), but it is small for the range of applicability of the two-level model and tends to be obscured by other affects in an experiment. Thus we have achieved the goal of obtaining σ_3 based upon fundamental physical quantities, μ_{21}, and the electric field or intensity of the optical beam.

This theory is only valid if the energy or intensity is smaller than a value where significant changes in the population of state 1 occurs. Since we have dealt with a constant envelope, it is appropriate to use the effective saturation intensity given by

$$\frac{1}{I_{s3}} = \frac{\sigma_{\text{eff}}\tau_{21}}{(h\nu_{21}/3)} = I_{s1}^{-1}\left\{1 + \left[\frac{\hbar\Omega}{\hbar\omega_{21}}\right]^4\left[\frac{\nu_{21}}{\Delta\nu_h}\right]^2\right\} \tag{14.9.20}$$

where I_{s1} is the one-photon saturation intensity computed earlier (6.97 TW/cm^2). If we use $\hbar\Omega/e = 0.1$ eV and the other values chosen earlier, then the brace is 2.3×10^4 and thus the intensity should be less than the effective saturation intensity of 6.97 TW/cm$^2 \div 2.3 \times 10^4 =$ 301 MW/cm^2 for the theory to be applicable. This is not that large of an intensity, a fact that suggests that nonlinear effects can be observed with modest sized lasers.

14.10 RAMAN EFFECTS

14.10.1 Phenomena

Raman scattering is a very important spectroscopic tool for diagnostics purposes, and the stimulated Raman effect is a most useful practical technique for shifting the wavelength (or

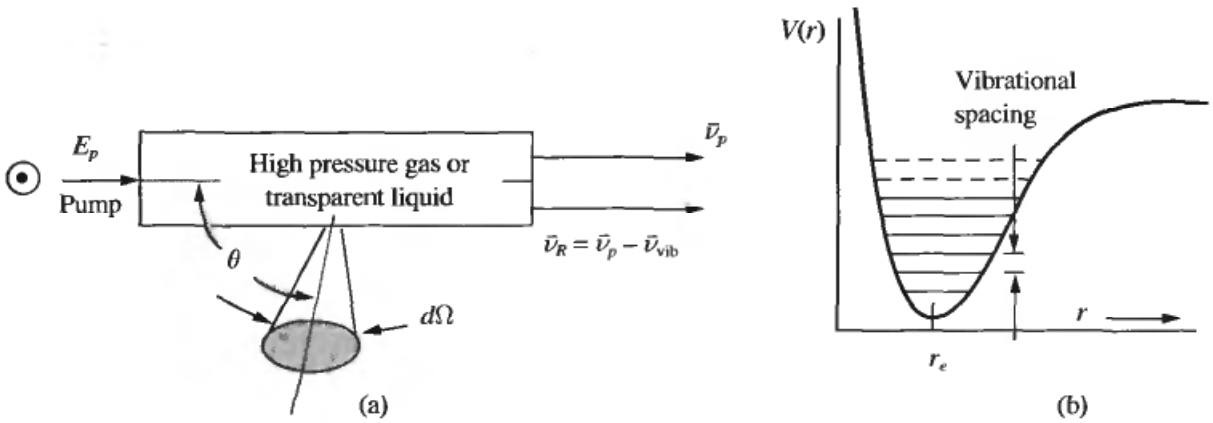


FIGURE 14.4. (a) The geometry used for the Raman effect. (b) The vibrational levels of a gas.

frequency) of pump laser. Figure 14.4 indicates a gas* (say, N_2 or H_2) being employed in the stimulated Raman effect to "shift" the frequency of the pump $\bar{\nu}_p$ by the vibrational spacing $\bar{\nu}_{vib}$ of the molecules.* Alternatively, we observe the scattering signal (usually at 90° with respect to the optical axis) to identify trace amounts of gas or other active modes in a solid. The amount of scattered signal is very small, usually less than 10^{-8} of the primary beam, but, with the advent of intense laser sources, is readily detected.

We use a relative intense source for the pump,† and detect a small fraction being scattered into a solid angle $d\Omega$ at θ (usually chosen to be 90°) at a frequency shifted by the vibrational spacing of the medium. If $\bar{\nu}_p$ is the wavenumber of the pump frequency, and $\bar{\nu}_{vib} = \bar{\nu}_{21}$ is that of the molecule, then the Raman signal appears at

$$\bar{\nu}_{R1} = \bar{\nu}_p - \bar{\nu}_{21} \tag{14.10.1a}$$

If the pump is strong enough, we can obtain gain at the optically shifted frequency given earlier. If the intensity of the first Stokes line is significant, it can generate a second Stoke line at

$$\bar{\nu}_{R2} = \bar{\nu}_{R1} - \bar{\nu}_{21} \tag{14.10.1b}$$

(A Stokes shift describes the situation where the emitted radiation is at a longer wavelength (lower frequency) than the pump, whereas anti-Stokes corresponds to the reverse.) If the population in state 2 becomes large enough, either by virtue of a thermal process or by a transfer to 2 from 1, then we can obtain an anti-Stokes line at

$$\bar{\nu}_{as} = \bar{\nu}_p + \bar{\nu}_{21} \tag{14.10.1c}$$

and so on.

*(A liquid or a solid can also be used).

*Although we will emphasize the vibrational modes of a gaseous molecule, since that is the most common situation, rotational levels and electronic levels as well as phonons in a solid can also be used.

†Raman used the strong UV lines from a mercury arc.

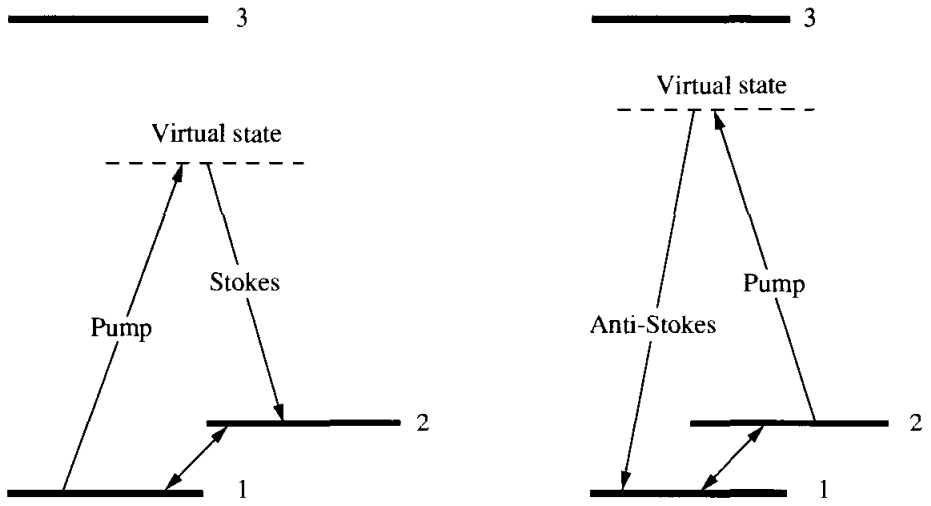


FIGURE 14.5. The generation of the Stokes and anti-Stokes lines by the Raman effect.

There are a couple of practical issues that make Raman scattering an extremely important spectroscopic tool for the study of vibrational modes of atoms in gases, liquids, and solids.

1. The Raman effect shifts the spectral region of interest from the inconvenient far infrared (FIR) region ($\bar{\nu} = 100\ \text{cm}^{-1}$ to $500\ \text{cm}^{-1}$) to the convenient visible or near UV region. In the FIR, detectors are slow, insensitive, usually cooled to LN_2 (77 K) or liquid He (4 K) temperature and therefore require a bulky dewar. Furthermore, some optical materials for windows and lenses are rather lossy by visible wavelength standards, are opaque and thus are difficult to align, or are hydroscopic and thus "degrade" in a humid atmosphere.

   In the visible portion of the spectrum, we have the advantages of quartz (or glass) as the nearly perfect optical material, which can be polished to a fantastic accuracy and which is nonhydroscopic, and very sensitive and fast photomultiplier detectors that can produce a measurable signal with as few as 5 to 10 photons. The penalty to be paid is that very high wavelength discrimination is required so that the small Raman signal can be observed close to the intense pump. Hence, double (or even triple) monochromators are employed.
2. Quite often, the vibrational modes are "inactive" in the FIR in the sense of having no allowed electric dipole moment (a forbidden transition) connecting the two states. Hence, even if we did put up with the aggravation of working in that part of the spectrum, some transitions would not appear in absorption. (The transition between adjacent vibrational states in N_2, H_2, and O_2 are examples.) However, these same modes are "Raman active" and show up clearly as either a scattered or stimulated radiation.
3. The frequency of the vibrational modes gives an unambiguous identification of the bonding partners and the type of bond. For instance, the Raman shift from graphic carbon is $1580\ \text{cm}^{-1}$ whereas that for diamond material is $1332\ \text{cm}^{-1}$ [22].

14.10.2 A Classical Analysis of the Raman Effect.

Consider Fig. 14.5 in which a strong pump, at ν_p, interrogates a medium with energy levels 1, 2, and 3 where electric dipole transitions are allowed between $1 \rightarrow 3$ and $2 \rightarrow 3$ but are forbidden between $1 \rightarrow 2$.

The Raman effect can be thought of as a transition from a real state (1 or 2) to virtual level (moderated by state 3) and then from that virtual state to a real state (2 or 1). The intensity of the scattered Stokes or anti-Stokes signal is proportional to the pump intensity, a Raman cross section (to be derived in the next section), and the population in the initial state. If $\Delta E_{21} \gg kT$, then almost all of the atoms reside in state 1, and hence the Stokes line is much stronger than the anti-Stokes line. If the virtual level is close to real state 3, then the cross section is greatly enhanced. Usually, $\Delta E_{31} \gg h\nu_p$, and thus state 3 can be considered the "rest" of the manifold of states in the atom or molecule.*

The quantum description is rather involved so it is worthwhile taking liberties with the physics of the atom with a combination of quantum theory and classical mechanical models to develop an intuitive feeling for the effect.

Consider a harmonic oscillator model of a molecule with $V(x) = k(x - x_e)^2/2$ describing the potential energy as one atom is displaced from its "equilibrium position" $r_e = x_e$ as is shown in Fig. 14.4(b). The allowed quantum energies are $\hbar\omega_e(v + 1/2)$, where v is the vibrational quantum number and only transitions with $\Delta v = \pm 1$ are permitted. A classical differential equation describing these facts can be "derived" by setting md^2x/dt^2 equal to the negative gradient of the potential energy and identifying ω_e^2 as the "spring constant," k, divided by the mass.

$$m\frac{d^2x}{dt^2} = -\nabla V(x) = -kx \quad \text{or} \quad \frac{d^2x}{dt^2} + \omega_e^2 x = 0 \tag{14.10.2a}$$

and if we include damping (due to whatever cause) we have

$$\frac{d^2x}{dt^2} + \frac{1}{T}\frac{dx}{dt} + \omega_e^2 x = 0 \tag{14.10.2b}$$

If a group of such vibrating molecules were interrogated with two electromagnetic waves at frequencies ω_1 and ω_2, then there is the possibility of a slight modification of potential $V(x)$ owing to the stored electric energy, $\mathbf{D} \cdot \mathbf{E}/2$, which may have a harmonic content close to ω_e. To illustrate, consider a group of molecules being irradiated by a single electromagnetic wave $\mathbf{E}\exp(j\omega t)$. The macroscopic displacement flux is given by

$$\mathbf{D} = \epsilon_0(1 + N\alpha)\mathbf{E} \tag{14.10.3a}$$

where α is the specific polarizability of each molecule and $(1 + N\alpha)$ is the relative dielectric constant of the gas. A strong field "moves" one atom with respect to the other and changes the specific polarization, which can be evaluated by expanding α in a Taylor series about its linear value.

$$\alpha = \alpha_0 + (\partial\alpha/\partial x)_0 \delta x$$

*Thus, we will *not* be able to select terms based upon a resonance involving $\nu_{3,1} \pm \nu$ or $\nu_{3,2} \pm \nu$ since there is little difference between any of those factors.

and

$$\mathbf{D} = \epsilon_0 \left\{1 + N[\alpha_0 + (\partial\alpha/\partial x)_0 \partial x]\right\} \mathbf{E} \tag{14.10.3b}$$

Thus the stored energy becomes

$$w_e = \frac{1}{2}\mathbf{D}\cdot\mathbf{E} = \frac{1}{2}\epsilon_0 \left\{1 + N[\alpha_0 + (\partial\alpha/\partial x)_0 \partial x]\mathbf{E}\right\}\cdot\mathbf{E} \tag{14.10.4}$$

where $$\mathbf{E} = \mathbf{E}_p \frac{e^{j\omega_p t} + e^{-j\omega_p t}}{2} + \mathbf{E}_R \frac{e^{+j\omega_R t} + e^{-j\omega_R t}}{2}$$

We add the excess electrostatic energy in (14.10.4) to $V(x)$ in (14.10.2a), differentiate the total with respect to δx to obtain the excess force (which eliminates the usual linear response terms), and again identify $\omega_e^2 = k/m$ to obtain an equation for δx

$$\frac{d^2\delta x}{dt^2} + \frac{1}{T}\frac{d\delta x}{dt} + \omega_e^2 \delta x = \frac{1}{2}\epsilon_0[(\partial\alpha/\partial x)_0]\frac{1}{4m}[\mathbf{E}_p \cdot \mathbf{E}_R^* e^{j(\omega_p - \omega_s)t} + \text{cc}] \tag{14.10.5}$$

where only the frequency component at $(\omega_p - \omega_R)$ is kept since this is the only term that might be resonant with natural vibrational frequency ω_e and thus cause a significant amplitude of the extra displacement δx. While we have implied that the two fields were applied externally, the analysis is applicable without change to the case where one of the fields – say, (E_p, ω_p), is applied – and the other, (E_R, ω_R), is generated internally and satisfies the frequency condition $\omega_s = \omega_p - \omega_R$.

Once (14.10.5) is accepted, the trip is all downhill to obtain a classical expression for the Raman gain coefficient. The essential steps are

1. Solve (14.10.5) for $\delta x(t) = [x_R e^{+j(\omega_p - \omega_s)t} + x_R^* e^{-j(\omega_p - \omega_R)t}]/2$ to find:

$$x_R = \frac{\epsilon_0(\partial\alpha\partial x)_0 \mathbf{E}_p \cdot \mathbf{E}_R^*}{4m\left\{[\omega_e^2 - (\omega_p - \omega_R)^2 + j(\omega_p - \omega_R)/T\right\}} \tag{14.10.6}$$

2. Compute the polarization *at* ω_s induced by the field *at* ω_p, which arises from the product of $[\partial\alpha/\partial x]_0 \delta x(t)]E(t)$ in

$$P(t) = \epsilon_0 N[\alpha_0 + (\partial\alpha/\partial x)_0 \delta x(t)]E(t) = \frac{(P_R(\omega_s)e^{+j(\omega_p - \omega_R)t} + P_R^*(\omega_s)e^{-j(\omega_p - \omega_R)t})}{2}$$

to obtain

$$P_R(\omega_s) = \epsilon_0 \left\{ \frac{N\epsilon_0(\partial\alpha/\partial x)^2 E_p^2}{8m\left\{[\omega_e^2 - (\omega_p - \omega_R)^2] - j[(\omega_p - \omega_R)/T]\right\}} \right\} E_R \tag{14.10.7}$$

3. This equation has a familiar complex form in terms of $\Delta\omega = \omega_p - \omega_s$:

$$\begin{aligned} P_R(\omega_1) &= \epsilon_0[\chi_R(\Delta\omega, E_p^2)]E_R(\omega_R) \\ &= \epsilon_0 \left\{[\chi_R'(\Delta\omega, E_p^2) - j\chi_R''(\Delta\omega, E_p^2)]\right\} E_R \end{aligned} \tag{14.10.8}$$

After making the usual high Q approximation, we obtain:

$$\chi'_R(\Delta\omega, E_p^2) = \frac{\epsilon_0 N(\partial\alpha/\partial x)_0^2[\omega_e - \Delta\omega]}{16m\omega_e\left\{[\omega_e - \Delta\omega]^2 + (1/2T)^2\right\}} E_p^2 \tag{14.10.9a}$$

and

$$\chi''_R(\Delta\omega, E_p^2) = -\frac{\epsilon_0 N(\partial\alpha/\partial x)_0^2(1/2T)}{16m\omega_e\left\{[\omega_e - \Delta\omega]^2 + (1/2T)\right\}} E_p^2 \tag{14.10.9b}$$

Notice that the imaginary part of the Raman susceptibility is negative indicating gain at ω_s.* By the simple application of (13.3.5) we have

$$\frac{1}{I_R}\frac{dI_R}{dz} \triangleq \gamma_k(\omega_s) \triangleq \frac{[k(\omega_R) = \omega_R n/c]N\epsilon_0(\partial\alpha/\partial x)_0^2(1/T)}{32n^2m\omega_e\left\{[\omega_e - \Delta\omega]^2 + (1/2T)^2\right\}} E_p^2 \triangleq N\left\{\sigma_R\left(\frac{E_p^2}{2\eta_0}\right)\right\} \tag{14.10.10}$$

which defines the peak Raman gain cross section when $\Delta\omega$ is set equal to ω_e. (NOTE: The dimensions of σ_R are area/(watt/area) or cm^4 sec/joule.)

This (mostly classical) phenomenological treatment exhibits much of the same functional dependencies found in the quantum treatment. For instance, notice that the Raman gain coefficient varies as the square of the electric field or the intensity of the pump, is linear in the density of molecules, and has a Lorentzian line shape as indicated by the quantity in the braces.

Unfortunately, (14.10.10) and the quantum counterpart to be derived in the next section (14.10.29) are of limited usefulness because of two factors:

1. Seldom do we know the atomic parameters well enough to evaluate the expressions. (How do we compute $(\partial\alpha/\partial x)_0$ for the classical model presented here or determine all of the dipole moments required for the next section?)
2. Usually the model is too simple. The classical version just presented is an example of a drastic simplification, but we are forced to limit the number of levels considered even in the quantum calculation.

Fortunately, we can use experimental results to bypass the uncertainties in any approach.

Most of the experimental work involved measuring the Raman "scattering" cross section using the geometry shown in Fig. 14.4(a) in which a fraction of the pump power scattered into a solid angle $d\Omega$ (determined by the collection optics) at an angle θ with respect to the incident wave (usually $\theta = 90°$ for experimental convenience), and into a given polarization is measured. For all isotropic substances, the scattering appears to originate as radiation from an electric dipole radiator induced by the electric field of the pump, and thus there is a body of knowledge on the differential scattering cross section (at

*Many define the Raman susceptibility as being the factor multiplying E_R^2, and thus χ would have a dimension of L^2/V^2 or L^2/watt if E_s^2 is expressed in term of the intensity.

λ_p). This experimental data can be used directly for diagnostic purposes and, with minimal assumptions, be adapted for the prediction of the Raman gain.

We first invoke conservation of photons by stating that the *loss* of photons of the pump per unit of distance dz owing to the Raman effect results in the appearance of the same number of photons at Stokes frequency $\nu_R = \nu_p - \nu_v$ into 4π steradians. Thus the loss of the pump caused by scattering is

$$\frac{1}{h\nu_p}\frac{\partial I_p}{\partial z} = \frac{\partial \Gamma_p}{\partial z}$$

$$= -\Gamma_p \int_0^{2\pi}\int_0^{\pi} N \frac{d\sigma_R(\theta,\phi)}{d\Omega}\,[d\Omega = \sin\theta\, d\theta\, d\phi] = -N\sigma_R\Gamma_p \qquad (14.10.11)$$

where $d\sigma_R/d\Omega$ is the Raman scattering cross-section. If we further assume the induced dipole mentioned earlier, then we can relate $d\sigma(\theta, \phi)/d\Omega$ to a value measured at one angle, usually at $\theta = 90°$.

$$\frac{d\sigma_R(\theta,\phi)}{d\Omega} = \left.\frac{d\sigma_R}{d\Omega}\right|_{90°} \cdot \sin^2\theta \qquad (14.10.12)$$

Thus the scattering cross section for the pump is found by combining (14.10.11) and (14.10.12).

$$\sigma_R = \left.\frac{8\pi}{3}\frac{d\sigma_R}{d\Omega}\right|_{90°} \qquad (14.10.13a)$$

where $(d\sigma_R/d\Omega)_{90°}$ has the dimensions of cm^2/(molecule-steradian). The scattered flux of photons at the Stokes frequency (per unit of volume) is

$$\frac{P(\text{scattered})}{h\nu_R\ (\text{vol})} = N\left\{\frac{8\pi}{3}\left[\frac{d\sigma_R}{d\Omega}\right]_{90°}\right\}\frac{I_p}{h\nu_p}$$

or

$$\frac{P(\text{scattered})}{(\text{vol})} = N\left\{\frac{\nu_R}{\nu_p}\frac{8\pi}{3}\left[\frac{d\sigma_R}{d\Omega}\right]_{90°}\right\} I_p \qquad (14.10.13b)$$

The scattering cross sections have been *measured* at various pump wavelengths for many substances with some common values given in Table 14.1. The cross section is usually specified by comparison to that of molecular nitrogen (N_2) for which $d\sigma/d\Omega = (3.31 \pm 1.1) \times 10^{-31}$ cm^2/ster mol at $\lambda = 4880$ Å. The cross section varies as λ_R^{-4}.

The loss of pump power by scattering is insignificant for laboratory size dimensions.

Example

For instance, consider the scattering by air (80% N_2 and 20% O_2) at STP (the sum of the densities $= 2.69 \times 10^{19}$ cm^{-3}) at 4880 Å. Using the data from Table 14.1a, the 90° scattering cross section $(d\sigma/d\Omega)_{90°}$ for $N_2 = 3.3 \times 10^{-31}$ cm^2/ster molecule($\bar{\nu}_v = 2331$ cm^{-1}) and that for $O_2 = 1.3 \times 3.3 \times 10^{-31} = 4.29 \times 10^{-31}$ cm^2/ster molecule($\bar{\nu}_v = 1556$ cm^{-1}).

TABLE 14.1a Data on Raman Scattering in Gases

Gas	Shift (cm^{-1})	Cross section ($d\sigma/d\Omega$) relative to N_2 (at 90°)
N_2	2331	1.0×
O_2	1556	1.3×
H_2	4161	2.1×
CO	2145	1.0×
NO	1877	0.27×
CO_2	1388	1.4×
	1286	0.89×
SO_2	1151	5.2×
	519	0.12×
H_2S	2611	6.4×
NH_3	3334	5.0×
ND_3	2420	3.0×
CH_4	2914	6.0×
C_2H_6	993	1.6×
C_6H_6	3062	7.0×
	992	9.1×

Data from Fenner et al. [20].

Thus the attenuation due to N_2: $[\mathbf{N}_2] = 0.8 \times 2.69 \times 10^{+19} = 2.152 \times 10^{+19}$ cm^{-3} $\therefore$ $[N_2]\sigma_R = 7.1 \times 10^{-12}$ cm^{-1} and that for oxygen is $[O_2]\sigma_R = 2.3 \times 10^{-12}$ cm^{-1} for a sum of

$$\therefore \Sigma N\sigma = 9.42 \times 10^{-12} \text{ cm}^{-1}$$

which represent the attenuation by dust-free air by Raman scattering, completely ignorable if we are only concerned with the pump.* However, if the argon ion laser at 4880 Å($\bar{\nu}_v$ = 20, 492 cm^{-1}) carried 10 watts of power, then 70 pW of power is scattered into 4π at λ = 5506.4 Å due to nitrogen) for every centimeter of path. This sounds small, but that power amounts to $1.75 \times 10^{+8}$ photons at 5506.4 Å. With photomultipliers capable of detecting as few as 10 or so photons, we have considerable leeway in reducing the solid angle of the collecting optics and allowing for spectral discrimination. This last is most important since the "dust," even in a clean room, will contribute a much larger number of scattered photons but at λ = 4880 Å. The above numbers indicate that we need a wavelength discrimination of $1 : 10^{11}$ at least.

The coherent Raman gain can also be predicted from these experimental numbers by recalling a very important basic relationship good for all radiative processes.

$$\frac{\text{stimulated rate (vol.)}^{-1} \text{ at } \nu_R}{\text{spontaneous rate (vol.)}^{-1} \text{ at } \nu_R} = \frac{g(\nu_R)I_R/(c/n_g)}{8\pi n^2 n_g h\nu_R^3/c^3} = \left\{ \frac{g(\nu_R)c^2}{8\pi n^2 h\nu_R^3} \right\} I_R \quad (14.10.14)$$

a statement that can immediately be verified by returning to the Einstein A and B coefficients description of a simple laser. Thus the gain coefficient for stimulated emission at the Stokes

*The blue sky is a result of the preferential scattering (λ^{-4}) of the blue photons coming from the sun (Rayleigh scattering).

TABLE 14.1b Data on Raman Scattering (Liquids and Solids)

Material	Shift (cm^{-1})	Line Width (cm^{-1})	Cross section ($\lambda_0 = 6943$ Å) $N(d\sigma/d\Omega) \times 10^8$	Gain Coefficient $\gamma_R \times 10^3$ (cm^{-1}) ÷ (MW/cm^2)
Liquids				
O_2	1552	0.117	0.48 ± 0.14	14.5 ± 4
N_2	2326.5	0.067	0.29 ± 0.09	17 ± 5
Benzene (C_2H_6)	992	2.15	3.06	2.8
CS_2	655.6	0.5	7.55	24
Nitrobenzene ($C_6H_5NO_2$)	1345	6.6	6.4	2.1
Bromobenzene (C_6H_5Br)	1000	1.9	1.5	1.5
Chlorobenzene (C_6H_5Cl)	1002	1.6	1.5	1.9
Toluene ($C_6H_5CH_3$)	1003	1.94	1.1	1.2
Solids				
Li^7NbO_3	256	23	381	8.9
	258	7	262	28.7
	637	20	231	9.4
	643	16	231	12.6
Li^7TaO_3	201	22	238	4.4
	215	12	167	10
SiO_2	467	—	—	0.8
Graphite	1580	—	5000	—
Diamond	1332.5	1.7–7.4	90	—

Data from W. Kaiser and M. Maier [21] and D.S. Knight and W. White [22].

frequency is given by

$$\gamma_R(\nu_R) \triangleq \frac{1}{I_R}\frac{dI_R}{dz} \triangleq \frac{\text{stim. rate (vol.)}^{-1}}{I_R} = \left\{\frac{\text{spont. rate (vol.)}^{-1}}{I_R}\right\} \cdot \frac{g(\nu_R)c^2}{8\pi n^2 h\nu_R^3} \cdot I_R$$

or

$$\gamma_R(\nu_R) = \left\{\text{spontaneous rate} = N\left\{\frac{\nu_R}{\nu_p}\frac{8\pi}{3}\left[\frac{d\sigma_s}{d\Omega}\right]_{90^\circ}\right\}I_p\right\}\left\{\frac{g(\nu_R)c^2}{8\pi n^2 h\nu_R^3}\right\}$$

or

$$\gamma_R(\nu_R) = N\left(\frac{d\sigma_s}{d\Omega}\right)_{90} \cdot \frac{\lambda_R^2}{3n^2} \cdot \frac{I_p}{h\nu_p} g(\nu_R) \tag{14.10.15}$$

where (14.10.13b) was used for scattered power per unit of volume. The density N is that of the state 1, and it has been assumed that state 2 was empty. In principle, we could relate the various factors in (14.10.10) to that in (14.10.15) to determine $(\partial\alpha/\partial x)^2$, but it really is not worth the exercise – (14.10.15) is correct – we forced it to agree with experimental values.

14.10.3 Density Matrix Description of the Raman Effect

The density matrix formalism will also predict the Raman gain coefficient in terms of the matrix elements connecting the various states. The problem to be addressed was shown in Fig. 14.5. We presume a system in which the single photon routes, $1 \leftrightarrow 3$ and $2 \leftrightarrow 3$, are allowed by the electric dipole route, whereas the $1 \leftrightarrow 2$ route is forbidden ($\mu_{21} = 0$). While this last assumption is not necessary, it is quite common, and the results illustrate both the uniqueness and the importance of the Raman effect. We also presume that the photon energies, $\hbar\omega_p$ and $\hbar\omega_R$, are not necessarily resonant with any of the natural energy differences, and thus the usual ploy of neglecting antiresonant field quantities will not be applicable.

The basic equation to be addressed is (14.6.7) with relaxation terms included.

$$\frac{\partial \rho_{nm}}{\partial t} + (\text{relaxation terms}) = \frac{1}{j\hbar} \sum_k (\rho_{km} H_{nk} - \rho_{nk} H_{km}) \tag{14.10.16}$$

where

$$H_{nk} = -\mu_{nk} E_T(t) = -\mu_{kn} E_T(t) = H_{kn} \quad \text{(since } H \text{ is Hermitian and real)}$$

$$H_{31} = -\mu_{31} E_T(t) = -\mu_{13} E_T(t) = H_{13} \text{ and } H_{32} = -\mu_{32} E_T(t) = -\mu_{23} E_T(t) = H_{23}$$

$$H_{21} = -(\mu_{21} = 0) E_T(t) = H_{12} = 0 \quad \text{(by assumption)}$$

$$H_{11} = E_1 = 0;\ H_{22} = E_2 = \hbar\omega_{21};\ H_{33} = E_3 = \hbar\omega_{31};\ \hbar\omega_{32} = E_3 - E_2$$

where

$$E_T(t) = E_p \cos\omega_p t + E_R \cos\omega_R t \quad \text{(the total electric field)} \tag{14.10.17}$$

It is a good practice exercise for the student to write out (14.10.16) in all of its glory (and gory) detail and include relaxation rates in the manner of Sec. 14.7. For the diagonal elements, we obtain

$$\frac{\partial \rho_{33}}{\partial t} + \frac{\rho_{33}}{\tau_3} = -j \frac{\mu_{31} E_T}{\hbar} (\rho_{31} - \rho_{31}^*) - j \frac{\mu_{32} E_T}{\hbar} (\rho_{32} - \rho_{32}^*) \tag{14.10.18a}$$

$$\frac{\partial \rho_{22}}{\partial t} + \frac{\rho_{22}}{\tau_2} = + \frac{\phi_{32} \rho_{33}}{\tau_3} + j \frac{\mu_{32} E_T}{\hbar} (\rho_{32} - \rho_{32}^*) \tag{14.10.18b}$$

$$\frac{\partial \rho_{11}}{\partial t} = (1 - \phi_{32}) \frac{\rho_{33}}{\tau_3} + \frac{\rho_{22}}{\tau_2} + j \frac{\mu_{31} E_T}{\hbar} (\rho_{31} - \rho_{31}^*) \tag{14.10.18c}$$

where τ_n^{-1} represents the decay rate of the population in state n (i.e., ρ_{nn}, $n = 1$, 2, or 3) and ϕ_{32} and ϕ_{31} are the branching ratios. Obviously, state 1 does not decay but does receive population from above. As a check, we add (14.10.18a) to (14.10.18c) and obtain $\partial(\rho_{11} + \rho_{22} + \rho_{33})/\partial t = 0$ consistent with $tr[\rho] = \text{constant} = 1$. For the off-diagonal elements, we obtain

$$\frac{\partial \rho_{32}}{\partial t} + \left(\frac{1}{T_{32}} + j\omega_{32} \right) \rho_{32} = -j \frac{\mu_{32} E_T}{\hbar} (\rho_{33} - \rho_{22}) + j \frac{\mu_{31} E_T}{\hbar} \rho_{21}^* \tag{14.10.19a}$$

$$\frac{\partial \rho_{31}}{\partial t} + \left(\frac{1}{T_{31}} + j\omega_{31}\right)\rho_{31} = -j\frac{\mu_{31}E_T}{\hbar}(\rho_{33} - \rho_{11}) + j\frac{\mu_{31}E_T}{\hbar}\rho_{21}^* \tag{14.10.19b}$$

$$\frac{\partial \rho_{21}}{\partial t} + \left(\frac{1}{T_{21}} + j\omega_{21}\right)\rho_{21} = -j\frac{\mu_{31}E_T}{\hbar}\rho_{32}^* + j\frac{\mu_{32}E_T}{\hbar}\rho_{31} \tag{14.10.19c}$$

where T_{nm}^{-1} is the decay rate of the off-diagonal elements (which control the polarization).

Notice that the difference $(\rho_{22} - \rho_{11})$ does not appear explicitly in the equation of motion for ρ_{21} (14.10.19c) nor does ρ_{21} explicitly appear in any of the population equations (14.10.18). It does affect the populations of 3, 2, and 1 through its influence on (ρ_{31}, ρ_{32}), which do appear in the density equations.

Note also that we are faced with a rather messy situation owing to the presence of two electromagnetic fields at frequencies (ω_p, ω_R) driving the elements ρ_{31} and ρ_{32} in first order in (E_p, E_s) and product terms $(\rho_{31}E_T, \rho_{32}E_T)$ in (14.10.19c). These product terms create a new driving term at a frequency $(\omega_p - \omega_R)$, which is resonant with the forbidden transition $2 \leftrightarrow 1$. Other frequencies are generated, but the problem is rapidly getting out of hand: We need some simplifying approximations: Assume

1. We neglect saturation and thus the fields do not affect the populations to any extent. Thus, (14.10.18a) to (14.10.18c) are ignored.
2. The population in state 3 is negligible $(\rho_{33} = 0)$, and 1 and 2 start and remain in thermodynamic equilibrium as a consequence of assumption 1.

$$\frac{\rho_{22}}{\rho_{11}} = \exp[-\Delta E_{21}/kT]$$

$$\therefore \qquad \rho_{11} = \frac{1}{1 + \exp[-\Delta E_{21}/kT]} \sim 1 \tag{14.10.20a}$$

$$\rho_{22} = \frac{\exp[-\Delta E_{21}/kT]}{1 + \exp[-\Delta E_{21}/kT]} \ll 1 \tag{14.10.20b}$$

3. Since the population in state 2 is much smaller than 1, we neglect the generation of the anti-Stokes field corresponding to

$$\hbar\omega_p + (\text{state } 2) \rightarrow (\text{another virtual state}) \rightarrow \hbar(\omega_p + \omega_{21}) = \hbar\omega_{AS}$$

4. We set $T_{21} = T_2$ in (14.10.19c) and ignore all other T^{-1} because those rates are small compared to the explicit optical frequencies in (14.10.19a) and (14.10.19b).
5. We anticipate a linear response to the fields E_p and E_R as

$$P_L^{(1)} \sim \epsilon_0 \left\{\chi_L(\omega_p)E_p + \chi_L(\omega_R)E_R\right\} \tag{14.10.21a}$$

6. We also anticipate a nonlinear response of the form*

$$P_{NL}^{(2)} = \epsilon_0 \left\{\chi_{NL}(\omega_p, E_p^2)E_p + \chi_{NL}(\omega_R, E_p^2)E_R\right\} \tag{14.10.21b}$$

*The superscript is intended to indicate the order of the solution with (1) being the usual linear response and (2) being a correction to second order.

where it is presumed that the electric field of the pump is much larger than that of the Stokes' wave so that it is the primary cause of the nonlinearity. These nonlinear factors "correct" the linear response described by (14.10.21a). As we will see, the last term in (14.10.21b) provides gain for the Stokes wave whereas the first contributes added attenuation on the pump to account for the conversion of energy from ω_p to ω_s.

Since the nonlinear response is the cause of the Raman effect, we must be very cautious in the approach to the solution. It is best to start out with the computation of the linear response of (14.10.19a) and (14.10.19b). After the terms $E_T\rho_{21}$ or $E_T\rho_{21}^*$ are neglected, the procedure is identical to that followed in Sec. 14.8 for the simple two level atom. We obtain

$$\frac{\rho_{32}^{(1)}}{(\rho_{33}-\rho_{22})} = \Omega_{32p}\left[\frac{e^{-j\omega_p t}}{\omega_{32}-\omega_p}+\frac{e^{+j\omega_p t}}{\omega_{32}+\omega_p}\right]+\Omega_{32s}\left[\frac{e^{-j\omega_s t}}{\omega_{32}-\omega_s}+\frac{e^{+j\omega_s t}}{\omega_{32}+\omega_s}\right] \tag{14.10.22a}$$

$$\frac{\rho_{31}^{(1)}}{(\rho_{33}-\rho_{11})} = \Omega_{31p}\left[\frac{e^{-j\omega_p t}}{\omega_{31}-\omega_p}+\frac{e^{+j\omega_p t}}{\omega_{31}+\omega_p}\right]+\Omega_{31s}\left[\frac{e^{-j\omega_s t}}{\omega_{31}-\omega_s}+\frac{e^{+j\omega_s t}}{\omega_{31}+\omega_s}\right] \tag{14.10.22b}$$

where $E_T(t) = E_p\cos\omega_p t + E_s\cos\omega_s t$ was assumed and $\Omega = \mu E/2\hbar$ as usual. For most Raman cases, none of the denominators are nearly equal to zero, and hence the rotating wave approximation cannot be used to discard any of the frequency components by comparison to others.

Now we insert the linear response expressed by (14.10.22) into (14.10.19c), and now there is a definite role for the rotating wave approximation. Note that a factor $E_{Tx}(t)\cdot[\mu_{31x}\rho_{32}^* + \mu_{32x}\rho_{31}]$ appears as the driving force on the right side. This product will generate terms varying as

$$\text{RHS of (14.10.19c)} \rightarrow \exp[\pm j(\omega_p \pm \omega_s)t + \exp[\pm j2(\omega_p+\omega_s)t]$$

Only the factor varying as $\exp[-j(\omega_p-\omega_s)t] \rightarrow \exp[-j(\sim\omega_{21})t]$ can provoke a resonant response in ρ_{21} so that it, in turn, can make a significant contribution in the nonlinear (and heretofore neglected) terms in the equations for ρ_{32} and ρ_{31}. We keep that term and ignore the rest. Furthermore, $|\rho_{33}-\rho_{22}| \ll |\rho_{22}-\rho_{11}|$, allowing us to ignore the term $E_T\rho_{32}^*$ in (14.10.19c). Thus, we multiply (14.10.22b) by $+j\mu_{32x}E_{Tx}/\hbar$ and keep only those terms varying as $\exp[-j(\omega_p-\omega_s)t]$ which leads to a resonant behavior in ρ_{21}.

$$\frac{\partial\rho_{21}}{\partial t}+\left(\frac{1}{T_2}+j\omega_{21}\right)\rho_{21} = +j\left[\frac{\Omega_{32p}\Omega_{31s}}{\omega_{31}+\omega_s}+\frac{\Omega_{32s}\Omega_{31p}}{\omega_{31}-\omega_p}\right](\rho_{33}-\rho_{11})e^{-j(\omega_p-\omega_s)t} \tag{14.10.22c}$$

Hence

$$\rho_{21} = \sigma_{21}e^{-j\Delta\omega t} \qquad \text{with} \quad \Delta\omega = \omega_p-\omega_s \tag{14.10.23a}$$

$$\text{where } \sigma_{21} = +j\left(\frac{1}{\Re}\right)\Omega_{32p}\Omega_{31s}\left[\frac{1}{\omega_{31}+\omega_s}+\frac{1}{\omega_{31}-\omega_p}\right](\rho_{33}-\rho_{11}) \tag{14.10.23b}$$

$$\text{with} \quad \Re = \frac{1}{T_2} + j(\omega_{21} - \Delta\omega) \text{ complex Raman line shape.} \tag{14.10.23c}$$

Now we use (14.10.23) to evaluate the term $(\mu_{32x}E_{Tx}/\hbar)\rho_{21}$ in (14.10.19a) and (14.10.19b) and start the solution process all over again.

By now, we should realize a bit of rhythm to this exercise: The product of ρ_{21} and E_T will generate frequencies at $\pm(\omega_p \pm \Delta\omega)$ and $\pm(\omega_s \pm \Delta\omega)$. The terms at $\pm(\omega_p - \Delta\omega) = \pm[\omega_p - (\omega_p - \omega_s)] = \pm\omega_s$ will modify the response of ρ_{31} at $\pm\omega_s$ and the terms $\pm(\omega_s + \Delta\omega) = \pm(\omega_s + (\omega_p - \omega_s) = \pm\omega_p$ will modify that at ω_p. Let us examine only those terms at the Stokes frequency (ω_s) that are proportional to E_p^2 (i.e., Ω_p^2), which represents the nonlinear terms responsible for the Raman gain.

$$\frac{\partial\rho_{32}^{(2)}}{\partial t} + j\omega_{32}\rho_{32}^{(2)} = -j\Omega_{31p}\sigma_{21}^* e^{-j\omega_s t} \Rightarrow \rho_{32}^{(2)} = \frac{-\Omega_{31p}\sigma_{21}^*}{\omega_{32} - \omega_s} e^{-j\omega_s t} \tag{14.10.24a}$$

$$\frac{\partial\rho_{31}^{(2)}}{\partial t} + j\omega_{31}\rho_{32}^{(2)} = +j\Omega_{32p}\sigma_{21} e^{+j\omega_s t} \Rightarrow \rho_{31}^{(2)} = \frac{+\Omega_{32p}\sigma_{21}^*}{\omega_{31} + \omega_s} e^{+j\omega_s t} \tag{14.10.24b}$$

Thus the second-order or nonlinear contribution to the complex polarization at ω_s is given by

$$P_{NL}^{(2)} = N\left\{\mu_{32x}[\rho_{32}^{(2)}]^* + \mu_{31x}[\rho_{31}^{(2)}]\right\} e^{j\omega_s t} = N\left\{-\frac{\mu_{32x}\Omega_{31p}\sigma_{21}}{\omega_{32} - \omega_s} + \frac{\mu_{31x}\Omega_{32p}\sigma_{21}}{\omega_{31} + \omega_s}\right\} e^{j\omega_s t}$$

where only those terms with a $e^{j\omega_s t}$ time variation have been kept.

$$\begin{aligned} P_{NL}^{(2)} &= \frac{N\mu_{32x}^2\mu_{31x}^2}{8\hbar^3} E_p^2\left[\frac{1}{\omega_{31} + \omega_s} + \frac{1}{\omega_{31} - \omega_p}\right]\left[\frac{1}{\omega_{31} + \omega_s} - \frac{1}{\omega_{32} - \omega_s}\right] \\ &\quad \times \left(\frac{j}{\Re}\right)(\rho_{33} - \rho_{11})E_s e^{j\omega_s t} \\ &\triangleq \epsilon_0(\chi_R' - j\chi_R'')\frac{E_s}{2} e^{j\omega_s t} \end{aligned} \tag{14.10.25}$$

where the definition of Ω and the complex polarization has been used. Thus the complex susceptibility is given by

$$\begin{aligned} \chi_R' - j\chi_R'' &= \frac{N\mu_{31x}^2\mu_{32x}^2 E_{px}^2}{4\epsilon_0\hbar^3}\left\{\frac{(2\omega_s + \omega_{21})(2\omega_{31} - \omega_{21})}{(\omega_{31} + \omega_s)^2(\omega_{31} - \omega_p)(\omega_{32} - \omega_s)}\right\} \\ &\quad \times \left[\frac{(\omega_{21} - \Delta\omega) + j1/T_2}{(\omega_{21} - \Delta\omega)^2 + (1/T_2)^2}\right](\rho_{11} - \rho_{33}) \end{aligned} \tag{14.10.26}$$

We again retreat to (13.3.5) to obtain the gain coefficient at the Stokes frequency in terms of the imaginary part of the susceptibility [recall (13.3.5)]

$$\gamma_R(\omega_s) = k(\omega_s)\left\{-\frac{\chi_R''(\omega_s, E_p^2)}{n^2}\right\}$$

Substituting $n = 1$ and $k = \omega_s/c$ for a gas, $\mu_{31x}^2 = \mu_{31}^2/3$ into (14.10.26), and extracting the imaginary part of χ yields the gain coefficient

$$\frac{\gamma}{N} = \frac{\omega_s}{c}\frac{\mu_{31}^2\mu_{32}^2E_p^2}{36\epsilon_0\hbar^3}\cdot\left\{\left[\frac{(2\omega_s+\omega_{21})(2\omega_{31}-\omega_{21})}{(\omega_{31}+\omega_s)^2(\omega_{31}-\omega_p)(\omega_{32}-\omega_s)}\right]\right.$$
$$\left.\times\left[\frac{1/T_2}{(\omega_{21}-\Delta\omega)^2+(1/T_2)^2}\right](\rho_{11}-\rho_{33})\right\} \tag{14.10.27}$$

This has the same functional form as (14.10.15) derived from the relationship between the spontaneous and stimulated Raman effects. If we equate the expression given by (14.10.15) to (14.10.27), then we obtain the scaling of the differential scattering cross section with the Raman wavelength:

$$\left.\frac{d\sigma}{d\Omega}\right|_{90^\circ} \propto \lambda_s^{-4} \tag{14.10.28}$$

Finally, we should recognize that state 3 represents the "rest" of the atom, and we should sum over all of the rest of the states. Doing so generates another formidable appearing equation that is of limited use since we seldom know all or even most of the coefficients involved. Hence we are forced to retreat to (14.10.15) for the exact (i.e., experimental) value.

We can use the above for the case of a resonant Raman effect where ω_p is close to ω_{31} and ω_s is therefore close to ω_{32}. We retreat to (14.10.25), multiply the second term in each bracket by j in the numerator and denominator, and replace $j(\omega_{31}-\omega_p)$ by $1/T_{31}+j(\omega_{31}-\omega_p)$ and $j(\omega_{32}-\omega_s)$ by $1/T_{32}+j(\omega_{32}-\omega_s)$ to avoid solving the problem again. Those two brackets become

$$\left[\frac{1}{\omega_{31}+\omega_s}+\frac{jT_{31}}{1+j(\omega_{31}-\omega_p)T_{31}}\right]\left[\frac{1}{\omega_{31}+\omega_s}-\frac{jT_{32}}{1+j(\omega_{32}-\omega_s)T_{32}}\right] \tag{14.10.29}$$

For exact resonance, $\omega_{31} = \omega_p$, $\omega_{32} = \omega_s$, and $\omega_p - \omega_s = \omega_{21}$. Then the second term in each bracket is much larger than the first and we obtain

$$\left.\frac{\gamma}{N}\right|_{\text{res}} = \frac{\omega_s}{c}\frac{\mu_{31x}^2\mu_{32x}^2}{36\epsilon_0\hbar^3}T_{21}T_{32}T_{31}E_p^2(\rho_{11}-\rho_{33}) \tag{14.10.30}$$

which is an enormous enhancement over that predicted by (14.10.27). However, the normal one-photon absorption process will change the population in state 3 (and 2), making some of the assumptions leading to this point invalid. We should start all over again, but that task is saved for more advanced works.

14.11 PROPAGATION OF PULSES: SELF-INDUCED TRANSPARENCY

14.11.1 Motivation for the Analysis

Sections 14.1 to 14.8 demonstrated that the rate equation approach yielded a satisfactory description of the interaction of the electromagnetic field with an atomic system provided we interpreted the parameters of the gain equation correctly. In view of the fact that we must depend upon experiments for these parameters, the rate equation approach is guaranteed to yield the correct answers, most of the time. It is the purpose of this section to examine an exception – where the two approaches yield widely differing answers – and thus to identify circumstances where the more formal density matrix is needed.

Consider the case of an absorbing medium being interrogated by a pulse of optical radiation tuned to the center of the absorption profile. Aside from the minor difference of considering a medium with absorption (rather than gain), this is precisely the problem addressed in Sec. 9.6 and all of the theory developed there should be applicable provided we use G_0 to be less than 1. What that previous theory predicts is sketched in Fig. 14.6 for the case of a "square" input pulse. The output pulse is delayed by the normal propagation delay of $l_g/(c/n)$ and distorted in time relative to the arrival of the initial part of the pulse. The initial temporal part reflects the attenuation with the later fraction yielding essentially unity transmission *if* the energy in the pulse is much greater than the saturation energy, $w_s = h\nu/2\sigma$.

While this "bleaching" could be called "self-induced transparency," the latter is reserved for analysis given next. We will find that the time shift is much greater than $l_g/(c/n)$ (by orders of magnitude), and the transparency condition is specified by the "area" A of the pulse.

$$A = \lim_{t\to\infty} \frac{\mu}{\hbar} \int_{-\infty}^{t} E(t')\,dt' = (0, 2, 4, \ldots)\pi \tag{14.11.1}$$

where $E(t')$ is the envelope of the electric field in the medium. In other words, the theory to follow predicts that a pulse will propagate *without attenuation* and *without distortion* if the area is chosen properly. If we do not properly match the correct temporal envelope,

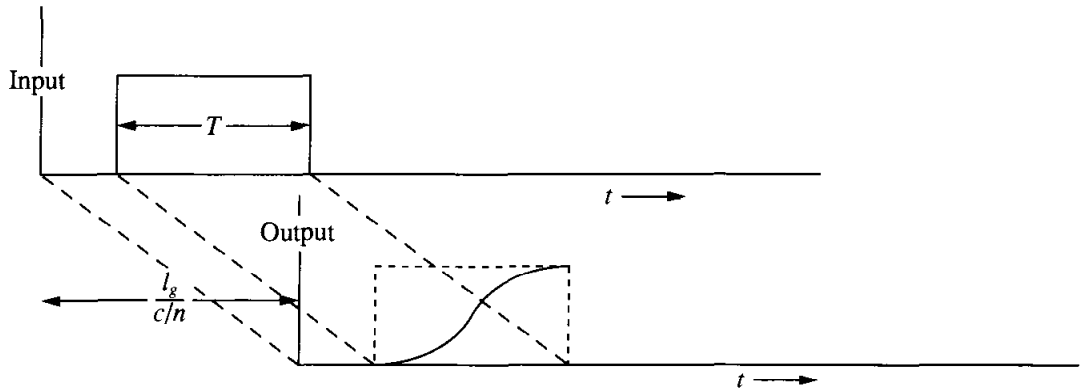


FIGURE 14.6. Intensity dependent "bleaching" according to Sec. 9.6.

numerical calculations show that the pulse will form itself and then propagate without further attenuation at a level or area corresponding to the next lowest value of $(2m\pi)$. That is considerably different from what is predicted by Sec. 9.6.

It requires the assumption that the time duration of the pulse be much less than T_2 (which is usually less than τ), and this provides the key to the range of its applicability. If the pulse is very short compared to T_2, then its spectral width is always much larger than $1/\pi T_2$, which is the homogeneous bandwidth $(\Delta\nu_h)$ of the region of attenuation. Implicit in the analysis of Sec. 9.6 is the assumption that the spectral content of the input is much less than $\Delta\nu_h$, and thus the two theories are complementary. However, "short pulses" are very much in demand, and thus it behooves us to examine the response of an atomic system under these circumstances.

14.11.2 A Self-Consistent Analysis of the Field-Atom Interaction

(A) Preliminary Steps. Let us first identify the system of equations facing us. Since we are dealing with the electromagnetic field, we must solve the inhomogeneous wave equation that includes the polarization p_a contributed by the atoms.*

$$\frac{\partial^2 e}{\partial z^2} - \left(\frac{n}{c}\right)^2 \frac{\partial^2 e}{\partial t^2} = \mu_0 \frac{\partial^2 p_a}{\partial t^2} \tag{14.11.2}$$

The polarization p_a is related to the density matrix by (14.6.9)

$$p_a = N\mu\langle\rho_{21} + \rho_{12}\rangle = N\mu\langle\rho_{21} + \rho_{21}^*\rangle \tag{14.11.3}$$

The small letters for the electric field and polarization remind us that the quantities are implicit functions of time and space without the factor $\exp j\omega t$ suppressed. The averaging symbols indicate that we need to average our results (at the appropriate point of our development) over the inhomogeneous distribution of center frequencies of the atomic system. This merely involves multiplying by and integrating over the line shape function describing that distribution. It is a chore to be faced, but for now we can proceed as if the averaging is not needed.

We need the electric field to compute the time and space evolution of the density matrix (cf., (14.7.11) and (14.7.12), rewritten below with a reordering of the factors $(\rho_{22} - \rho_{11}) = -(\rho_{11} - \rho_{22})$ to imply absorption, and with τ and T_2 dropped.

$$\frac{\partial}{\partial t}(\rho_{11} - \rho_{22}) = j\frac{2\mu e(z,t)}{\hbar}(\rho_{21} - \rho_{21}^*) \tag{14.11.4}$$

$$\frac{\partial \rho_{21}}{\partial t} + j\omega_{21}\rho_{21} = +j\frac{\mu e(z,t)}{\hbar}(\rho_{11} - \rho_{22}) \tag{14.11.5}$$

To try to combine (14.11.2) to (14.11.5) is an extremely messy arithmetic problem and is complicated by the fact that (14.11.2) involves second (partial) derivatives with respect to z

*The field and the polarization vary rapidly in time in addition to the normal harmonic variation. Thus we use small letters to represent the complete expression and reserve the capital letters for the envelope.

and time, fortunately (14.11.4) and (14.11.5) involve only time. However, it is obvious that we need (desperately) some simplifying algorithms and sensible approximations.

We start by expressing the (real) electric field and (real) polarization by an envelope function (of z and t) times the traditional plane wave expression

$$e(z, t) = \frac{1}{2} E(z, t) \cdot \left\{ e^{j[\omega t - kz - \phi(z,t)]} + e^{-j[\omega t - kz - \phi(z,t)]} \right\} \tag{14.11.6}$$

$$p_a(z, t) = \frac{1}{2} \left\{ [U + jV] e^{-j[\omega t - kz - \phi(z,t)]} + [U - jV] e^{+j[\omega t - kz - \phi(z,t)]} \right\} \tag{14.11.7}$$

where both U and V are real functions of z and t.

There are no approximations being made in these two steps. It merely expresses the unknown quantities in terms of an envelope rather than $e(z, t)$ and $p_a(z, t)$. The next step does use approximations, and the resulting simplification justifies the strategy: Substitute (14.11.6) and (14.11.7) into the wave equation (14.11.2), equate real and imaginary parts, and use the slowly varying envelope approximation (as in Chapter 3 for the derivation of the paraxial wave equation).

Assume

$$\begin{aligned}
&\frac{\partial}{\partial t} \text{ of } (E, U, V, \text{ or } \phi) \ll \omega(E, U, V, \text{ or } \phi) \\
&\frac{\partial}{\partial z} \text{ of } (E, U, V, \text{ or } \phi) \ll k(E, U, V, \text{ or } \phi) \\
&\frac{\partial^2}{\partial t^2} \text{ of } (E, U, V, \text{ or } \phi) \ll \omega \frac{\partial}{\partial t} \text{ of } (E, U, V, \text{ or } \phi) \\
&\frac{\partial^2}{\partial z^2} \text{ of } (E, U, V, \text{ or } \phi) \ll k \frac{\partial}{\partial z} \text{ of } (E, U, V, \text{ or } \phi)
\end{aligned} \tag{14.11.8}$$

where $k = \omega n/c$. The wave equation becomes

$$\boxed{\frac{\partial E}{\partial z} + \frac{n}{c} \frac{\partial E}{\partial t} = -\frac{\omega c \mu_0}{2n} V} \tag{14.11.9a}$$

$$\boxed{E \left(\frac{\partial \phi}{\partial z} + \frac{n}{c} \frac{\partial \phi}{\partial t} \right) = \frac{\omega c \mu_0}{2n} U} \tag{14.11.9b}$$

Now we use the definition of the polarization, (14.11.3) and the representation given by (14.11.7) to express ρ_{21} in terms of the variables $U(z, t)$ and $V(z, t)$.

$$\begin{aligned}
p_a = N\mu(\rho_{21} + \rho_{21}^*) = \frac{1}{2} \Big\{ &[U(z, t) + jV(z, t)] e^{-j(\omega t - kz - \phi)} \\
&+ [U(z, t) - jV(z, t)] e^{+j(\omega t - kz - \phi)} \Big\}
\end{aligned}$$

Thus

$$\rho_{21} = \frac{1}{2N\mu} [U(z, t) + jV(z, t)] e^{-j(\omega t - kz - \phi)} \tag{14.11.10}$$

Now we use (14.11.10) to convert the dynamical equations for the density matrix (14.11.4) and (14.11.5), to one expressing the dynamics of U and V.

$$\begin{aligned} \frac{\partial \rho_{21}}{\partial t} &= \frac{1}{2N\mu}\left\{\left[\frac{\partial U}{\partial t} + j\frac{\partial V}{\partial t}\right] - j\left[\omega - \frac{\partial \phi}{\partial t}\right][U + jV]\right\} e^{-j(\omega t - kz - \phi)} \\ + \qquad & \qquad + \\ j\omega_{21}\rho_{21} &= \frac{1}{2N\mu}\left\{j\omega_{21}U - \omega_{21}V\right\} e^{-j(\omega t - kz - \phi)} \\ \| \qquad & \qquad \| \\ j\frac{\mu \mathbf{e}(z,t)}{\hbar}(\rho_{11} - \rho_{22}) &= j\frac{\mu E(z,t)}{2\hbar}[e^{j(\omega t - kz - \phi)} + e^{-j(\omega t - kz - \phi)}](\rho_{11} - \rho_{22}) \end{aligned} \tag{14.11.11}$$

We neglect the term, $\exp[+j(\omega t - kz - \phi)]$ in the spirit of the rotating wave approximation, cancel the common factors of $\exp[-j(\omega t - kz - \phi)]$, and equate real and imaginary parts. We also recognize that the right side of the last line of (14.11.11) is imaginary since $e(z, t)$ is always real and $N(\rho_{11} - \rho_{22})$ is the real population difference. Thus, the real and imaginary parts of (14.11.11) leads to:

$$\frac{\partial U}{\partial t} = \left(\Delta\omega + \frac{\partial \phi}{\partial t}\right) V \tag{14.11.12}$$

$$\frac{\partial V}{\partial t} = \left(\Delta\omega + \frac{\partial \phi}{\partial t}\right) U + \frac{\mu E}{\hbar} W \tag{14.11.13}$$

where $$\Delta\omega = (\omega_{21} - \omega)$$

and $$W = N\mu(\rho_{11} - \rho_{22}) \tag{14.11.14}$$

Following the same prescription for (14.11.4) yields

$$\frac{\partial W}{\partial t} = -\frac{\mu E}{\hbar} V \tag{14.11.15}$$

So far, it might appear that we are spinning our wheels and converting one set of unknowns into another. However, the variables U, V, W and their dynamics can be interpreted in terms of a very simple and familiar equation. Hence, let us break for a moment to examine (14.11.12), (14.11.13), and (14.11.15).

(B) Vector Form of the Density Matrix. All three of these equations can be expressed by one vector equation by defining a right-handed orthogonal coordinate system $(\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \hat{\mathbf{a}}_3)$ with corresponding components projected onto those axes.

Let

$$\mathbf{r} = \frac{1}{N\mu}\left\{U\hat{\mathbf{a}}_1 + V\hat{\mathbf{a}}_2 + W\hat{\mathbf{a}}_3\right\} \tag{14.11.16a}$$

and

$$\mathbf{T} = \frac{-\mu E}{\hbar}\hat{\mathbf{a}}_1 - (\Delta\omega + \frac{\partial \phi}{\partial t})\hat{\mathbf{a}}_3 \tag{14.11.16b}$$

Then (14.11.12), (14.11.13), and (14.11.15) can be written in a very compact manner.

$$\frac{\partial \mathbf{r}}{\partial t} = \mathbf{T} \times \mathbf{r} \tag{14.11.17}$$

While compaction of mathematics is always worthwhile, the form of (14.11.17) should be familiar: It describes the *precession* of vector quantity **r** about the vector **T** in the manner shown in Fig. 14.7 in the same way that a gyroscope precesses around the gravitational field or a magnetic dipole $\mathbf{m}_d$ precesses around a magnetic field. In the latter case, we (should) know that the time rate of change of the angular momentum, **L**, is equal to the torque, which, in the case of the magnetic dipole, is $\mathbf{m}_d \times \mathbf{B}$

$$\frac{d\mathbf{L}}{dt} = \mathbf{m}_d \times \mathbf{B} \tag{14.11.18a}$$

The magnetic moment $\mathbf{m}_d$ of an electron is related to its angular momentum $\mathbf{L}_d$ by

$$\mathbf{m}_d = -\gamma \mathbf{L}_d \tag{14.11.18b}$$

where $\gamma = |\mathbf{m}_d/\mathbf{L}_d|$ is the gyromagnetic ratio $(e/2m_e)$ with the minus sign in (14.11.18b) indicating that the angular momentum and the magnetic dipole moment are antiparallel. Thus, (14.11.18a) becomes

$$\frac{d\mathbf{m}_d}{dt} = -\gamma \mathbf{m}_d \times \mathbf{B} \tag{14.11.18c}$$

which is the same form as (14.11.17). It is worthwhile to inspect Fig. 14.7 to obtain some rather obvious results that can then be applied to the pseudovector **r** representing the components of the density matrix. In Fig. 14.7(a), we presume a magnetic dipole oriented at an angle θ with respect to a static magnetic field $\mathbf{B}_0$ and thus (14.11.18a) predicts a precession frequency of $\omega_0 = \gamma \mathbf{B}_0$ around the field lines. In Fig. 14.7(b) we presume an additional

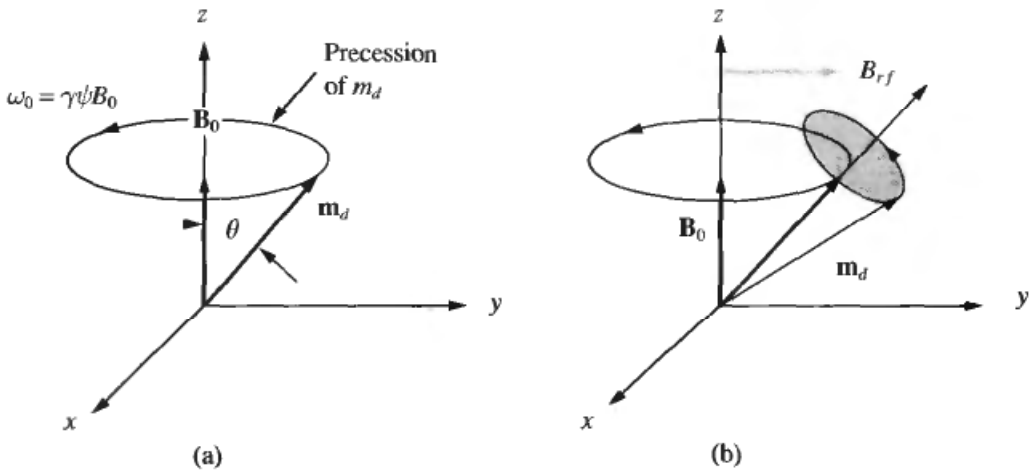


FIGURE 14.7. (a) The magnetic moment is showing precessing around the static magnetic field. (b) The magnetic moment "tips" from its equilibrium position if a near resonant AC magnetic field is applied perpendicular to $\mathbf{B}_0$.

AC magnetic field $\mathbf{B} = \mathbf{B}_{rf} \cos \omega_0 t$, which is perpendicular to the static value and synchronized to the natural precession frequency yielding a net "torque" tending to tip the magnetic dipole.

For instance, it should be obvious that the z component of $\mathbf{m}_d$ is constant with time in Fig. 14.7(a) while the x and y components vary harmonically. Applying this same line of reasoning to the vector $\mathbf{r}$ of (14.11.17), we find that

Example

If $E = 0$ for $t > t_0$ (i.e., the pulse is over)

$$U(\Delta\omega, z, t) = U_0(\Delta\omega, z, t_0) \cos[\Delta\omega(t - t_0)] + V_0 \sin[\Delta\omega(t - t_0)] \quad (14.11.19a)$$

$$V(\Delta\omega, z, t) = -U_0(\Delta\omega, z, t_0) \sin[\Delta\omega(t - t_0)] + V_0 \cos[\Delta\omega(t - t_0)] \quad (14.11.19b)$$

$$W(\Delta\omega, z, t) = W(\Delta\omega, z, t_0) \quad (14.11.19c)$$

where U_0, V_0, and $W(\Delta\omega, z, t_0)$ are the values of those components of $\mathbf{r}$ after the electric field returns to zero at $t > t_0$ and are functions of the detuning $\Delta\omega$, space, and t_0. In other words, the pseudovector representing the real and imaginary parts of the density matrix and the population difference will precess around the direction of the electric field even after the field has vanished.

Another transparent case is that of resonance between the applied AC field and the natural precession frequency. For the case shown in Fig. 14.7(b) we see that there is now a net torque tending to tip the vector $\mathbf{m}_d$ away from alignment with the z axis.

Example

If $\Delta\omega + \partial\phi/\partial t = 0$, $E \neq 0$ with initial conditions $U(z, -\infty) = V(z, -\infty) = 0$, then a simple exercise in differentiation will show that the following are solutions to (14.11.12), (14.11.13), and (14.11.15)

$$U(0, z, t) = 0 \quad (14.11.20a)$$

$$V(0, z, t) = W_0 \sin\theta(z, t) \quad (14.11.20b)$$

$$W(0, z, t) = W_0 \cos\theta(z, t) \quad (14.11.20c)$$

where

$$\theta(z, t) = \frac{\mu}{\hbar} \int_{-\infty}^{t} E(z, t')\, dt' \quad (14.11.21)$$

In the limit of $t \to \infty$, the quantity $\theta(z, t)$ is called the area, A, of the electric field of the pulse and plays a central role in the theory to follow. We will need (14.11.16) through (14.11.21) for this theory, and the serious student should verify their correctness by direct substitution into the differential equations. (While we can disregard the one-to-one analogy with the magnetic dipole behavior, that ignores an enormous body of literature on magnetic resonance, which was the precursor to the optical phenomena.)

14.11.3 "Area" Theorem

This is truly an amazing result first predicted and observed by S.L. McCall and E.L. Hahn (Ref. 23), which provides dramatic proof of the necessity of a full quantum description of

the field-atom interaction. The area theorem states that

$$\boxed{\frac{dA}{dz} = -\frac{\alpha}{2}\sin A} \tag{14.11.22}$$

where

$$\boxed{A = \lim_{t\to\infty}\frac{\mu}{\hbar}\int_{-\infty}^{t} E(z, t')\,dt'} \tag{14.11.23}$$

and

$$\alpha = \frac{\omega\pi\mu_0 N\mu^2 g(\nu_{21})}{n\hbar} = A_{21}\frac{\lambda_0^2}{8\pi n^2}g(\nu_{21})N_1 \tag{14.11.24}$$

The quantity α is the normal absorption (or gain) coefficient experienced by a CW optical signal tuned to the center of an optical transition and with N_2 assumed to be 0. The proof of (14.11.22) and (14.11.23) is a matter of patience with differential calculus, so it is best to consider the implications. (We should also remind ourselves that we are considering pulses that are short compared to T_2 or τ and thus have a spectral width *larger* than the homogeneous line width. Hence, there is a (slight) danger of applying this result in inappropriate situations.)

Equation (14.11.22) states that if A equals a multiple of π, the right side is zero and thus the area – and hence the electric field – *does not change with distance.*

$$\frac{dA}{dz} = 0 \qquad \text{if} \qquad A = m\pi \qquad m = 0, 1, 2, 3, 4\cdots \tag{14.11.25}$$

It is easy to convince ourselves that pulse areas with even multiples of π are stable whereas odd integers represent an unstable area (presuming α is positive). This implies that an input pulse will coalesce to the stable values. If $m > 1$, the pulse area will go to a 2π pulse. If $m < 1$, the area goes to 0π. If A is very small such that $\sin A \sim A$, then

$$\frac{dA}{dz} = -\frac{\alpha}{2}A \qquad \text{or} \qquad A(z) = A(z=0)\exp\left[-\frac{\alpha}{2}z\right] \tag{14.11.26a}$$

and the intensity E^2 related to A^2 varies as

$$I(z, t) = \frac{E^2(z, t)}{2\eta_0} \Rightarrow I(z=0, t)e^{-\alpha z} \tag{14.11.26b}$$

and we recover our normal result. If, however, the area is an even value of π, then the pulse propagates without attenuation (nor loss of *any* energy) even though α may be exceedingly large.

The physical reason for this counter intuitive result lies in the "tipping" picture presented earlier. Suppose the area $A = 2\pi$. If the medium started out in the normal, noninverted state before the pulse arrived, then $\rho_{22} - \rho_{11} = -1$ (i.e., all atoms are in the lower state). When $\theta = \pi$ (i.e., halfway through the pulse) the sign of W is changed from W_0 to $-W_0$ or $\rho_{22} - \rho_{11} = +1$ by (14.11.20c) and now we have a complete inversion. For

the latter part of the pulse, θ increases from π to 2π, and (14.11.20c) returns W to its initial value. There is *no net change in the population difference* and thus *there is no net energy lost.*

This is true provided the pulse shape is a proper one, a topic that we have carefully avoided. To obtain the characteristic pulse shape, we require a *simultaneous* solution to (14.11.9), (14.11.12), (14.11.13), (14.11.15), and (14.11.22) along with the definitions. These are repeated next.

From the wave equation

$$\frac{\partial E}{\partial z} + \frac{n}{c}\frac{\partial E}{\partial t} = -\frac{\omega c \mu_0}{2n}\langle V \rangle \tag{14.11.9a}$$

$$E\left(\frac{\partial \phi}{\partial z} + \frac{n}{c}\frac{\partial \phi}{\partial t}\right) = \frac{\omega c \mu_0}{2n}\langle U \rangle \tag{14.11.9b}$$

From the density matrix equations

$$\frac{\partial U}{\partial t} = \left(\Delta\omega + \frac{\partial \phi}{\partial t}\right) V \tag{14.11.12}$$

$$\frac{\partial V}{\partial t} = -\left(\Delta\omega + \frac{\partial \phi}{\partial t}\right) U + \frac{\mu E}{\hbar} W \tag{14.11.13}$$

$$\frac{\partial W}{\partial t} = \frac{-\mu E}{\hbar} V \tag{14.11.15}$$

$$\frac{dA}{dz} = -\frac{\alpha}{2}\sin A \tag{14.11.22}$$

where

$$W = N\mu(\rho_{11} - \rho_{22}) \tag{14.11.14}$$

and

$$A = \lim_{t\to\infty} \frac{\mu}{\hbar}\int_{-\infty}^{t} E(z, t')\, dt' \tag{14.11.23}$$

$$\alpha = A_{21}\frac{\lambda_0^2}{8\pi n^2} g(0) N_1 \tag{14.11.24}$$

The faint of heart would not expect a simple analytic solution to this system of six equations, but, amazingly enough, one does exist and will be given, the details of the solution being quite drawn out.

We might discern that there is a bit of apparent inconsistency involved in the area theorem and the consequences just discussed. We have neglected the dephasing term, $1/T_2 \to 0$, and population relaxation rate, $1/\tau \to 0$, in the density matrix equations, yet the normal absorption coefficient, including the line shape, appears in the result. As we shall see, the averaging over the distribution of center frequencies introduces the line shape, but, strictly speaking, $\alpha \to 0$ if the relaxation rates are zero. To correct this deficiency, we would need to insert a term $[-U/T_2]$ in (14.11.12) and a $[-(W - W_0)/\tau]$ term in (14.11.15). However, this is a complexity that we do not deserve, especially if the pulse duration is much less

than those characteristic times. Thus we keep our assumptions and trust a more involved numerical analysis that indicates that our procedure is quite good.

As was mentioned before, another issue must be addressed. All transitions are inhomogeneously broadened to a degree, and we must average the quantities U and V over this distribution of center frequencies or equivalently, a distribution of detuning factors before inserting into (14.11.9a) and (14.11.9b). Expressing those quantities as functions of the detuning $\Delta\omega$ yields

$$\langle V \rangle = \int_{-\infty}^{+\infty} V(\Delta\omega) g(\Delta\omega)\, d(\Delta\omega) \tag{14.11.27a}$$

$$\langle U \rangle = \int_{-\infty}^{+\infty} U(\Delta\omega) g(\Delta\omega)\, d(\Delta\omega) \tag{14.11.27b}$$

where $g(\Delta\omega)$ is the distribution of center frequencies or detuning. If we recall (14.8.14) and (14.8.15), we see that χ'' (which is represented by V) is an even function of the detuning $\Delta\omega = \omega_{21} - \omega$, whereas χ' (represented by U) is an odd function. We anticipate this same general behavior here (and can check for consistency later on). This logic path leads to

$$\langle U \rangle = 0 \tag{14.11.28a}$$

(but does not mean that $U = 0$) and thus (14.11.9) is solved by

$$\phi(z, t) = 0 \tag{14.11.28b}$$

This eliminates one of the equations facing us, and we now turn to the pulse solution.

14.11.4 Pulse Solution

If the conclusions of the previous section are correct – the pulse with area $= 2m\pi$ propagates without distortion or attenuation – then all quantities U, V, W, and E must be functions of a common argument, a local time given by

$$T = t - \frac{z}{v} \tag{14.11.29}$$

where the velocity v is to be determined. Hence,

$$\frac{\partial f}{\partial t} = \frac{df}{dT} \quad \text{and} \quad \frac{\partial f}{\partial z} = \frac{df}{dT}\frac{\partial T}{\partial z} = -\frac{1}{v}\frac{df}{dT} \tag{14.11.30}$$

[If there were no atoms present, then the field would propagate with a velocity of c/n (as usual) by (14.11.9a) with the other equations being solved by the trivial solution of $U = V = W = 0$.] With this transformation of variables, our basic equations become a bit simpler and ordinary differentials can be used.

$$\frac{dE}{dT} = \frac{\omega_0 c \mu_0}{2n\left(\frac{1}{v} - \frac{n}{c}\right)} \langle V \rangle \tag{14.11.31a}$$

$$\frac{dU}{dT} = (\Delta\omega) V \tag{14.11.31b}$$

$$\frac{dV}{dT} = -(\Delta\omega)U + \left(\frac{\mu E}{\hbar}\right) W \tag{14.11.31c}$$

$$\frac{dW}{dT} = -\left[\frac{\mu E}{\hbar}\right] V \tag{14.11.31d}$$

It may be verified, by direct substitution, that the following are solutions to (14.11.31a) to (14.11.31d)

$$E(z, t) = \frac{2\hbar}{\mu t_p} \operatorname{sech}\left\{\frac{t - z/v}{t_p}\right\} \tag{14.11.32a}$$

$$U(z, t) = 2N\mu \frac{(\Delta\omega)t_p}{1 + (\Delta\omega t_p)^2} \operatorname{sech}\left\{\frac{t - z/v}{t_p}\right\} \tag{14.11.32b}$$

$$V(z, t) = -2N\mu \frac{1}{1 + (\Delta\omega t_p)^2} \tanh\left\{\frac{t - z/v}{t_p}\right\} \operatorname{sech}\left\{\frac{t - z/v}{t_p}\right\} \tag{14.11.32c}$$

$$W(z, t) = N\mu\left[1 - \frac{2}{1 + (\Delta\omega t_p)^2}\right] \operatorname{sech}^2\left\{\frac{t - z/v}{t_p}\right\} \tag{14.11.32d}$$

and

$$\frac{1}{v} = \frac{n}{c} + \frac{\alpha t_p^2}{2\pi g(0)} \int_{-\infty}^{+\infty} \frac{g(\Delta\omega)}{1 + (\Delta\omega t_p)^2}\, d(\Delta\omega) \tag{14.11.33}$$

where the characteristic pulse width t_p is arbitrary, but must be much less than T_2 or τ. Thus a $\operatorname{sech}(t/t_p)$ is a "natural" pulse shape for the electric field and a $\operatorname{sech}^2(t/t_p)$ for the intensity.

A big surprise is the drastic reduction of the velocity of propagation. If we assume a simple bell-shape curve for $g(\Delta\omega)$, a Lorentzian for mathematical convenience (but not to be construed as homogeneously broadened lineshape), then the distribution of detunings is given by

$$g(\Delta\omega) = \frac{(\Delta\omega_a/2)}{\pi\left\{(\Delta\omega)^2 + (\Delta\omega_a/2)^2\right\}} \tag{14.11.34}$$

The integral in (14.11.33) can be evaluated in terms of elementary functions. Multiply and divide the integral by t_p^2:

$$\text{integral} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{(\Delta\omega_a t_p/2)}{(\Delta\omega t_p)^2 + (\Delta\omega_a t_p/2)^2} \cdot \frac{d(\Delta\omega t_p)}{1 + (\Delta\omega t_p)^2} \tag{14.11.35}$$

Let $x = \Delta\omega t_p$, $a = \Delta\omega_a t_p/2$, and use a partial fraction expansion.

$$\text{integral} = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{a\,dx}{(x^2 + a^2)(x^2 + 1)} = \frac{a}{\pi(a^2 - 1)}\left[\int_{-\infty}^{+\infty} \frac{dx}{x^2 + 1} - \int_{-\infty}^{+\infty} \frac{dx}{x^2 + a^2}\right]$$

$$= \frac{1}{(a + 1)} = \frac{2}{(\Delta\omega t_p + 2)} \tag{14.11.36}$$

Now $g(0) = 2/(\pi \Delta\omega_a)$. Hence (14.7.33) becomes

$$\frac{1}{v} = \frac{n}{c} + \frac{\alpha t_p}{2}\left[\frac{\Delta\omega_a t_p}{\Delta\omega_a t_p + 2}\right] \tag{14.11.37}$$

There are obviously two natural limits to be considered. If $\Delta\omega_a t_p \gg 1$ – a broad transition – then (14.11.37) becomes

$$\frac{1}{v} = \frac{n}{c} + \frac{\alpha t_p}{2} \tag{14.11.38}$$

Example

Let us use some typical data for neon, say, the 5944 Å transition in neon ($2p_4 \rightarrow 1s_5$) which has a radiative lifetime of 19.53 ns $= \tau/2$ and which is Doppler broadened with $\Delta\nu_D \approx 1.5$ GHz. The A_{21} coefficient for the transition is $10.5 \times 10^{+6}$ sec^{-1}. (See Table 10.5 for data). The lower state is metastable, and hence we do not normally have an inversion, and a typical density of this lower state is 10^{12} cm^{-3}. Thus, the absorption coefficient is

$$\alpha = A_{21}\frac{\lambda^2}{8\pi}\left\{\left(\frac{4\ln 2}{\pi}\right)^{1/2}\frac{1}{\Delta\nu_D}\right\}N_1 = 0.92 \text{ cm}^{-1}$$

If the pulse width $t_p = 5$ ns, then

$$\frac{1}{v} = \frac{1}{3 \times 10^{10}} + 2.206 \times 10^{-9} = 2.24 \times 10^{-9} \text{ sec/cm}$$

or

$$v = 4.47 \times 10^{+8} \text{ cm/sec} \sim c/67$$

In other words, the velocity of pulse is roughly 1/67 of that of propagating in free space. Note that $\Delta\omega_a = 2\pi\Delta\nu_d$ (approximating the Doppler-broadened Gaussian line by an equivalent Lorentzian) and is equal to $9.4 \times 10^{+9}$ rad/sec. Thus, $\Delta\omega_a t_p = 47$, which is quite adequate for the limits of (14.11.38). The pulse width is considerably shorter than the population inversion relaxation time of 39 ns, and is smaller than T_2 if the pressure is low enough. For the other extreme, we should back up and include the relaxation terms in the density matrix, a mathematical complexity that we leave for more advanced books.

We can also ask how intense a pulse is required. We use (14.3.11) to relate A_{21} to $|\mu_{21}|^2 = 3|\mu_{21x}|^2$ and find that $\mu_{21x} = 5.1 \times 10^{-30}$ (c–m). Thus, the peak field of the sech(t/t_p) pulse need only to be 82.7 V/cm or we need a peak power of 9.1 W/cm^2. The energy in the pulse in only a paltry 90.6 nJ. In contrast, the bleaching theory of Sec. 9.6 would suggest that there is only a small reduction in the attenuation of 0.92 cm^{-1} when the energy (per unit of area) was $h\nu/2\sigma = 182$ nJ/cm^2 nearly twice that required for the propagation of a 2π pulse with no attenuation.

It should be clear, then, that there are cases that the rate equation approach will not yield the correct answer. The danger flag goes up when the envelope of the field is fast compared to the relaxation time T_2. The rate equation approach will not handle multiphoton and super-fast phenomena such as that discussed here, but since it is much simpler and appealing to one's intuition, it should always be done first.

PROBLEMS

14.1. Use the perturbation theory of Sec. 14.3 to compute the A coefficient for the $ep \rightarrow 2s$ transition in hydrogen. (Ans.: $A = 2.245 \times 10^7\ s^{-1}$.) What is the wavelength of the radiation? (Ans.: $H_\alpha = \lambda_0 = 6562.86$ Å.)

14.2. Consider a particle bound in a one-dimensional potential well [i.e., $V(x) = 0$; $0 < x < d$; $V(x) - \infty$ otherwise]. Compute the transition probabilities between the quantum levels.

14.3. The density matrix formulation of the field-atom interaction yielded the following expression for the imaginary part of the susceptibility χ'', assuming homogeneous broadening.

$$\chi'' = \frac{(\mu_{21})^2 T_2}{3\epsilon_0 \hbar} \cdot \frac{\Delta N^0}{1 + (\omega_{21} - \omega)^2 T_2^2 + 4\Omega^2 T_2 \tau}$$

where

$$\Delta N^0 = N(\rho_{11} - \rho_{22})^0, \text{ which corresponds to } N_1 - N_2$$

$$\Omega = \text{Rabi flopping frequency}$$

$$(\mu_{21x})^2 = (\mu_{21})^2/3$$

Show that this equation and 13.3.5 lead to the same expression for the (saturated) gain coefficients as was derived from the rate equations (8.3.15) provided one uses the correct interpretation of the factors T_2 and τ. In particular, manipulate the expression to show that it contains

(a) The homogeneous line width $\Delta\nu_h$

(b) The Lorentzian line shape

$$g(\nu) = \frac{\Delta\nu_h}{2\pi[(\nu_0 - \nu)^2 + (\Delta\nu_h/2)^2]}$$

(c) The stimulated emission cross section $\sigma(\nu_0) = A_{21} \cdot (\lambda^2/8\pi^2) \cdot (2/\Delta\nu_h)$

(d) The saturation intensity I_s and the homogeneous saturation law

(e) The "hole" width $\Delta\nu_H$

(f) The coefficient for spontaneous emission A_{21} (14.3.11)

14.4. The frequency dependence of the real part of the complex susceptibility for a homogeneously broadened line is given by

$$\frac{\chi'}{n^2} = A_{21} \cdot \frac{\lambda_0^3}{16\pi^2 n^3} \cdot \left[\frac{g_2}{g_1} N_1 - N_2\right] \cdot \frac{\nu_0 - \nu}{\pi\left[(\nu_0 - \nu)^2 + \left(\frac{\Delta\nu_h}{2}\right)^2\right]}$$

If the distribution of center frequencies for a Doppler line were approximated by a "square" distribution:

$$\rho(f) = \frac{1}{\Delta\nu_s} \qquad \text{for} \quad \nu_0 - \Delta\nu_s/2 < f < \nu_0 + \Delta\nu_s/2$$

find the susceptibility for the Doppler-broadened transition using these approximations.

14.5. The experimental data presented in the paper by P.W. Smith "Mode Selection in Lasers," *Proc. IEEE*, p. 428, Fig. 13, 1972, on the homogeneous line width of a He^3:Ne^{20} laser line at 6328 Å can be represented by an equation of the form:

$$\Delta\nu_h(\text{MHz}) = 33.3 + 111.1p$$

where p is the pressure (in torr) of the 7 : 1 gas mixture. Other data for this transition in a typical laser application are as follows: $A_{21} = 6.56 \times 10^6$ sec^{-1}, $A_2 = \Sigma A_{2j} = 12.8 \times 10^6$ sec^{-1}, $A_1 = 51.2 \times 10^6$ sec^{-1}, T=500 K.

(a) Should the graph of $\Delta\nu_h$ go through zero at zero pressure? Was that an experimental error? If not, then what should be the zero pressure intercept?

(b) What is T_2 for a pressure of 3 torr? What is the lifetime of the upper state under these circumstances?

(c) At what pressure would the homogeneous line width be equal to the Doppler width?

(d) What is the electric dipole moment for the $3s_2 - 2p_4$ transition? Express the result in debyes (1 D = 3.33×10^{-30} m).

(e) What is the saturation intensity for a pressure of 3 torr?

(f) What is the Rabi flopping frequency (in hertz) for a stimulating wave with an intensity of 100 W/cm^2?

14.6. The purpose of the following problem is to contrast the simple saturation of absorption predicted by the theory of Sec. 9.6 with that using the Rabi theory for the dynamics of the occupation probabilities of two states. We will consider the propagation of a "square" pulse (width = 1.5 ns) of a dye laser tuned to the center of sodium D line at 5889.95 Å ($3^2S_{1/2} \rightarrow 3^2P_{3/2}$). Let [Na]= 1.5×10^{12} cm^{-3}, a homogeneous line width of 1 GHz, and $A_{21} = 6.3 \times 10^7$ sec^{-1}. Let us first make some preliminary calculations:

(a) What are the degeneracies of the two states?

(b) What is the dipole moment μ_{21x} for the transition?

(c) What is the stimulated emission cross section?

(d) What is the "small" signal attenuation coefficient (in cm^{-1}) and what is the small signal transmission coefficient of a 1 cm path?

Let us first examine the predictions of Sec. 9.6. (Do not forget to account for the different degeneracies.)

(e) What is the electric field of a 1.5 ns square pulse carrying an energy equal to the saturation value?

(f) Assume $w_1 = w_s$. Use the theory of Sec. 9.6 to predict the transmitted energy and compare to the prediction of (d).

Now let us use the theory for the atomic populations.

(g) Evaluate the Rabi frequency (in MHz) for the electric field computed in (e).

(h) Assume that the pulse is intense enough such that $\Omega\tau_p = \pi$. What is the electric field for this pulse?

(i) What is the energy in the pulse of (h)?

(j) Use an argument based upon the conservation of total energy (of the pulse plus that stored as potential energy in the atoms) to show that there is *no* attenuation.

(k) The self-induced transparency theory of McCall and Hahn predicts unity transmission if A = area of a hyperbolic secant pulse (i.e., $E_0 \operatorname{sech}[t/\tau_p]$) equals 2π, where

$$A = \lim_{t\to\infty} \frac{\mu_{21x}}{\hbar} \int_{-\infty}^{t} E(t')\, dt'$$

Evaluate the peak electric field and the energy in such a pulse.

14.7. We can obtain a deceptively simple differential equation for the atomic polarization, some additional insights into the dynamics of the two-level system and an evaluation of the AC Stark effect by the following steps:

(a) Start with the differential equation for ρ_{21}:

$$\frac{\partial \rho_{21}}{\partial T} + \left[\frac{1}{T_2} + j\omega_{21}\right]\rho_{21} = -j\frac{\mu_{21x}E_x(t)}{\hbar}[\rho_{22} - \rho_{11}] \tag{1}$$

(b) Add the complex conjugate of (1) to (1). (2)

(c) Subtract the complex conjugate of (1) from (1). (3)

(d) Differentiate (2) (4)

(e) Use (2) and (3) to eliminate $\rho_{21} - \rho_{21}^*$ in (4) (5)

(f) Multiply by $N\mu_{21x}$, identify P_{ax} and $\Delta N = N_2 - N_1$ to find

$$\frac{\partial^2 P_a}{\partial t^2} + \left[\frac{2}{T_2}\right]\frac{\partial P_a}{\partial t} + \left[\omega_{21}^2 + \left(\frac{1}{T_2}\right)^2\right]P_a = -2\omega_{21}\frac{\mu_{21x}^2 E_x(t)}{\hbar}\Delta N(t) \tag{6}$$

(g) Multiply the differential equation for $(\rho_{22} - \rho_{11})$ by N and substitute the results of (2) and (3) to obtain a differential equation for $[\Delta N - \Delta N^0]$.

$$\frac{\partial[\Delta N - \Delta N^0]}{\partial t} + \frac{[\Delta N - \Delta N^0]}{\tau} = \frac{2\mathbf{E}}{\hbar\omega_{21}} \cdot \left[\frac{\partial \mathbf{P}_a(t)}{T_2} + \frac{\mathbf{P}_a(t)}{T_2}\right] \tag{7}$$

where ΔN^0 is the small signal inversion.

(h) Assume a time variation of the field, polarization, and inversion in the following form: A real field

$$E_x(t) = \frac{E_{0x}e^{j\omega t}}{2} + \frac{E_{0x}e^{-j\omega t}}{2}$$

A real polarization

$$P_{ax}(t) = \frac{P_{0x}e^{j\omega t}}{2} + \frac{P_{0x}^* e^{-j\omega t}}{2}$$

A real density

$$\Delta N(t) = \Delta N_0 + \frac{N_2 e^{j2\omega t}}{2} + \frac{N_2^* e^{-j2\omega t}}{2}$$

where ΔN_0 is steady saturated inversion $\neq \Delta N^0$.

Substitute into the above, equate harmonic terms, and derive an expression for the shift in resonant frequency caused by the square of the optical electric field. This is called the AC Stark effect. (NOTE: It is not necessary to completely determine P_{0x}, but rather identify a term ($\propto E_{0x}^2$) that shifts the resonant frequency from the small signal value.)

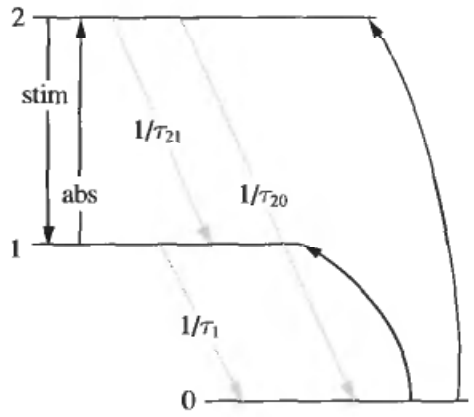
14.8. We can integrate (14.11.22) directly by using a few calculus tricks. Express $\sin A = 2\sin(A/2)\cos(A/2) = 2[\tan(A/2)][\sec^2(A/2)]^{-1}$; separate variables (A, z); substitute $u = \tan(A/2)$; and integrate between A_0 (at $z = 0$) and $A(z)$ to find an explicit formula relating the two areas. What are the different limits of this relationship as $z \to \infty$ for positive and negative α?

14.9. The parameter τ can be interpreted as the time constant for the relaxation of the population *difference* back toward its equilibrium value when stimulated emission ceases. Use the rate equations for the simple three-level model shown on the diagram below to relate τ to τ_1, τ_2, τ_{20}, and τ_{21} by using this definition:

$$\frac{1}{\tau} = \frac{1}{\Delta N}\frac{d(\Delta N)}{dt}$$

$$= \frac{1}{\left\{[N_2 - N_1] - [N_2^0 - N_1^0]\right\}}\frac{d}{dt}\left\{[N_2 - N_1] - [N_2^0 - N_1^0]\right\}|_{\text{stim}=0}$$

Consider two extreme cases as a starting point: $\phi_{21} = 0$ (i.e., no spontaneous decay from 2 to 1) and $\phi_{21} = 1$ (i.e., all of the spontaneous decay from 2 is to 1). Then let ϕ_{21} be arbitrary.



REFERENCES AND SUGGESTED READINGS

1. G. Herzberg, *Atomic Spectra and Atomic Structure*, 2nd ed. (New York: Dover, 1944).
2. A.C.G. Mitchell and M.W. Zemansky, *Resonance Radiation and Excited Atoms* (New York: Cambridge University Press, 1971), especially Chap. 3.
3. R.M. Eisberg, *Fundamentals of Modern Physics* (New York: John Wiley & Sons, 1961), Chap. 9.

4. E. Merzbacher, *Quantum Mechanics* (New York: John Wiley & Sons, 1975), Chaps. 19 and 20.
5. A. Yariv, *Quantum Electronics*, 2nd ed. (New York: John Wiley & Sons, 1971), Chaps. 2 and 3.
6. W.S. Chang, *Principles of Quantum Electronics* (Reading, Mass.: Addison-Wesley, 1969), Chaps, 5 and 6.
7. A. Maitland and M.H. Dunn, *Laser Physics* (Amsterdam: North-Holland, 1969), Chaps. 2 and 3.
8. M.O. Scully and M. Sargent III, "The Concept of the Photon," *Phys. Today*, 38–47, Mar. 1972.
9. A. Matveyev, *Principles of Electrodynamics* (New York: Reinhold, 1966), Chap. 7.
10. G. Herzberg, *Spectra of Diatomic Molecules* (Princeton, N.J.: D. Van Nostrand, 1950).
11. A. van der Ziel, *Solid State Physical Electronics*, 2nd ed. (Englewood Cliffs, N.J.: Prentice Hall, 1968), Chap. 2.
12. R.L. White, *Basic Quantum Mechanics* (New York: McGraw-Hill 1966), Chap. 11.
13. Willis E. Lamb, "Theory of Optical Maser," *Phys. Rev. A134*, A1429–1450, June 15, 1964.
14. See also some of the collected papers in *Laser Theory*, Ed. Frank S. Barnes (New York: IEEE Press, 1972). Part I, Historical Papers, is especially recommended.
15. M. Sargent, III, M. Scully, and W. Lamb, Jr., *Laser Physics* (Reading, Mass.: Addison-Wesley, 1974).
16. R.H. Pantell and H.E. Puthoff, *Fundamentals of Quantum Electronics* (New York: John Wiley & Sons, 1969).
17. R.P. Feynman, R.B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Vol. III (Reading, Mass.: Addison-Wesley, 1965).
18. U-Fano, "Description of States in Quantum Mechanics by Density Matrix and Operator Techniques," *Rev. Mod. Phys. 20*, 74–93, 1957.
19. Y. Kato and H. Takuma, "Experimental Study on the Wavelength Dependence of the Raman Scattering Cross Section," *J. Chem. Phys. 54*, 5398, 1971.
20. W.R. Fenner, H.A. Hyatt, J.M. Kellman, and S.P.S. Porto, "Raman Cross Sections of Some Simple Gases," *J. Opt. Soc. 63*, 73, 1973.
21. W. Kaiser and M. Maier, "Stimulated Rayleigh, Brillouin and Raman Spectroscopy," in *Laser Handbook*, Ed. F.T. Arecchi and E.O. Schultz-Dubois, (Amsterdam: North Holland, 1972).
22. D.S. Knight and W. White, "Characterization of Diamond Films by Raman Spectroscopy," *J. Mater. Res. 4*, 385–393, 1989.
23. S.L. McCall and E.L. Hahn, "Self-Induced Transparency by Pulsed Coherent Light," *Phys. Rev. Lett. 8*, 908, 1967.
24. R. Loudon, *The Quantum Theory of Light* (Oxford: Clarendon, 1983).



Spectroscopy of Common Lasers

15.1 INTRODUCTION

This chapter briefly reviews the notation commonly used to describe the quantum state of atoms and molecules. Most students will have had some introduction to atomic spectra so only a brief résumé will be given here. However, many will not have had an exposure to molecular vibrational-rotational structure; hence, a more detailed explanation is given. In any case, our goal is to learn the *language* of spectroscopic notation in a minimum of time with a minimum of mathematics. For a more detailed explanation of the underlying principles, one should consult Refs. 1 through 4.

This chapter will provide you with more than a set of dry rules, regulations, selections, rules and formulas. We will also try to examine those issues about various lasers that can only be appreciated after the spectroscopy is understood. Sections devoted to those special issues are marked with an asterisk (*).

15.2 ATOMIC NOTATION

15.2.1 Energy Levels

With the exception of atomic hydrogen (and possibly helium), an atom composed of z protons and z electrons is much too complicated to expect an analytic solution for the

quantum levels. However, various perturbation and coupling schemes have evolved that have proved to be remarkably accurate in predicting the trends and rules of the spectra. Actually, the process worked in reverse order: The energy levels and facts about the spectra were known from experiment, which, in turn, led to schemes reproducing the known answer.

One of the most successful is the LS (or Russel–Saunders) coupling scheme, in which a quantum state of an atom is labeled in the following manner:

$$Nl^k \quad {}^{2S+1}L_j \tag{15.2.1}$$

where L is the total orbital angular momentum quantum number, S the total spin angular momentum, and J the magnitude of the total angular momentum. $\mathbf{J} = \mathbf{L} + \mathbf{S}$ according to the vector model covered in Herzberg [1]. By convention, letter symbols are used for L according to the following scheme:

Letter	S	P	D	F	G	H	I	$\cdots$
Value of L	0	1	2	3	4	5	6	

The number N denotes the orbit number (in the Bohr sense) and l the angular momentum states of the last k active electrons. Quite often l^k is omitted.

For instance, the ground state of mercury is 6^1S_0 and the first five excited states are 6^3P_0, 6^3P_1, 6^3P_2, 6^1P_1, and 7^3S_1. According to the table above, we have the following information provided:

$6^1S_0 : N = 6, L = 0, J = 0, S = 0$ (read "six singlet-S-zero")

$6^3P_0 : N = 6, L = 1, J = 0, S = 1$ (read "six triplet-P-zero")

$6^3P_1 : N = 6, L = 1, J = 1, S = 1$ (etc.)

$6^3P_2 : N = 6, L = 1, J = 2, S = 1$

$6^1P_1 : N = 6, L = 1, J = 1, S = 0$

$7^3S_1 : N = 7, L = 0, J = 1, S = 1$

In an energy-level diagram, states with different multiplicities, $2S + 1$, are separated (along the $L = 0$ or the S column) and various states with common L are arranged in columns. Figure 15.1, for mercury, illustrates these conventions. Also shown are some transitions that illustrate the selection rules discussed below.

15.2.2 Transitions: Selection Rules

The theory behind the selection rules is buried in reams of mathematics—see Chapter 14 for a small part of it—and for every "rule" there appears to be an "exception."* However, a

*The word "exception" is not really appropriate, but it fits the situation for now. The rules are given for electric dipole transitions, whereas the exceptions do not fit that category. See the discussion following (15.2.3).

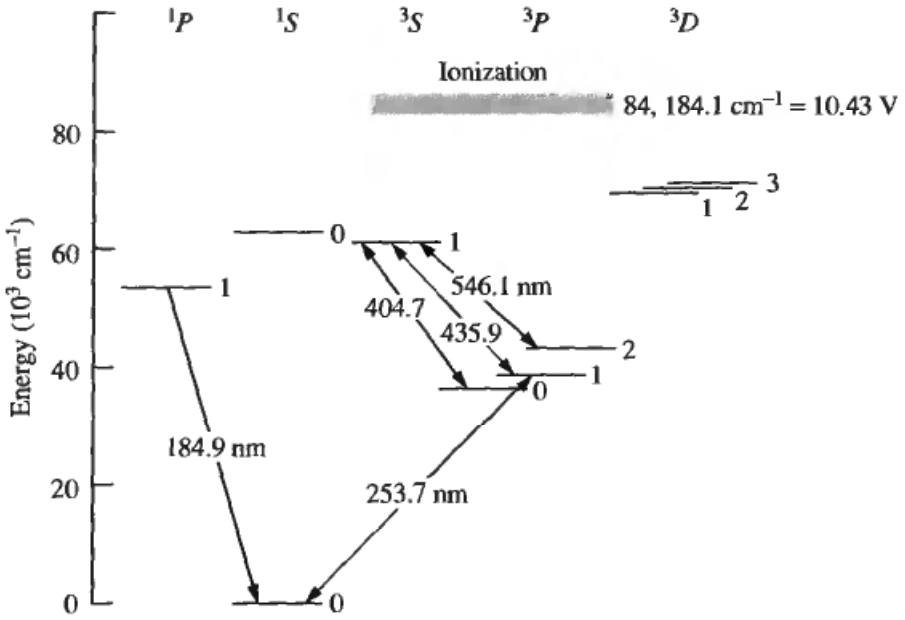


FIGURE 15.1. Partial energy-level diagram for mercury. [Data from Moore [5], (vol. 3).]

good starting point is to commit the following three rules to memory:

$$\Delta J = \pm 1, 0 \qquad \text{but} \qquad J = 0 \nleftrightarrow J = 0 \tag{15.2.2}$$

Thus J can change by ± 1 or 0 (but not 2, 3, etc.), and transitions between states with $J = 0$ are forbidden. It is a "rule" that has few exceptions.

$$\Delta L = \pm 1 \tag{15.2.3}$$

A well-known laser transition is iodine, $5^2P_{1/2} \rightarrow 5^2P_{3/2}$, at $\lambda = 1.315\ \mu$m violates this rule, since $\Delta L = 0$. Such a transition is a magnetic dipole transition.

$$\Delta S = 0 \tag{15.2.4}$$

In other words, transitions involving a spin change are forbidden. This rule seems to be the first to fall. In the light elements such as helium, it is rigorously obeyed, but more and more deviations are found in the heavier elements. For instance, the 253.7-nm transition in mercury violates this rule and yet is one of the strongest lines in the spectra. Indeed, it is that transition that is used to excite the phosphor in a fluorescent lamp.

As an example, the $6^3P_1 \rightarrow 6^1S_0$ transition is allowed according to (15.2.2) ($\Delta J = +1$), is allowed according to (15.2.3) ($\Delta L = +1$), but is forbidden according to (15.2.4). Obviously the transition does take place, proving that atoms cannot read or obey rules. Note that two of the first three excited states, 6^3P_2, and 6^3P_0, do not have a radiative transition back to the ground state. For the first, $\Delta J = 2$, and $J = 0 \rightarrow J = 0$ for the second. Such states are called metastable.

15.3 MOLECULAR STRUCTURE: DIATOMIC MOLECULES

15.3.1 Preliminary Comments

When two atoms are combined to form a diatomic molecule, such as H_2, N_2, or CO, the system has quantum states associated with *rotation* and *vibration* in addition to the *electronic* structure. In other words, the atoms may vibrate with respect to each other and rotate about an axis, all while the electrons in the combined system can undergo changes also.

To a first approximation, assume that these three types of "motion" are independent of each other. To be more precise, assume that the wave function can be factored into a product of rotational, vibrational, and electronic wave functions. The advantage of this approximation is that each motion can be treated separately and then combined in the end. This also implies that the energies associated with the different types of motion are additive.

If two atoms, A and B, are separated by a long distance, they retain their separate identities with their associated electronic structure and literally ignore one another. If we move them closer together, various forces come into play to either attract or repel the other atom. At last, the horizontal axis on an energy-level diagram means something—we plot this interaction potential as a function of separation of the nuclei, as shown in Fig. 15.2. In Fig. 15.2, two different interaction potentials are shown: The solid curve represents the potential energy for a bound molecular state, and the dashed one represents a repulsive state.

As shown in this figure, the minimum energy is when the two atoms are separated by a distance r_e. As indicated in Fig. 15.2, the value of this minimum depends on whether one of the atoms is in an electronic excited state; for now, let us focus on the case where both are in the ground state. While the two atoms are in the potential well, we have a bound diatomic molecule, with a binding—called the dissociation energy—of the difference between $V(r = \infty)$ and the lowest allowed energy state near the "equilibrium" distance r_e.

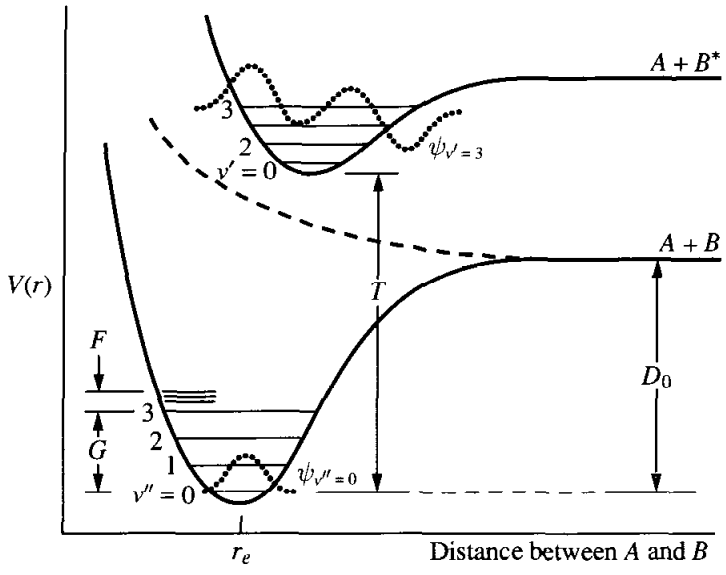


FIGURE 15.2. Potential-energy diagram for a hypothetical molecule AB. By convention, the vibrational energy is denoted by the letter symbol G, rotational energy by F, and electronic energy by T. For the CO molecule, $r_e = 1.128322$ Å.

As soon as we start talking about distances of a few angstroms, typical of r_e, it is not possible to identify a fixed position as an equilibrium position of two atoms, for this would require zero velocity. The uncertainty principle forbids such absolute determinacy. Rather, we must solve for energy levels in this more or less parabolic potential well. Every elementary text in quantum mechanics solves that problem, so we will just use one of the major results of that analysis:

> The energy levels are more or less uniformly spaced in the well, with the minimum energy level lying above the minimum of the well.

These levels are the vibrational levels that correspond to a classical vibration of the two nuclei bound together by a spring. If the well were exactly parabolic, the energy levels would be spaced by a multiple of $\hbar$ times the vibration frequency, ω_{vib}, and the minimum would be $(\hbar\omega_{\text{vib}}/2)$ above the minimum of the parabola. However, it is not a parabola, so some modification must be made to the above.

Thus we have a bound vibrating molecule. It takes little imagination to recognize that some more energy could be tied up in rotation. Thus in a given electronic manifold, specified by $V(r)$, there is vibrational and rotational energy in addition to the electronic value. We can have transitions of the following types:

1. Rotational: the vibrational quantum number and the electronic state do not change.
2. Vibrational-rotational (VR): the electronic state does not change.
3. Electronic: everything changes.

Pure rotational transitions occur in the far infrared to microwave portion of the spectrum (λ = cm to 15 μm), VR transitions occur in the 20 to 2 μm region, and electronic ones are in the range 1 to 10 eV. In the material given next, we show how to compute the various levels given the molecular data. In view of the complexity of type 3, only the procedure for naming the states will be given.

15.3.2 Rotational Structure and Transitions

The minimum data necessary to specify the rotational energy levels are the rotational constant B_e and how this constant depends on the particular vibrational level to yield B_v.

$$B_v = B_e - \alpha_e\left(v + \tfrac{1}{2}\right) \tag{15.3.1}$$

where v is the vibrational quantum number and B_e and α_e are part of the specification of the molecule. Then the energy level of the rotational state, F, depends on B_v and the angular momentum quantum number J according to*

$$F(J) = B_v J(J+1) \tag{15.3.2}$$

*There is a small correction to (15.3.2) of the form, $-D_v J^2(J+1)^2$, but it is usually neglected. See Herzberg [2] for more detail.

Rotational transitions occur between adjacent states according to $\Delta J = \pm 1$ ($\Delta J = 0$ means that there was not a transition for this case). Labeling this transition by the J value *of the lower state* yields the following formula for the transition frequencies (measured in cm^{-1}):

$$F(J+1) - F(J) = \bar{\nu}(J) = 2B_v(J+1) \tag{15.3.3}$$

Typical values for B_e and α_e (for the ground electronic state of CO) [6] are

$$B_e = 1.931271\ \text{cm}^{-1} \qquad \alpha_e = 0.017513\ \text{cm}^{-1}$$

15.3.3 Thermal Distribution of the Population in Rotational States

It is obvious from these typical numbers that the spacing between rotational levels is very small compared to energies of random motion of the gas molecules ($kT = 208.5\ \text{cm}^{-1}$ at 300° K). Consequently, collisions between molecules are very effective in establishing a thermal population of the rotational states. The population in a rotational state J of vibrational manifold v is dictated by Boltzmann statistics according to

$$\frac{N_{v,J}}{N_v} = \frac{(2J+1)\exp\left\{-[hcB_vJ(J+1)/kT]\right\}}{\sum_{J=0}^{\infty}(2J+1)\exp\left\{-[hcB_vJ(J+1)kT]\right\}} \tag{15.3.4}$$

where $(2J+1)$ is the degeneracy of the quantum state J. The numerator of (15.3.4) is the statistical weight of the rotational state J times the Boltzmann factor, whereas the denominator is the sum of similar factors for all states in this vibrational manifold. Thus a vibrational population N_v is apportioned among many—say, 50 to 100—different rotational states. With very little error, we can approximate the infinite series by an integral:

$$\sum_{J=0}^{\infty}(2J+1)\exp\left\{-[hcB_vJ(J+1)/kT]\right\}$$

$$\to \int_0^{\infty}(2J+1)\exp\left\{-[hcB_vJ(J+1)/kT]\right\}dJ = \frac{kT}{hcB_v} \tag{15.3.5}$$

$$\therefore \qquad \frac{N_{v,J}}{N_v} = \frac{hcB_v}{kT}(2J+1)\exp\left[-\frac{hcB_vJ(J+1)}{kT}\right] \tag{15.3.6}$$

This function is plotted in Fig. 15.3 for the $v = 0$ state in the CO molecule by using the data given previously. Note that there is very little population in the low and very high values of J. Most of the populations appear at energies at $\sim kT$.

At first glance, we might think that there was a built-in population inversion between some of the rotational states, say, between $J = 4$ and $J = 3$. There is, in the sense of strict numbers, but it does not help build a laser. The quantity that counts insofar as the laser gain is the density divided by the degeneracy:

$$\gamma = \sigma\left(N_2 - \frac{g_2}{g_1}N_1\right) = \sigma g_2\left(\frac{N_2}{g_2} - \frac{N_1}{g_1}\right)$$

This difference is always negative for all values of J for a thermal distribution.

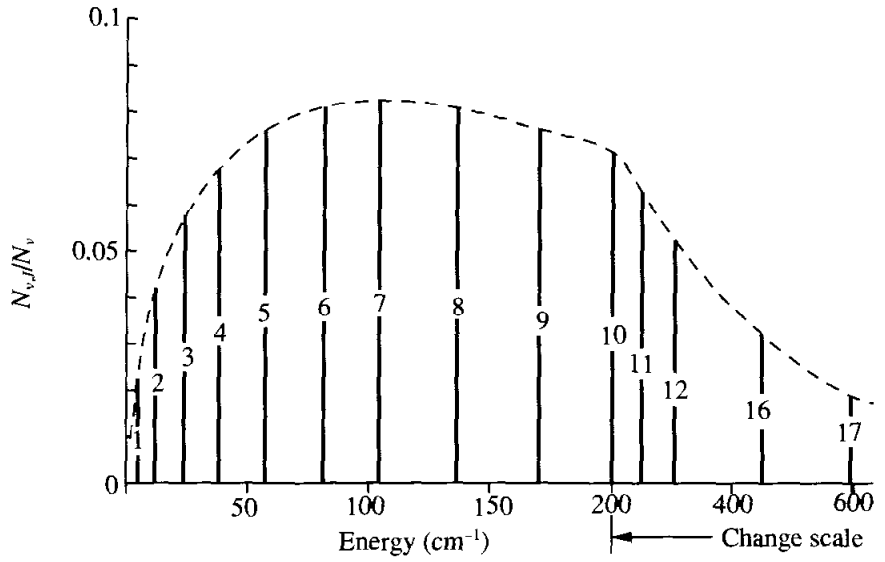


FIGURE 15.3. Rotational distribution of the population in the $v = 0$ state of CO at 300° K.

15.3.4 Vibrational Structure

The vibrational energy level $G(v)$, can be computed from

$$G(v) = \omega_e \left(v + \tfrac{1}{2}\right) - \omega_e x_e \left(v + \tfrac{1}{2}\right)^2 + \omega_e y_e \left(v + \tfrac{1}{2}\right)^3 + \cdots \qquad (15.3.7)$$

where ω_e, $\omega_e x_e$, and so on are all measured constants of a molecule in a particular electronic state. In the ground electronic state of CO, these constants [5] are

$$\omega_e = 2169.8233 \text{ cm}^{-1}$$

$$\omega_e x_e = 13.2939 \text{ cm}^{-1}$$

$$\omega_e y_e = 1.57 \times 10^{-5} \text{ cm}^{-1}$$

There is a temptation to factor ω_e from the expression, but since the products, $\omega_e x_e$ and $\omega_e y_e$, . . ., are always specified, it is a futile exercise.

We should recognize (15.3.7) as a Taylor series expansion that converges very rapidly, inasmuch as the higher-order coefficients decrease dramatically. The first term is the energy level of a particle in a parabolic potential well, and the remainder is the correction for the fact that $V(r)$ in Fig. 15.2 is not a parabola. The quantity $c\omega_e$ is roughly the vibrational frequency of the classical mass-spring system. Thus the heavier the atoms, the smaller the vibrational frequency. For example, ω_e for molecular hydrogen is 4395.2 cm^{-1}, that of D_2 is 3118.4 cm^{-1}, and the ratio is $2^{1/2}$ to within 0.3%.



The total energy of a rotation state J in a vibrational manifold v is given by the sum (15.3.7) and (15.3.2):

$$E(v, J) = G(v) + F(J) \tag{15.3.8}$$

15.3.5 Vibration-Rotational Transitions

These transitions are between rotational states of adjacent vibrational manifolds, according to the following selection rules:

$$\begin{aligned} \Delta v &= \pm 1 \\ \Delta J &= \pm 1, 0, \qquad J = 0 \nleftrightarrow J = 0 \end{aligned} \tag{15.3.9}$$

Transitions $\Delta v = 2$ do occur but are much weaker.

If the J value of the lower state is one greater than the upper, the family of transitions is called the P branch; if it is the same, the family is called the Q branch, and if it is less, it is called the R branch.

$$\begin{aligned} P \text{ branch}: &\quad \Delta J = -1, &\quad J_{\text{upper}} &= J_{\text{lower}} - 1 \\ Q \text{ branch}: &\quad \Delta J = 0, &\quad J_{\text{upper}} &= J_{\text{lower}} \\ R \text{ branch}: &\quad \Delta J = +1, &\quad J_{\text{upper}} &= J_{\text{lower}} + 1 \end{aligned} \tag{15.3.10}$$

In most diatomic molecules, the Q branch is *not* present in the VR spectra (NO is the only exception). In homonuclear diatomic molecules, such as H_2, N_2, O_2, and so on, none of the VR transitions are observable. The classical reason for this fact is that no amount of stretching of two identical molecules can form an electric dipole. Consequently, *all* vibrational levels of homonuclear molecules are metastable. That fact plays a major role in the spectacular performance of the CO_2/N_2 laser.

It is more a matter of perseverance than intelligence to apply (15.3.9) to compute the wave number of the transitions. If we label the transition by the lower-state J values, we have

$$\bar{\nu} = G(v') - G(v'') + F(J') - F(J'')$$

For the P branch, $J' = J'' - 1$, or

$$\bar{\nu}_{P(J)} = \bar{\nu}_0 - (B_{v'} + B_{v''})J + (B_{v'} - B_{v''})J^2 \tag{15.3.11}$$

For the R branch, $J' = J'' + 1$, or

$$\bar{\nu}_{R(J)} = \bar{\nu}_0 + 2B_{v'} + (3B_{v'} - B_{v''})J + (B_{v'} - B_{v''})J^2 \tag{15.3.12}$$

where the single prime represents the upper state, the double prime represents the lower, and $\bar{\nu}_0$ is the "center" of the band ignoring rotation.

$$\bar{\nu}_0 = G(v') - G(v'') \tag{15.3.13}$$

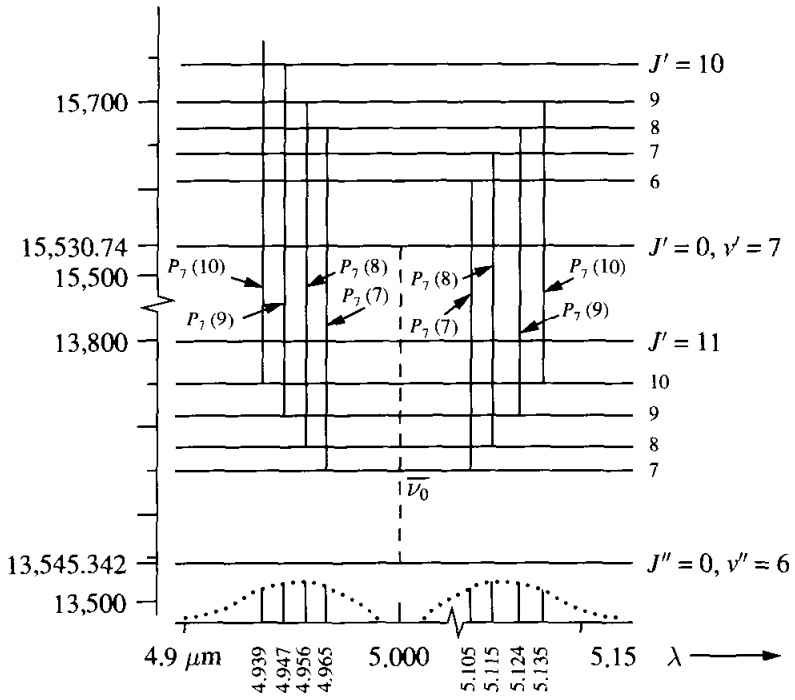


FIGURE 15.4. Example of the VR spectra of CO on the 7 → 6 band.

An example of VR transitions in CO is shown in Fig. 15.4. We also introduce the shorthand notation used in lasers to name a transition: For example, $P_7(10)$ means that it is in the P branch, originates at $v = 7$, $J = 9$ and terminates on $v = 6$, $J = 10$.

15.3.6 Relative Gain on *P* and *R* Branches: Partial and Total Inversions

Whereas gas collisions are very effective in maintaining an equilibrium distribution among rotational states, they are far less so for vibrational states. Indeed, in an electric discharge laser, the population in vibration states tends to equilibrate at the electron temperature, which can be many times (10 to 50) the gas temperature. Moreover, it is quite common to speak of three temperatures—translational, rotational, and vibrational—to describe the distribution of atoms in those states. This leads to some important consequences for lasers:

1. The gain on P branch from a given J level is always higher than the R branch originating at the same upper level.
2. One can always obtain gain on the P branch for some value of J even though the total density of molecules in the upper vibrational state is *less* than that in the lower state provided that $T_v > T_R$. This is called a *partial inversion.*

Both statements can be proved by a patient examination of the laser-gain equation:

$$\gamma(\nu) = \sigma(\nu)\left(N_2 - \frac{g_2}{g_1}N_1\right) = \sigma(\nu)g_2\left(\frac{N_2}{g_2} - \frac{N_1}{g_1}\right) \tag{7.5.2}$$

We identify the upper state 2 with the rotational state J' of the v' vibrational state, J'', v'' with that of the lower state 1, and recall that $g_2 = 2J' + 1$, $g_1 = 2J'' + 1$. Substituting (15.3.6) into the laser-gain equation leads to a rather long and painful expression:

$$\gamma(\nu) = \sigma(\nu)(2J' + 1)\left\{N_{v'} \frac{hcB_{v'}}{kT_r} \exp\left[-\frac{hcB_{v'}J'(J' + 1)}{kT_r}\right] - N_{v''} \frac{hcB_{v''}}{kT_r} \exp\left[-\frac{hcB_{v''}J''(J'' + 1)}{kT_r}\right]\right\} \tag{15.3.14}$$

This equation looks much worse than it is. Let us simplify by ignoring any difference in the rotational constants (i.e., let $B_{v''} = B_{v'} = B$) and expressing the gain for P and R branches in terms of the J value for the *upper state* (even though a transition is usually labeled by the J value of the lower state).

For the P branch, $J' = J$, $J'' = J + 1$:

$$\frac{\gamma_P}{N_{v'}\sigma(\nu)} = \frac{hcB}{kT_r}(2J + 1)\left\{\exp\left[-\frac{hcBJ(J + 1)}{kT_r}\right] - \frac{N_{v''}}{N_{v'}} \exp\left[-\frac{hcB(J + 1)(J + 2)}{kT_r}\right]\right\} \tag{15.3.15a}$$

For the R branch, $J' = J$, $J'' = J - 1$:

$$\frac{\gamma_R}{N_{v'}\sigma(\nu)} = \frac{hcB}{kT_r}(2J + 1)\left\{\exp\left[-\frac{hcB(J)(J + 1)}{kT_r}\right] - \frac{N_{v''}}{N_{v'}} \exp\left[-\frac{hcB(J - 1)(J)}{kT_r}\right]\right\} \tag{15.3.15b}$$

These functions are plotted in Fig. 15.5 for different values of the vibrational ratio $N_{v'}/N_{v''}$. Note that at a common J value of the upper state the gain of the P branch is always higher than that of the R branch, and furthermore that gain can still be achieved on the P branch even when $N_{v'}/N_{v''} < 1$.

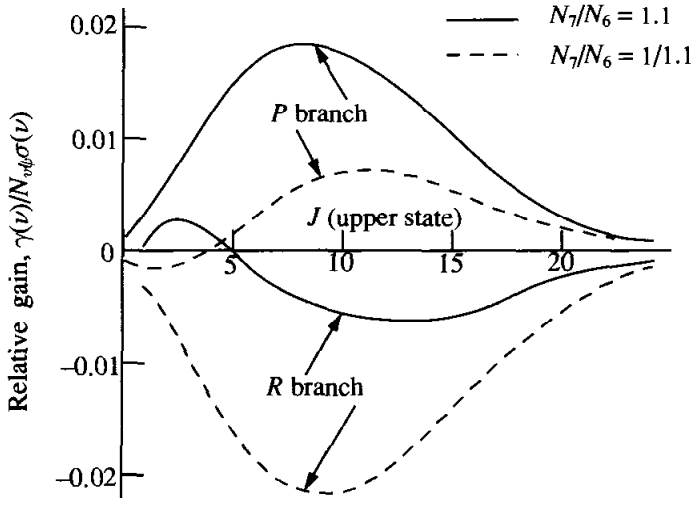


FIGURE 15.5. Relative gain on the $7 \rightarrow 6$ transition of CO for different values of the vibrational population ratio.

If $N_{v'}/N_{v''} < 1$, then gain is obtained over only a fraction of the rotational states, and hence its name, *partial inversion*; and if the ratio is greater than 1, then it is called a *total inversion*. While we can always obtain mathematical gain on high rotational states in the P branch, the value of the gain becomes vanishingly small because of the small population at high J values. Thus, unless special precautions are taken, a laser will always tend to oscillate on the P branch at the expense of the R-branch inversion.

15.4 ELECTRONIC STATES IN MOLECULES

15.4.1 Notation

The notation for the electronic manifold of a molecule is similar to that for an atom, with capital Greek letters replacing the letters. In addition, the *lowest* electronic manifold is given the name X followed by spectroscopic formula for the configuration, with the multiplicity $2S + 1$ preceding the L value as a superscript. For CO, the ground state is $X^1\Sigma$. As indicated, capital Greek letters, Σ, Π, Δ, . . ., are used rather than S, P, D, ...

The excited states with the *same* multiplicity are then given the names A, B, C, . . ., followed by spectroscopic formulas, as we ascend in energy. The excited state with *different* multiplicities are given the names a, b, c, ... The sole exception to this naming procedure is for N_2, where the roles of the capital and lowercase letters are interchanged (see Fig. 15.6).

There is such a rich variety of rules for electronic transitions that even a summary will not be attempted here. Suffice it to say that there are transitions between different vibrational

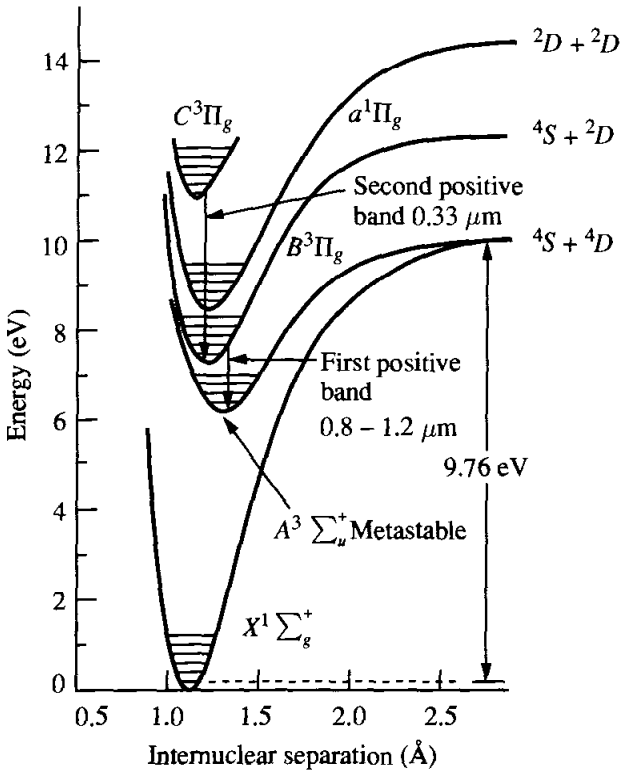


FIGURE 15.6. Energy-level diagram for N_2, showing the common laser transitions.

and rotational quantum numbers within each electronic state and that the vibrational and rotational constants depend on the electronic state. Because of the smallness of the typical rotational constants, there is a transition within every angstrom or so within a band.

15.4.2 The Franck–Condon Principle

There is one issue associated with changes in electronic states of a molecule that is very important but is simple in concept and easy to understand. The Franck–Condon principle states that all electronic processes, such as electronic transitions and electron-molecule inelastic collisions, must take place along a vertical path (i.e., r is a constant).

The classical explanation for this is that the bound electrons can readjust their orbits instantaneously compared to the more sluggish vibrational motion of the nucleii.

Thus the most probable absorption path in Fig. 15.2 is from $v'' = 0$ in the electronic ground state to $v' = 3$ in the excited electronic state. The same is true for inelastic electron collisions. Note that to dissociate the molecule into two free atoms, A and B, we must provide sufficient energy to go vertically from $v = 0$ to the repulsive electronic state (shown as dashed curve in Fig. 15.2). From that point, the two atoms will fly apart, sharing the excess energy.

If the molecule were in $v' = 0$ of the excited state, then it would most probably radiate back to $v'' = 3$, maybe to $v'' = 2$, but not to 0 or 1, because this would require a considerable change in "position" of the molecule during the transition.

This principle is most important in all types of lasers: gas, liquid (dye), and solid state. In the semiconductor lasers, the rule is called the "$\Delta k = 0$" or "k selection rule" and explains why indirect band-gap semiconductors such as Ge and Si do not lase, whereas a direct band-gap material, such as GaAs, does.

15.4.3 Molecular Nitrogen Lasers*

A partial energy-level diagram of molecular nitrogen laser is shown in Fig. 15.6. As noted previously, nitrogen is an exception to the rules for naming the levels, and thus the A, B, and C states have different multiplicities (a triplet) than does the ground state ($X^1\Sigma_g^+$). In a light gas such as N_2, the rule forbidding intersystem crossing is obeyed (i.e., singlet $\leftrightarrow$ triplet), and thus the lowest $A^3\Sigma_u^+$ state is metastable against radiation back to the ground level.

Not only is the lowest laser level metastable, the lifetime of the B state is *longer* than that of the C state. Thus, at first glance, almost everything is wrong, yet it lases rather spectacularly on the $C \rightarrow B$ band.

Part of the reason for its performance in spite of the lifetime odds against it is because of the favorable excitation route from X to C, allowed by the Franck–Condon principle of the preceding section. For instance, most of the excitation in a nitrogen discharge would be from the $v = 0$ vibrational level in the $X^1\Sigma_g^+$ state. Since the "equilibrium" position r_e of the C state is close to that of the ground state, electron impact excitation will proceed along that path rather than to the B or A states (see Table 15.1). For instance, if an electron has 15 eV of energy (sufficient to excite all the states of Table 15.1), then by experiment, excitation to the C state is twice as likely as to the B or A state. (See Ref. 8 for more details.)

TABLE 15.1 Data on N_2

State	$T(\text{cm}^{-1})$	$\omega_e(\text{cm}^{-1})$	$w_e x_e(\text{cm}^{-1})$	$r_e(\text{Å})$	τ
$X^1\Sigma_g^+$	0	2359.61	14.456	1.094	∞
$A^3\Sigma_u^+$	50, 206.0	1460.37	13.891	1.293	seconds
$B^3\Pi_g$	59, 626.3	1734.11	14.47	1.2123	10μs
$C^3\Pi_u$	89, 147.3	2035.1	17.08	1.148	40 ns

However, the real critical issue associated with the nitrogen laser has nothing to do with spectroscopy, atomic physics, stimulated emission, or any other esoteric subject but has most everything to do with the speed of the electrical discharge *circuit*. It has to be capable of switching very high voltages (15 to 40 kV is typical) with a very fast rise time ($\leq$ 10 ns). (That is a nontrivial job.) The speed is required to sustain the electron "temperature" at a high enough value to take advantage of the favorable excitation route allowed by the Franck–Condon principle.

If all the provisos are met, the N_2 lases quite spectacularly in the near UV on the C-B band, in the pulsed mode obviously. Gains are so high (50 to 75 dB/m are typical) that cavities are superfluous. One mirror is used to merely not "waste" the photons from one end.

PROBLEMS

15.1. Refer to Table 10.6. What electronic selection rule forbids the emission on the $3s_2 \rightarrow 2p_9$; $2p_4 \rightarrow 1s_3$? (Ans.: $\Delta J = \pm 1, 0$.)

15.2. Refer to Fig. 10.30. Which selection rule is violated for the 5145-Å and 5287-Å laser lines in the argon ion? (Ans.: $\Delta S = 0$.)

15.3. Why do all rare gases have two metastable states?

15.4. Use Moore [5] to name and evaluate the energies of the metastable states of the rare gases. Present your answer in the following format:

Gas	Spectroscopic Name	Energy Level (cm^{-1})	Energy Level (eV)
He			
Ne			
Ar			
Kr			
Xe			

15.5. Compute and verify the VR spectra of the $7 \rightarrow 6$ spectra of CO shown in Fig. 15.4.

15.6. Make a plot similar to Fig. 15.5. for the $7 \rightarrow 6$ transition in CO for the two different rotational temperatures, $T_r = 150$ K and $T_r = 300$ K. Does the CO laser benefit from cooling? Is this a general rule, or is it peculiar to CO?

15.7. Use Herzberg [2] to find the data necessary to analyze the $3 \rightarrow 2$ band of the HF laser.

15.8. Refer to Table 10.7 for data for the CO_2 laser.

(a) Construct a table of wave numbers for the P-branch transition in the 10.4 μm band.

(b) Suppose that there was a total inversion for this band and that the gain on the $P(22)$ transition was 0.5%/cm. What is the ratio $N_{v'}/N_{v''}$? (Assume Doppler broadening at 300 K.)

15.9. Compute the wavelength of the iodine laser. (Use Moore [5] for energy levels.)

15.10. Fig. 15.5. implies that there is a minimum value of J before positive gain is observed on a VR transition. If we express the vibrational population by a Boltzmann factor of the form

$$\frac{N_{v'}}{N_{v''}} = \exp\left[-\frac{G(v') - G(v'')}{kT_v}\right]$$

then the vibrational temperature T_v can be positive or negative corresponding to the types of inversions.

(a) What is that correspondence?

(b) Show that one can always obtain gain provided that J is larger than $J_{\min}$ given by

$$J > J_{\min} = \frac{G(v') - G(v'')}{2B}\frac{T_r}{T_v} - 1$$

[Eq. (2.11) of Polanyi (Ref. 7)].

15.11. The nominal wavelength of the resonance transition in mercury that excites the phosphor in a common fluorescent lamp is 2537 Å. Actually there are 10 separate lines within 0.05 Å of this nominal value because of the dependence of the center frequency on the isotope and because of the splitting induced by the interaction of the nuclear spins (of the odd isotopes 199 and 201) with the electronic states. The table below shows the mass of the isotope, its relative abundance, the nuclear spin I, the shift from the center of the Hg (198) transition, and the relative contribution (i.e., intensity) of each transition.

These lines are Doppler broadened around their center frequencies corresponding to a gas temperature of 40 C. Construct the spontaneous emission profile and present the results in graph format. The following questions are meant as an aid for the construction of the graph. [See Halstead and Reeves [9] and Anderson et al. [10]]

Isotope	Abundance	Spin I	Shift (cm^{-1})	Intensity
196	0.1%	0	+0.137	0.1%
198	9.89	0	0.0	9.89
200	23.77	0	−0.160	23.77
202	29.27	0	−0.337	29.27
204	6.85	0	−0.511	6.85
199	16.45	1/2(A)	−0.514	5.48
		(B)	−0.160	10.96
201	13.67	3/2(a)	−0.489	6.84
		(b)	−0.023	4.56
		(c)	+0.229	2.28

(a) What is the Doppler width of the Hg (198) transition? Is this width large or small compared to the shifts in the table above? Is it worthwhile worrying about the different widths for each transition?

(b) Plot the spectral distribution of intensity if the lines were a delta function at the respective center frequencies. Identify the spikes by the isotopes and/or the hyperfine component, that is, 201(a) and so on. Are there any natural groupings of transitions that are separated by less than a Doppler width and thus could be considered as one transition? What are the relative strengths of the groups and what are the center frequencies of each?

(c) Use the approximations suggested by (b) to construct a careful graph of the emission profile of the 2537-Å line.

(1) At what wave number (relative to the center of the 198 transition) does the spontaneous emission reach a peak?

(2) What is the FWHM of the composite line shape?

(3) Express the amplitudes of the peaks by a proportionality sequence L:M:N and so on.

REFERENCES AND SUGGESTED READINGS

1. G. Herzberg, *Atomic Spectra and Atomic Structure* (Englewood Cliffs, N.J.: Prentice Hall, 1937; New York: Dover, 1944).
2. G. Herzberg, *Spectra of Diatomic Molecules* (Princeton, N.J.: D. Van Nostrand, 1950).
3. G. Herzberg, *Infrared and Raman Spectra* (Princeton, N.J.: D. Van Nostrand, 1966).
4. E. U. Condon and G. H. Shortly, *The Theory of Atomic Spectra* New York: Cambridge University Press, 1957).
5. C. Moore, *Atomic Energy Levels*, NSRDS-NBS-35, Vols. 1–3. (Washington, D.C.: U.S. Dept. of Commerce, 1971).
6. P. Krupenie, *The Band Spectra of Carbon Monoxide*, NSRDS-NBS-5 (Washington, D.C.: U.S. Dept. of Commerce, 1966).

7. J. C. Polanyi, "Vibrational-Rotational Population Inversion," *Appl. Opt., Suppl. 2 Chem. Lasers*, 109–127, 1965.
8. L. A. Newman and T. A. DeTemple, "Electron Transport Parameters and Excitation Rates in N_2," *J. Appl. Phys. 47*, 1912–1915, 1976.
9. J. A. Halstead and R. A. Reeves, "Time Resolved Spectroscopy of the Mercury 6^3P_1 State," *J. Phys. Chem. 85*, 2777, 1981.
10. J. B. Anderson, J. Maya, M. W. Grossman, R. Laqueskenko, and J. F. Waymouth, "Monte Carlo Treatment of Resonant-radiation Imprisonment in Fluorescent Lamps," *Phys. Rev. A 31*, 2968, 1985.



Detection of Optical Radiation

16.1 INTRODUCTION

Detection of optical radiation is a topic that has a long history, preceding the invention of the laser by many years.* However, the communication capabilities of the laser have spurred renewed interest in making the detectors faster, more sensitive, and more convenient and versatile. This chapter discussed the characteristics of some of the common quantum detectors and the origin of noise in various systems.

We discuss first the manner in which these detectors convert a photon flux into an electrical current and ignore the question of noise. However, the latter question is most important and we will invest considerable effort in describing the origin of the noise. We then return to specific detector classes and discuss their use in a typical signal-processing environment.

16.2 QUANTUM DETECTORS

The term *quantum detector* implies that there is a direct correspondence between the number of photons absorbed and the number of electrical carriers (electrons or holes) generated and

*In fact, Einstein's explanation of the photoelectric effect can be considered as the start of the wide acceptance of the concept of a photon, but detectors go even farther back. Calibrated optical detectors had to be in existence so that the blackbody spectra could be *measured* and *then* explained by Planck.

subsequently used in the circuit. Note that this restriction eliminates thermal detectors* from our consideration.

The ratio of electrical current generated (carriers per second) to photons per second absorbed is the quantum efficiency:

$$\eta_{qe} = \frac{\text{number of carriers generated}}{\text{number of photons absorbed}} \tag{16.2.1a}$$

$$= \frac{i/e}{P_{\text{abs}}/h\nu} = \frac{i/e}{P_{\text{inc}}(1 - \exp[-\alpha l])/h\nu} \tag{16.2.1b}$$

assuming, of course, that every carrier generated is collected by the circuit. In a p-n diode, carriers generated by the absorption of photons beyond a diffusion length of the junction merely recombine and do not move and contribute to the circuit current.

Let us take some examples to illustrate the characteristic of some common detectors.

16.2.1 Vacuum Photodiode

Probably the easiest detector to analyze and to visualize this conversion of a photon flux into an electrical current is the vacuum photodiode shown in Fig. 16.1. Indeed, it was this device that gave convincing proof of the existence of a photon.

If the photon energy $h\nu$ is larger than the photoelectric work function of the cathode, then current will flow; if not *nothing* is obtained irrespective of the bias voltage or the intensity of the incoming beam. (If the intensity is large enough, it can *heat* the photocathode, which will then emit thermionic electrons. Under such extreme circumstances, a quantum detector becomes a thermal one.) A typical photo-response curve is shown in Fig. 16.1(b). If the wavelength is too long, then the photoelectric emission process ceases. This is a function of the cathode material and its preparation; the theory is most involved and is not appropriate to be discussed here. If the wavelength is too short, the transmission of the vacuum window degrades, and this fact accounts for the high-frequency cutoff.

The photons pass through the vacuum window and impinge on the photocathode and generate η_{qe} electrons, which in turn are attracted to and collected by the anode to complete the electric circuit. The bias voltage V_k is usually quite high—say, 300 to 5000 V—to eliminate the possibility that the negative space charge of the emitted electrons would limit the external current. For this same reason, the A-K gap distance is also made as small as practical given the constraint that the output capacitance should also be small. The anode will be of the form of a highly transparent metallic grid (or maybe just a few wires) so that few photons are intercepted by it.

There are a few practical issues brought out in Fig. 16.1(a) and Fig. 16(c). First of all, it is the *current* that is generated by the photons. Hence it is logical to consider a Norton equivalent circuit for this device. The output voltage therefore is a function of the load resistance. The speed of the device is usually determined by the time constant RC_T and not by the transit time of the electrons from cathode to anode, which is exceedingly fast. Sometimes, the cathode is biased negatively and the load is placed in the anode circuit,

*Such as heating water with the photons and measuring the temperature rise with a thermometer. Although such a detector is not of a "high tech" variety, it is easily calibrated.

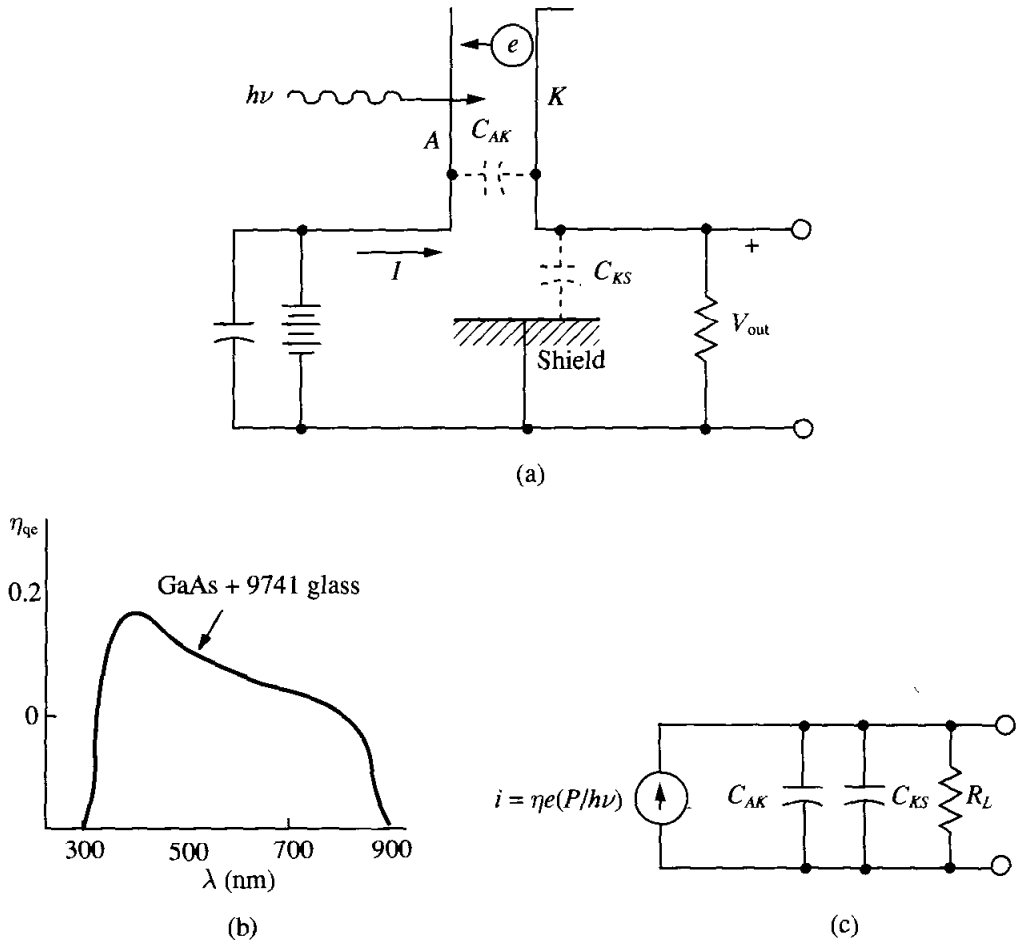


FIGURE 16.1. Vacuum photodiode: (a) bias circuitry; (b) typical variation of η_{qe}; and (c) equivalent circuit.

which eliminates the cathode-shield capacitance from the circuit. Even though the DC current can only flow in one direction, $A \rightarrow K$, we will be concerned with the high-frequency components of the individual electron impulses. Therefore we can also apply this equivalent circuit for the high-frequency component of the impulses.

Vacuum photodiodes have been largely supplanted by the much more convenient solid-state detectors [smaller, faster in some cases, more sensitive, and requiring less formidable power supplies (if at all) for bias]. They still are used when the optical signal is very large, large enough that focusing onto a typical solid-state detector would result in damage. If the voltage is high enough, then amperes of photoelectron current can be obtained. Obviously, the time duration must be small enough so that the tube dissipation does not destroy the anode. However, this section serves to introduce the premium detector of all types—the photomultiplier.

16.2.2 Photomultiplier

The major use of the vacuum photoelectric effect is in conjunction with the most perfect current amplifier in existence, the secondary emission amplifier. This amplifier is based on the experimental fact that many materials will emit, on the average, δ new electrons for every

electron impinging on them. If the kinetic energy of the incident electron is high enough—typically 100 to 200 V—then δ is greater than 1 and we obtain essentially "noise-free" current amplification.

The combination of the vacuum-photocathode technology with anodes (referred to as dynodes) constructed from materials selected to enhance δ yields a photomultiplier in the configuration shown in Fig. 16.2.

If one electron is emitted by the photocathode in response to a photon, it impinges on the first anode, called dynode 1. This element emits δ new electrons, and if the voltages are chosen correctly, then those δ electrons head for the second dynode, which emits δ^2 electrons, and so on, all in response to the initial photo-electron. It is obvious that if we have N dynodes with a secondary emission ratio of δ each, the current gain is

$$G = \frac{i_L}{i_K} = \delta^N \tag{16.2.2}$$

If $N = 12$ and $\delta = 4.0$, then G is huge, 1.68×10^7 (or 144 dB).

It must be emphasized that this gain is essentially "noise free." Detailed considerations of noise will be covered later in the chapter, but note here that *nothing* flows through the output resistor unless that first electron starts down the multiplier chain. That is not true for any other amplifier. For instance, the common AM or FM radio is an amplifier with gains comparable to that of a photomultiplier. The "hiss" heard when it is not tuned to a station is an output with no input; that is, noise. As we will see, that does not happen with a photomultiplier. The photomultiplier is extremely sensitive and can reach the limit of detecting single photons. To illustrate its capabilities consider the following example.

Suppose that a distant source at $\lambda = 5000$ Å reached our receiver (the photomultiplier) with 20 photons per microsecond. The incoming power is thus 7.95×10^{-12} W. If the gain were as computed above and the quantum efficiency of the photocathode were $\eta_{qe} = 0.15$, the load current would be

$$i_L = G\eta e\left(\frac{P}{h\nu}\right) \tag{16.2.3}$$

or 8.06×10^{-6} A, a value that is readily measured. If $R_L = 1M\Omega$, then $V_L = 8$ V, a large signal in response to 7.95 picowatts.

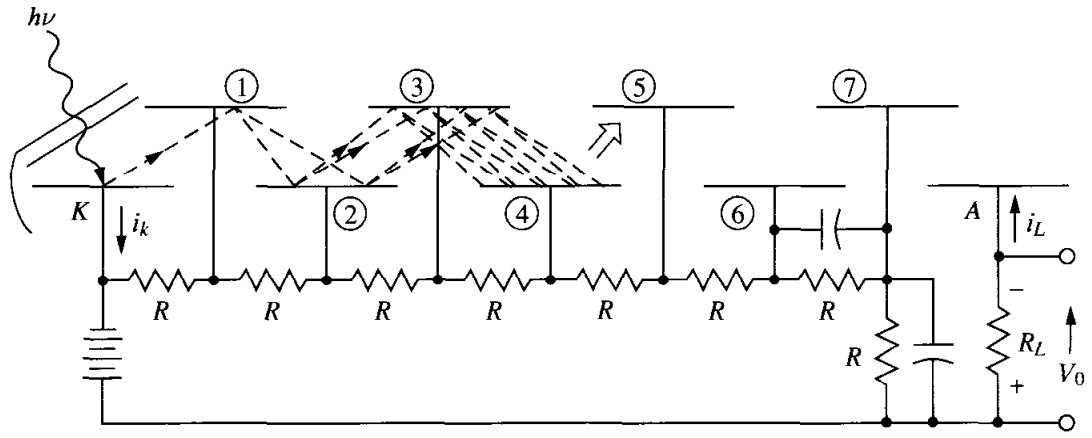


FIGURE 16.2. Seven-stage photomultiplier.

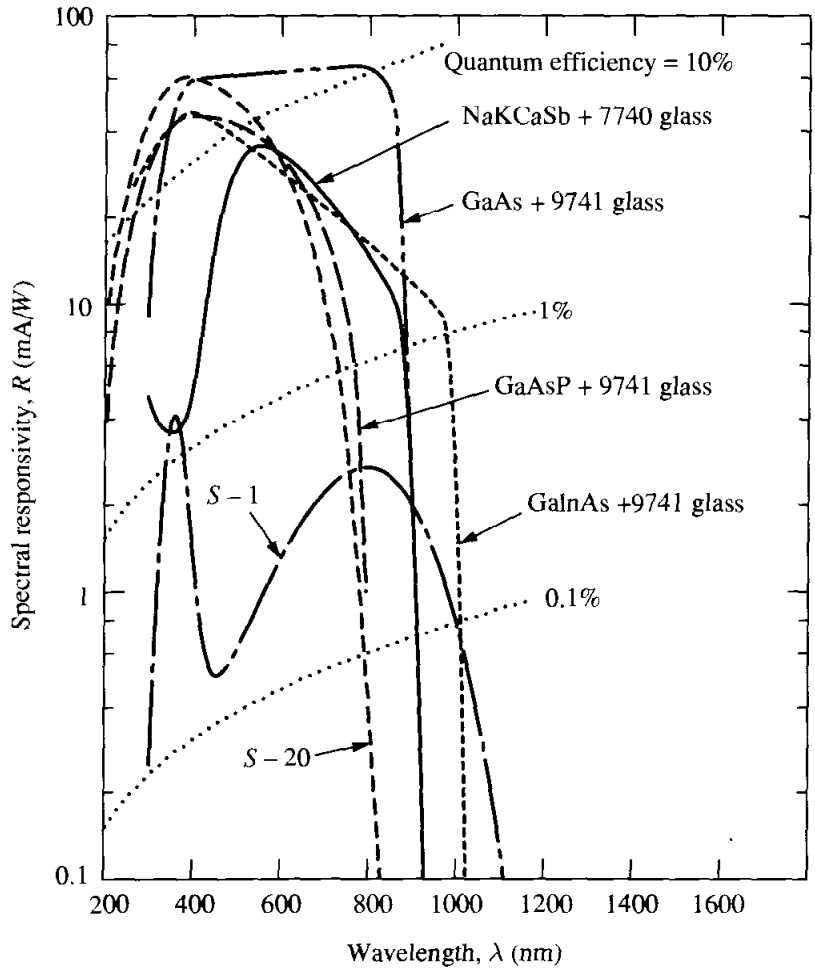


FIGURE 16.3. Typical response curves of photocathode/glass envelope combinations. Note that the response is plotted as current per watt and not as current per photon. (Data from Fig. 10.5 of *RCA Electro-Optics Handbook* [5].)

The dynode resistor chain in Fig. 16.2 is chosen to allot each dynode its proper share of a single power supply voltage. The capacitors are placed on the last few stages, where the dynode currents are the largest, to ensure that the voltage difference remains constant in the presence of a large pulse of photoelectrons.

Many of the problems associated with vacuum photodiodes are also present in photomultipliers; they are large and fragile and require high-voltage power supplies. However, photomultipliers hold a commanding lead in sensitivity. The most serious limitation of both is a fundamental one: The vacuum photoelectric effect ceases for $\lambda \gtrsim 1.1\ \mu$m. This is shown in Fig. 16.3 for various photocathode materials (and glass envelopes). Thus we must turn to solid-state detectors (or thermal ones) for radiation at longer wavelengths.

16.3 SOLID-STATE QUANTUM DETECTORS

16.3.1 Photoconductor

An elementary schematic for a photoconductive element used in a detector circuit is shown in Fig. 16.4. The element is chosen such that its "dark" resistance is much larger than the load R, and thus most of the bias voltage appears across the detector chip. Depending on the

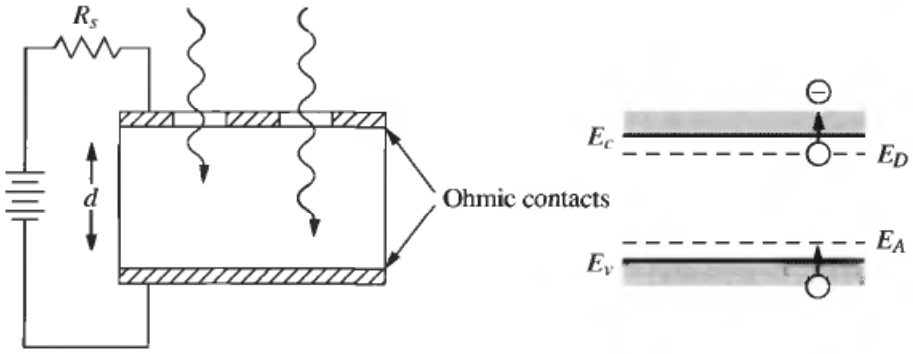


FIGURE 16.4. Photoconductive detector.

type of semiconductor material used and its doping, the incoming radiation might ionize a donor, creating an excess electron in the conduction band for an n-type semiconductor, an excess hole in the valence band by promoting an electron to an acceptor level in a p-type, or an electron-hole pair by band-to-band transitions.

In any case the optical signal changes the number of free carriers in the chip and thus its resistance drops, allowing more of the bias voltage to appear across the load.

Once the excess electron (or hole) is created by the absorption of a photon, it will drift under the influence of the field toward the appropriate contact and current flows in the external circuit. Since this is a majority-carrier conduction process, the other contact emits another carrier as soon as the first is collected. Thus current continues to flow until the carrier recombines with the donor (or acceptor).

If we define τ_0 as being the carrier lifetime, the number of excess carriers ΔN generated in this detector in response to a suddenly applied optical signal is given by a solution to

$$\frac{d\Delta N}{dt} = \eta\left(\frac{P_{\text{abs.}}}{h\nu}\right)u(t) - \frac{\Delta N}{\tau_0} \tag{16.3.1}$$

where η is the quantum efficiency and $u(t)$ the step function. The solution is

$$\Delta N = \eta\left(\frac{P_{\text{abs.}}}{h\nu}\right)\tau_0\left[1 - \exp\left(\frac{t}{\tau_0}\right)\right] \tag{16.3.2}$$

As each carrier moves across the gap d, it conducts a current $qv_d/d = q/\tau_d$, where τ_d is the time for the carrier to cross the gap. Thus the current in the external circuit is

$$i = q\eta\left(\frac{P_{\text{abs.}}}{h\nu}\right)\left(\frac{\tau_0}{\tau_d}\right)\left[1 - \exp\left(-\frac{t}{\tau_0}\right)\right] \tag{16.3.3}$$

Note that the factor τ_0/τ_d can be interpreted as a photoconductive gain. If $\tau_0 \gg \tau_d$, then the peak current is much larger than q times the carrier generation rate.

However, this advantage is a double-edged sword. If the carrier lifetime is long, then the time-response factor in (16.3.3) is slow. If we attempt to preserve a fast time response and gain, the drift time τ_d must be short. But there is a limit here also. The distance cannot be less than the absorption length of the photons. Otherwise, little optical power is absorbed, and thus few excess carriers are generated.

There are other limitations and difficulties with this type of detector, a penalty we must pay for the ability to detect longer wavelengths. If the donor (or acceptor) level-to-conduction (or valence) band gap is small to detect a small photon energy, then these donors are easily ionized by virtue of the thermal environment. For instance, Hg goes into Ge as an acceptor at 0.09 eV, and thus the longest wavelength that it can detect is 14 μm. At room temperature, $T = 300$ K, most the acceptors would be ionized, creating a large quantity of holes and lowering the resistance of the chip. Consequently, most photoconductive detectors are cooled to liquid-nitrogen (77 K) or liquid-helium (4.2 K) temperatures. Thus what started out to be an inexpensive small chip, say 2 mm^2 in area, now requires a bulky Dewar and an expensive and delicate vacuum window* to allow the IR radiation to irradiate the detector element.

16.3.2 Junction Photodiode

Many of the limitations and difficulties previously noted are alleviated by utilizing the highly developed semiconductor technology of p-n junctions with all of the attendant varieties, the simple p-n junction, the p-i-n diode, avalanche photodiode (APD), and the phototransistor. These devices are rugged, sensitive, fast, easily produced, easily biased, and obviously compatible with the concept of integrated optics, where an entire system, receiver, multiplexer, and transmitter are to be placed on a single chip of semiconductor material. Our attention will be limited to the first three types of detectors since these are the workhorses of the industry.

Let us imagine the processes that take place in the formation of the reverse-biased p-n junction photodiode shown in Fig. 16.5. We start with two neutral materials, doped to be p- and n-type, which are to be brought together to form the junction. At the instant of contact, electrons flow from the n region to the p region, leaving behind the immobile donor ions. Similarly, holes flow from p to n, leaving behind the negatively charged acceptors. Thus an internal space-charge field builds up to prevent further gross migration of the carriers.

Note the direction of this space-charge field. If a minority carrier in the n region—a hole—would wander into this space-charge region, then it would immediately be accelerated by this field and the hole would *drift* to the p region. Similarly, if a minority carrier on the p side—an electron—wanders into the space-charge region, it will be accelerated by the field and drift to the n side. Both processes contribute to the "drift" current crossing the junction, which is, of course, balanced at zero bias by an opposite current because of the motion of majority carriers against the retarding field.

The addition of an external reverse bias greatly reduces the motion of majority carriers across the junction. However, the wanderings of the *minority* carriers in the two regions still proceed as usual. Some still have the audacity to diffuse from the neutral material into the space-charge region and are thus swept across the junction. As Streetman ([1], p. 151) notes, the drift current is not so much "how fast the carriers are swept across the junction, but how often" they arrive there to be accelerated. In other words, the drift current is equal

*Most common materials, such as glass, are highly attenuating for wavelengths longer than 3 μm.

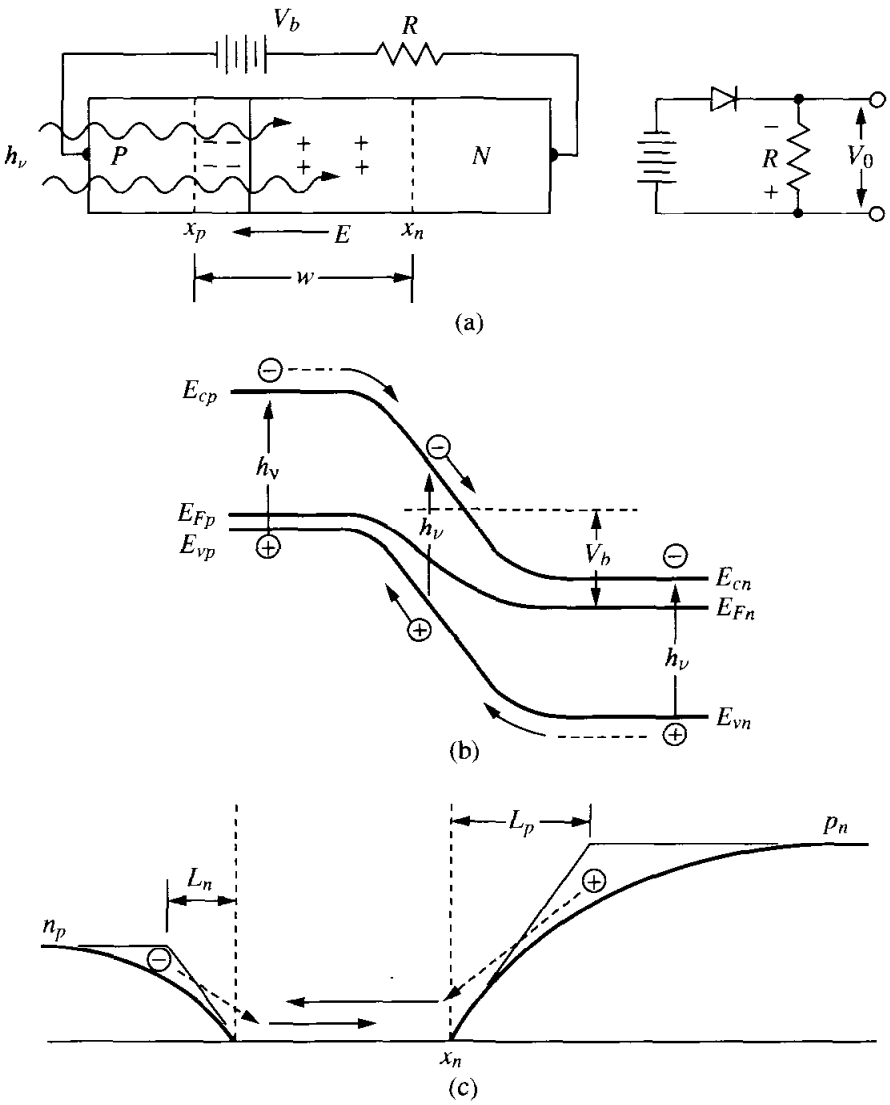


FIGURE 16.5. Solid-state photodiode: (a) reverse-biased *p-n* junction; (b) energy-band diagram; and (c) distribution of minority carriers.

to how often these minority carriers are generated within a diffusion length of the transition region or within that region itself.

If we are provided access to the diode by an optical window, we can change this generation rate by the incoming photons and thus obtain a current proportional to the *absorbed* photons. However, it should be noted that only those electron-hole pairs generated in the transition region or within a diffusion length contribute to the current. If we can neglect recombination in this depletion region, then the current is given by

$$I = e\eta\left(\frac{P_{\text{abs.}}}{h\nu}\right) \tag{16.3.4}$$

Now it is important to remember that only the minority carriers that are created within a diffusion length of the transition region contribute to the photocurrent. Those electron-hole pairs created deeply in the neutral n and p region recombine before they ever diffuse to the

junction to be influenced by the field and hence transport charge. This points out a major failing of the simple p-n junction diode: that one must wait for a relatively slow process, diffusion, to transport a minority carrier to the junction. Thus we could anticipate a rather slow time response if most of the carriers are generated in the neutral regions.

This is especially serious for indirect semiconductors such as silicon, in which the optical absorption coefficient is small. Since the transition width is small compared to the diffusion length, most of the pair generation takes place in the neutral regions, and thus we must depend on this slow process. This difficulty can be avoided by utilizing the p-i-n structure, which is discussed in Sec. 16.3.3.

Even though (16.3.4) appears to be very similar to the multiplicitive factor in (16.3.3), there are major differences in the physics. There is no possibility of photoconductive gain in a diode. Most of the applied bias appears across the depletion layers. Thus only those pairs formed within that region or that can diffuse to that region can be influenced by the field and carry electrical current.

We can utilize the solid-state physics of p-n junctions in a way other than the photoconductive mode described above. For instance, the optical photons generate an excess number of electron-hole pairs in the regions near the junction, and this contribution to the drift current is in addition to that caused by the thermal generation of electron-hole pairs. If the device is open circuited so that the net current is zero, then the built-in field must be reduced to allow more diffusion current to flow. If I_0 is the normal reverse current of the diode as a result of thermal generation of electron-hole pairs and I_{pc} is the photocurrent produced according to (16.3.4), then the open circuit voltage is given by

$$V_{oc} = \frac{kT}{e} \ln\left(\frac{I_{pc}}{I_0} + 1\right) \tag{16.3.5a}$$

This type of operation is called the *photovoltaic* mode and follows directly from the usual diode equation in the presence of optical generation of carriers:

$$I = I_0\left[\exp\left(\frac{eV}{kT}\right) - 1\right] - I_{pc} \tag{16.3.5b}$$

Usually, the photoconductive mode is preferred for a variety of practical reasons. First, note that the current-to-optical power relation is linear for the photoconductive mode (16.3.4) but logarithmic for the photovoltaic one. If we want speed of response, the choice is overwhelmingly in favor of the photoconductive mode.

This is because the depletion-layer capacitance caused by the space-charged layers of Fig. 16.5(a) appears in parallel with the load R_L. If the diode is reverse biased, then $C \propto (V)^{-1/2}$ (for an abrupt junction) and the inherent time constant of the detector $\tau = RC$ is minimized. On the other hand, this capacitance is quite large at the low voltages implied by photovoltaic operation. Incidentally, the voltage given by (16.3.5a) *cannot* be greater than the contact potential ($V_0 \sim E_g$) between p and n, irrespective of the implication that it can by (16.3.5b) (see Streetman [1] for more details). For the photovoltaic mode, we should also use a large R to ensure an open-circuit condition for (16.3.5). Hence, the detector time constant is seriously degraded.

16.3.3 *p-i-n* Diode

In all the equations so far, we have carefully insisted on using the *absorbed* optical power and knowing *where* it is absorbed. Obviously, if the photon passes through the detector, it cannot generate an electron-hole pair. The problem is shown in Fig. 16.6(a), where incoming radiation passes around the contacts through the bulk p region to be absorbed (partially) in and around the depletion region. The part of the radiation that is shadowed by the contacts does nothing, that absorbed deep in the p region and n regions does next to nothing (since the minority carrier so generated has a very small probability of reaching the junction before recombining), and only that absorbed in the depletion layer or within a diffusion length of it is useful. Since the diffusion lengths are usually much larger than the width of the depletion region, most of the useful absorption and pair production are in the neutral region, and we must depend on the slow diffusion process for transport of the current.

We would prefer for the carriers to be generated where the field is large, so that the charge transport (i.e., velocity) is due to the fast drift rather than the slow diffusion. This is accomplished by adding an intrinsic (or at least a high-resistivity) region between the p and n layers, as shown in Fig. 16.6(b). This is called a *p-i-n* diode. Most solid-state detectors for lasers in the visible to near-IR region are of this variety. If the intrinsic layer is thick compared to the optical absorption length, most of the photocurrent will be generated in the intrinsic layer, where the field is largest. Thus any carriers generated there are immediately swept to

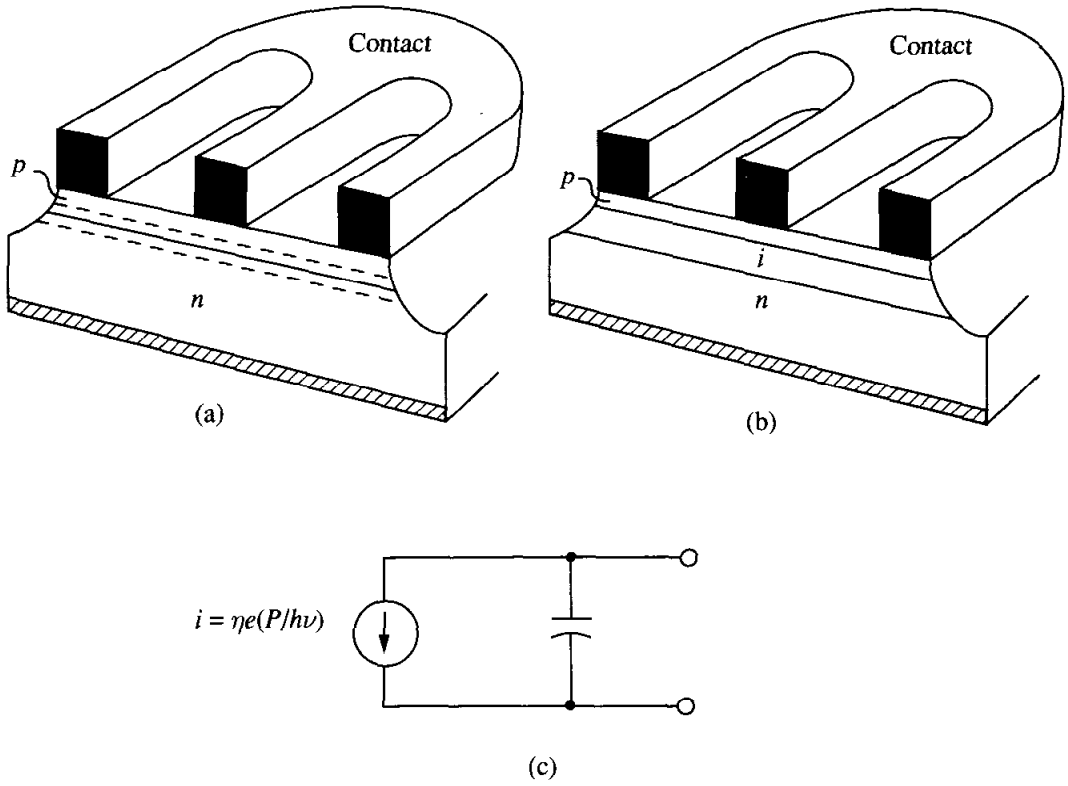


FIGURE 16.6. *p-n* and *p-i-n* diode: (a) *p-n* junction detector; (b) *p-i-n* detector; and (c) equivalent circuit.

the appropriate contact. There are other advantages, also. The intrinsic layer separates the depletion-layer charges of Fig. 16.5(a) and thus lowers the capacitance.

16.3.4 Avalanche Photodiode

We can utilize the fact that the optically produced electrons (and holes) can create secondary pairs as they drift across the junction and thus provide an avalanche of free carriers. Diodes using this effect are called avalanche photodiodes. These second-generation carriers move under the influence of the field and can create a third generation with all generations contributing to the external current. If M new pairs are generated for each primary pair created by the photon, then (16.3.4) must be multiplied by this factor, thus greatly enhancing the sensitivity of the device. Typical values of M range from 20 to 100.

We cannot use arbitrarily large values of M, because, as will be discussed later, this multiplication does contribute excess noise. Furthermore, if the electron-hole pair created by the optical photon avalanches, so does the pair created by thermal processes. Thus thermal runaway is a distinct possibility at high values of M.

16.4 NOISE CONSIDERATIONS

If we insist on large optical signals, there is really no problem with detection. Indeed, a crude standard detector for a high-powered CO_2 laser was the number of firebricks burned per second. Obviously, other "standards" can also be used. In this section, let us make sure that we have identified the problem with the detection of weak optical signals before we pull out our mathematical guns to attack the problem.

For weak signals, there are a whole host of statistical issues that must be addressed. For instance, we have indicated that the photomultiplier is able to detect $\sim$ 20 photons in a microsecond or an optical energy of 8×10^{-12} W. However, if the source is a distant laser, the statistical fluctuation of the number of photons into the solid angle of our detector must be taken into account. Furthermore, the quantum efficiency of the photon cathode, η, and the secondary emission ratio δ are statistical *averages*, not hard and fast numbers. To illustrate this last issue, consider a "Gedanken" experiment, using a computer, to predict the generation of electrons emitted from the cathode.

Assume a source emitting an on-off square wave with a precise number of 20 photons within the envelope of the on time. If the quantum efficiency were 100%, then we would have 20 "spikes" of current, which would then be amplified by the secondary emission amplifier. Since the gain was assumed to be 1.68×10^7 we now have $20 \times 1.68 \times 10^7$ "spikes" of current through our load resistor, each carrying a charge of 1.6×10^{-19} coulombs. If the time interval were 1 μs, this represents an average current of 53.7 μA. If the load resistor were 1 kΩ, then the *average* output voltage would be 53.7 mV, with the "spikes" being much larger. If there were no electrons emitted during the off time, then we would have an unambiguous indication of the presence of these photons. This is shown in Fig. 16.7(a).

The situation becomes less clear for $\eta < 1$ as shown in Fig. 16.7(b) and Fig. 16.7(c). The quantum efficiency is a statistical quantity with the average value as shown on the

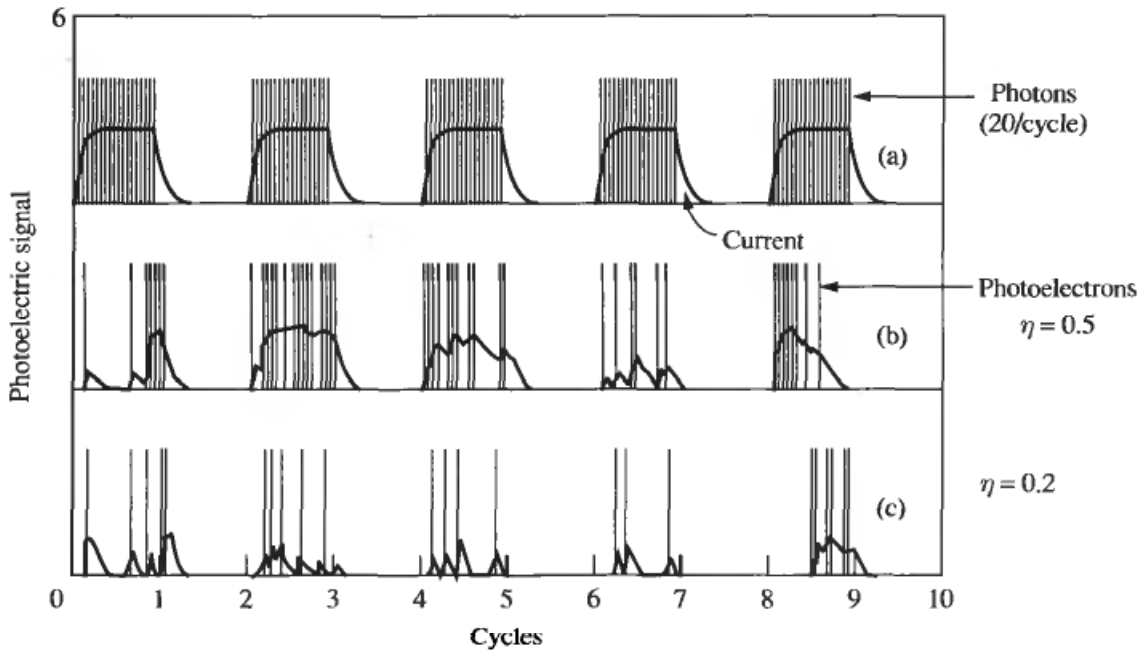


FIGURE 16.7. A computer experiment demonstrating the statistical nature of the conversion of photons into an electrical current (shown solid). See text for added detail.

side. For $\eta = 0.5$, we cannot get half of an electron. We get one half of the time, on the average. As a consequence our nice square-wave envelope became rather ragged because of *the statistical nature of the detection process*. The situation becomes much worse for $\eta = 0.2$. Indeed, the current "spikes" remind us of a noise generator.

Fig. 16.7 was generated by a computer by using its random number generator so as to emit an electron according to the statistics demanded by the quantum efficiency specification. If the source had a statistical fluctuation in number of photons during the on time, or there was thermionic emission from the cathode, then these would also appear as "spikes" in the current to be further amplified by the dynode chain. These spikes could appear anywhere and destroy the infinite contrast of the "data" implied by Fig. 16.7. If the detector were a solid-state device, such as an APD, then the thermal generation of *e–h* pairs would also contribute a statistical current that acts as a fluctuation of the baseline of the output. In any case, note that the pulses do not look the same. Sometimes they appear to replicate the envelope of the photons whereas others are a rather poor representation. Our task is to quantify the concept of a signal-to-noise ratio. But it should be clear that the *statistical nature of converting a photon into an electron generates its own brand of noise* in addition to any statistical fluctuation of the source. It is impossible to assign a partial blame to the detector, and hence both causes are usually lumped into a category called *quantum noise*.

Although we have used the photomultiplier as a convenient example, the same considerations apply to any quantum detector. The only problem with the others is that external "noisy" amplifiers must be used to amplify the output current to the level used here, and these amplifiers contribute their own bit of noise, to muddy the waters even further. But, in

all cases, *each* photoelectron (or hole) produces a *current impulse* of the form

$$i = e\frac{v}{d} = \frac{e}{\tau} \qquad 0 \le t \le \tau = \frac{d}{v} \tag{16.4.1}$$

where v is the free electron velocity of a vacuum photodiode or the drift velocity in a semiconductor and d the distance over which it must travel before being collected, giving a lifetime τ. Please note that τ is quite small, typically less than 10^{-9} sec. Thus these impulses have a nearly "flat" frequency spectral content. This fact plays a major role in the source of noise in optical detectors.

16.5 MATHEMATICS OF NOISE

The most essential bit of mathematics pertinent to noise consideration is the Fourier transform pair, which relates the spectral content of a video signal to its time response. Thus, if $v(t)$ is given, then $V(\omega)$ is related to it by

$$V(\omega) = \int_{-\infty}^{+\infty} v(t)e^{-j\omega t}\,dt \tag{16.5.1}$$

and, of course, the reverse sequence can be followed to find $v(t)$ if $V(\omega)$ is given:

$$v(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} V(\omega)e^{+j\omega t}\,d\omega \tag{16.5.2}$$

In practice, we are forced to measure the signals over a finite sampling interval T, and thus we can assume $v(t)$ to be zero (or, more properly, not observed) outside the interval $-T/2 \le t \le T/2$. We indicate this finite time slot by a subscript:

$$V_T(\omega) = \int_{-T/2}^{+T/2} v(t)e^{-j\omega t}\,dt \tag{16.5.3a}$$

and

$$v(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} V_T(\omega)e^{+j\omega t}\,d\omega \tag{16.5.3b}$$

Since $v(t)$ is real, we also know that $V_T(\omega) = V_T^*(-\omega)$.

Let us suppose that this video signal $v(t)$ is driving an amplifier with an input resistance R and we wish to compute the average power transferred to it. We shall now proceed to show that this power can be computed in the time domain or in the frequency domain by suitable manipulations of (16.5.3a) and (16.5.3b). We first multiply the voltage times the current and average over the sampling time T:

$$\langle P \rangle = \frac{1}{T}\int_{-T/2}^{+T/2} v(t)i(t)\,dt = \frac{1}{R}\left[\frac{1}{T}\int_{-T/2}^{+T/2} v(t)v(t)\,dt\right] \tag{16.5.4a}$$

Substitute (16.5.3b) into (16.5.3a) for the second $v(t)$:

$$\langle P \rangle = \frac{1}{R}\left\{\int_{-T/2}^{+T/2} v(t)\left[\frac{1}{2\pi}\int_{-\infty}^{+\infty} V_T(\omega)e^{+j\omega t}\,d\omega\right]dt\right\} \tag{16.5.4b}$$

Now interchange the order of integration and identify the complex conjugate of $V_T(\omega)$.

$$\langle P \rangle = \frac{1}{R}\left\{\frac{1}{2\pi T}\int_{-\infty}^{+\infty} V_T(\omega)\underbrace{\left[\int_{-T/2}^{+T/2} v(t)e^{+j\omega t}\,dt\right]}_{V_T^*(\omega)}d\omega\right\} \tag{16.5.4c}$$

$$= \frac{1}{R}\left[\frac{1}{2\pi T}\int_{-\infty}^{+\infty} |V_T(\omega)|^2\,d\omega\right] \tag{16.5.4d}$$

Now since $V_T(\omega) = V_T^*(-\omega)$, the quantity $|V_T(\omega)|^2$ must be even with respect to frequency—hence we need only consider positive ω. Thus our principal result is

$$\langle P \rangle = \frac{1}{R}\left[\int_0^{\infty} \frac{|V_T(\omega)|^2}{\pi T}\,d\omega\right] \tag{16.5.4e}$$

The integrand, $|V_T(\omega)|^2/\pi T$, can be interpreted as the energy per radian-frequency interval $d\omega$ (for a 1-Ω resistor) and is referred to as the spectral density function $S_T(\omega)$.

$$S_T(\omega) = \frac{1}{\pi T}|V_T(\omega)|^2 \tag{16.5.5}$$

Even though ω is the most convenient variable for theoretical calculations, the frequency ν (Hz) is the most common system specification. Thus we define $S_T(\nu)$ to yield the same answer as (16.5.4e):

$$\int S_T(\nu)\,d\nu \triangleq \int S_T(\omega)\,d\omega \tag{16.5.6a}$$

or

$$S_T(\nu) = 2\pi S_T(\omega) = \frac{2}{T}|V_T(\omega)|^2 \tag{16.5.6b}$$

We can attribute a circuit function to the mathematical quantity $S_T(\nu)\Delta\nu$ by defining an equivalent voltage generator capable of generating the correct average power in the load resistor when each frequency component is summed over the bandwidth of the postdetector system. If we had started our analysis with a current signal $i(t)$, anther definition of $S_T(\omega)$ would result, with $|I_T(\omega)|^2R$ replacing $(|V_T(\omega)|^2/R)$ in (16.5.4e), and we are naturally led to an equivalent current generator. Since $v = iR$, this is no more complicated than the standard technique of replacing a Thévenin equivalent voltage generator by a Norton circuit involving current sources. These equivalent circuits are shown in Fig. 16.8.

Please note: It is the *power-generating capacity* of the generators that is specified, and this capacity, in turn, depends on the impedance level and bandwidth of the circuit elements

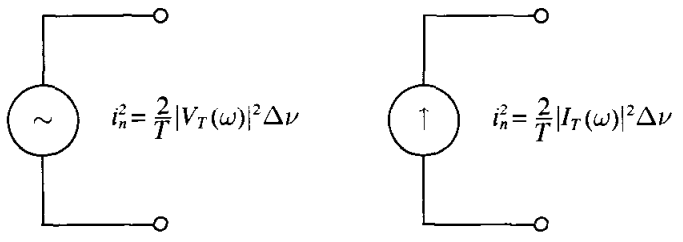


FIGURE 16.8. Equivalent-circuit representations of $S_T(\nu)\Delta\nu$.

placed on the output. The generator is not an entity in itself—*it is defined to give the right answer*—(16.5.4e) (or its counterpart in terms of current).

An example will point out the subtle implications contained in (16.5.4e). Consider the voltage signal produced on the output of the photomultiplier in the preceding section in response to the emission of a photoelectron. Obviously, the display as shown in Fig. 16.6 is the combined effect of the response of the detector *and* the circuitry (i.e., the distributed capacitance). If we looked more carefully at this pulse with a perfect display circuit—one without capacitance—we would observe an extremely short pulse, typically 1 ns or so. As implied by (16.4.1), current only flows while the carrier is in transit (*for any detector*) and τ is very small. If we neglect the change in velocity of the carrier in transit, the transform of the current is

$$\begin{aligned} I_T(\omega) &= \int_{-\tau/2}^{+\tau/2} i(t)e^{-j\omega t}\,dt = \int_{-\tau/2}^{+\tau/2} e\left(\frac{v}{d}\right)e^{-j\omega t}\,dt \\ &= e\left(\frac{v}{d}\right)\left[\frac{e^{j(\omega\tau/2)} - e^{-j(\omega\tau/2)}}{2j\omega\tau/2}\right]\tau \\ &= e\left[\frac{\sin(\omega\tau/2)}{(\omega\tau/2)}\right] \end{aligned} \tag{16.5.7}$$

since $\tau = d/v$. Thus the process of detecting the optical photons produces a current pulse whose frequency components are spread out over a whole band according to the spectral density function $S_T(\nu)$:

$$S_T(\nu) = \frac{2e^2}{T}\left[\frac{\sin(\omega\tau/2)}{(\omega\tau/2)}\right]^2\Bigg|_{\tau\to 0} \to \frac{2e^2}{T} \tag{16.5.8}$$

Equation (16.5.8) must be multiplied by G for the photomultiplier and τ interpreted as the pulse duration of the output current. Quite often we neglect the pulse duration and take the limit of (16.5.8) as $\tau \to 0$ and thus replace $(\sin x/x)^2$ by 1. This is the spectral density function of the detector without modification by the external circuitry.

The important point to remember was stated before, but it is repeated here for emphasis:

The very process of detecting a photon generates noise power over the whole band of (video) frequencies..

This is the critical point in the discussion of shot noise in Sec. 16.6.1.

Now, of course, we seldom deal with only one such pulse or video signal $i(t)$. Rather, we deal with a sequence of signals such as that depicted in Fig. 16.7. Thus, if $\overline{N}$ is the average pulse *rate*, then we would have $\overline{N}T$ pulses in the time interval T, and each pulse would be expressed by $i(t - t_j)$, where t_j is the time origin for each pulse.

$$i = \sum_{j=1}^{\overline{N}T} ef(t - t_j) \tag{16.5.9}$$

where the function $f(t)$ expresses the time behavior of one such event and e is the charge transported. It is important to realize what we can or cannot specify about this output.

We can surely specify the average or DC current produced by this detector—we measure it (with an ammeter). Thus, if the output consists of $\overline{N}$ pulses per second in our sampling interval T, the DC current is merely

$$I_{DC} = \frac{1}{T}\int_{-T/2}^{+T/2} i\,dt = \frac{1}{T}(\overline{N}T_e) = \overline{N}e \tag{16.5.10}$$

But we *cannot* predict when the pulses will arrive.* We can predict only the average rate of arrival given by $\overline{N}$. Thus the Fourier transform of the output current must reflect this ignorance.

$$\begin{aligned} I_T(\omega) &= \sum_{j=1}^{\overline{N}T} e\int_{-T/2}^{+T/2} f(t - t_j)e^{-j\omega t}\,dt \\ &= \sum_{j=1}^{\overline{N}T} eF(\omega)e^{-j\omega t_j} \end{aligned} \tag{16.5.11}$$

where $F(\omega)$ is the transform of a single event. To compute the spectral distribution of the power contained in (16.5.11), we need to form the product $I_T(\omega)I_T^*(\omega)$:

$$|I_T(\omega)|^2 = \left[e\sum_{j=1}^{\overline{N}T} F(\omega)e^{-j\omega t_j}\right]\left[e\sum_{k=1}^{\overline{N}T} F^*(\omega)e^{+j\omega t_k}\right] \tag{16.5.12a}$$

When $j = k$, the exponential factors cancel and the summation yields $\overline{N}T$.

$$|I_T(\omega)|^2 = e^2|F(\omega)|^2\left[\overline{N}T + \sum_{j\neq k}^{\overline{N}T}\sum_{k=1}^{\overline{N}T} e^{j\omega(t_k - t_j)}\right] \tag{16.5.12b}$$

If we admit our ignorance and confess that we do *not* know the time delays, then the average result of many repeated applications of (16.5.12b) is the first term with the double summation averaging to zero:

$$|I_T(\omega)|^2 = e|F(\omega)|^2(\overline{N}eT) \tag{16.5.13}$$

*The only way we could predict the arrival of the "next" photon is to have measured it, thereby ensuring that the "next" photon does not hit the detector.

Now we can identify the product $\overline{N}e$ as the DC current [i.e., (16.5.10)] and substitute this expression into (16.5.6b) to find the spectral density function:

$$S_{\dot{T}}(\nu) = \frac{2}{T}|I_T(\omega)|^2 = 2eI_{\mathrm{DC}}|F(\omega)|^2 \tag{16.5.14}$$

We are now in a position to handle many of the aspects of noise in a detection system.

16.6 SOURCES OF NOISE

Noise, by implication, is an undesirable feature of detection with it being a video current or voltage wave riding on top of the signal component. It is our purpose here to identify the sources of noise so as to estimate and to *accept the limiting performance* of our system. This limiting state is obtained when the noise as a result of the quantized nature of the electrical charge, $\pm 1.6 \times 10^{-19}$ C, generated by the absorption of quanta of light, overwhelms all other sources of noise. No amount of circuitry tricks, cooling, design of optics, or other such skullduggery will help in this limit of "shot noise" or "quantum noise."

There are other noise sources that can be minimized but never completely eliminated. Any detector must observe a background that has a finite temperature, and thus there are always a few unwanted optical photons impinging on the detector, which, in turn, generates these quantized electrical carriers and their associated video noise. The video circuit itself consists of a resistor at a finite temperature, which is part of an amplifier that introduces its own brand of noise. By careful design, these sources can be minimized.

16.1.1 Shot Noise

There are two viewpoints as to the origin of this type of noise. One viewpoint is that the noise is the result of the quantized nature of electromagnetic energy. The other is that it is the result of the quantization of the electrical charge. It is impossible to distinguish between the two opinions, since both are facts of life and one (the charge) is the direct result of the absorption of the other (the photon). However, since there are other physical processes that generate a free carrier, such as thermionic emission from a photocathode or thermal generation of electron-hole pairs in a semiconductor, it is convenient to assign this noise to the discrete nature of the electrical charge.

In either case, shot noise is the result of the probabilistic nature of the generation of the electrical charge within the detector. We can specify the average generation rate, but we cannot specify when the next charge will be emitted given that the first started at $t = 0$. This is precisely the case considered in Sec. 16.5. If we neglect the time interval during which the charges move to the collector, then $\overline{F(\omega)^2} = 1$ in (16.5.14), and the equivalent circuit is as shown in Fig. 16.9.

It may seem strange at first that in Fig. 16.9, a direct current causes an AC noise, until we remember that this average current is actually a sequence of many very short pulses. It is the spectral content of those pulses that is represented by the AC noise generator. Shown also in Fig. 16.9 is the realistic situation where there is some DC current even when the detector

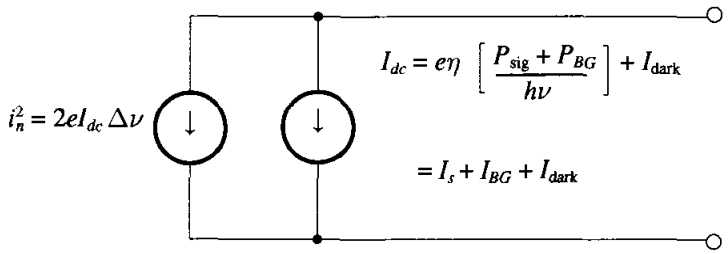


FIGURE 16.9. Equivalent circuit of the detector, showing the signal and noise current generators.

is not illuminated. As indicated, this dark current may be due to thermionic emission from the cathode of a photomultiplier, or it may be due to the thermal generation of electron-hole pairs in a semiconductor. Furthermore, there may be some current because of background photons riding along with our desired optical signal. But in either or any case, the current still comes in the form of impulses of charge and thus contributes to the noise power.

The signal power to a load R_L is $I_s^2 R_L$ and hence, the signal-to-noise ratio is

$$\frac{S}{N} = \frac{I_s^2 R_L}{i_N^2 R_L} = \frac{[e\eta(P_{\text{sig}}/h\nu)]^2}{2eI_{\text{DC}}\Delta\nu} \tag{16.6.1}$$

Even if we hope for the ideal situation where the dark current I_D and that due to the background I_{BG} is negligible, we are still faced with a finite S/N ratio due to the fact that the signal itself produces a DC current.

$$\left(\frac{S}{N}\right)_{\max} = \frac{[e\eta(P_{\text{sig}}/h\nu)]^2}{2e[e\eta(P_{\text{sig}}/h\nu)]\Delta\nu} = \eta\left(\frac{P_{\text{sig}}}{h\nu}\right)\frac{1}{2\Delta\nu} \tag{16.6.2}$$

This is the best that can be done. Indeed, (16.6.2) is so obvious that we could have stated it based on common sense without all of the mathematical harangue of this chapter. If $\eta_{qe} = 1$ (the best), then $S/N = 1$ when there is one photon per sampling-time interval $T/2 = 1/(2\Delta\nu)$. Since we always require two samples to distinguish between a 1 or a 0 bit, (16.6.2) states that the maximum data rate with a video channel of bandwidth $\Delta\nu$ is $\Delta\nu/2$ bits per second (for $S/N = 1$).

There are other sources of noise, but (16.6.2) is the limiting value of S/N. Let us postpone further discussion until these sources are identified and combined with what we just discussed.

16.6.2 Thermal Noise

Thermal noise is due to the finite temperature of the elements of the detector system. As such, it can be partially alleviated by cooling these elements to dry-ice (195 K), liquid-nitrogen (77 K), or even liquid-helium (4 K) temperatures.

This radiation goes by many different names: (1) it is sometimes referred to as "white" noise, because its spectral distribution is uniform or "flat" at video frequencies at normal temperatures (the same is true of shot noise); (2) it is sometimes referred to as "Johnson" or "Nyquist" noise, after early pioneers in the field; and (3) it is also called "blackbody" noise, because it can be derived from Planck's blackbody formula discussed earlier.

Recall that the beginning of the quantum era was the successful explanation of the radiation emerging from a small "hole" in a heated cavity. The radiation emerging from this cavity is due to the energy in those cavity modes, which couple to the small aperture.

Although our initial discussion of this fact in Chapter 7 was in terms of a cavity that is huge compared to a wavelength, the same considerations apply at video frequencies for elements that are small compared to a wavelength. The maximum power that can be transferred from a blackbody at a temperature T in a bandwidth $\Delta\nu$ is the product of the following factors: (1) the number of modes in that bandwidth $\Delta\nu$, (2) the energy per photon, (3) the photons per mode, and (4) the bandwidth $\Delta\nu$.

At video frequencies, there is usually only one mode—the TEM mode extending down to zero frequency—and it surely has only one orientation (or polarization) of the field. Thus we obtain the low-frequency limit pertinent to our detector problem.

$$p_n = 1\left[\frac{h\nu}{\exp(h\nu/kT) - 1}\right]\Delta\nu \tag{16.6.3}$$

At normal temperatures (273 K) and reasonable frequencies (e.g., less than 1 GHz) the photon energy is much less than the characteristic thermal energy, $h\nu \ll kT$, and the approximate form of (16.6.3) is often used.

$$p_n = kT\Delta\nu \tag{16.6.4}$$

It is this last form that is often referred to as "Johnson" noise, but it is an approximation to the more correct formula, (16.6.3).

Now, this is the maximum power that can be transferred from one blackbody circuit element to another blackbody element in the form of thermal noise. Figure 16.10 illustrates the restrictions involved in applying (16.6.3). Resistor R_A at a temperature T_A emits noise power according to (16.6.3), to be absorbed by R_B, which, in turn, radiates power back to A. Only if $R_A = R_B$ will the two systems be "black" to each other's radiation, thereby absorbing all of the incident power.

Thus the characterization of an element being "black" is identical to the specification of a system being matched for maximum power transfer. To account for the fact that many systems are not matched, we construct an equivalent circuit with appropriate generators and *noiseless* resistors, which yields the correct answer for the power transfer when matched [i.e., (16.6.3)] and also properly allows for a mismatch. Such a model is shown in Fig. 16.11.

We *define* a voltage (or current) generator of such a magnitude to yield the proper power transfer according to the laws of transmission-line theory, such that it agrees with

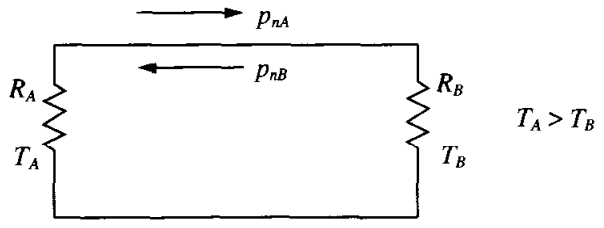


FIGURE 16.10. Interchange of power between two resistors.

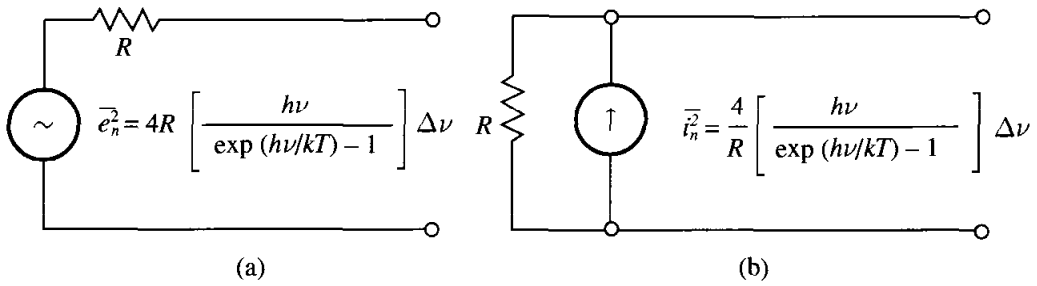


FIGURE 16.11. Equivalent circuits for thermal noise.

the maximum power specified by Planck. Thus the mean of the squared rms voltage is

$$\overline{e_n^2} = 4R\left[\frac{h\nu}{\exp(h\nu/kT) - 1}\right]\Delta\nu \approx 4kTR\,\Delta\nu \tag{16.6.5}$$

and the corresponding value for the current generator is

$$\overline{i_n^2} = \frac{4}{R}\left[\frac{h\nu}{\exp(h\nu/kT) - 1}\right]\Delta\nu \simeq \frac{4kT}{R}\Delta\nu \tag{16.6.6}$$

It is only the resistive part of the circuit that can accept power and, by the same token, can generate this noise power. A reactive element such as a capacitor affects the bandwidth $\Delta\nu$ under consideration but does not contribute to or accept the noise power.

16.6.3 Noise Figure of Video Amplifiers

Most optical systems require a video amplifier to amplify the output to a level where it can be used for communication and control purposes. In doing so these amplifiers contribute their own noise power, which must be *added* to the amplified value from the detector, thereby degrading the signal-to-noise ratio. It is important that we account for this reality.

The primary causes of noise in a video amplifier are shot noise from charge transport in the active devices (i.e., transistors) and amplified thermal noise from the resistive components. The net result is that there is a noise component to the voltage (or current) output of the amplifier even when there is no signal input or the input resistor is cooled to 0 K. This is the *excess* noise contributed by the amplifier.

The concept of a noise figure (of merit), F, was invented to describe this excess noise in a systematic manner. If the output noise power is divided by the gain, we obtain an equivalent noise source at the input terminals of the amplifier, which can now be considered noiseless. A noise temperature T_A is now defined to correctly specify the excess noise at the input terminals with its specified input resistance, which, in turn, contributes its own noise by virtue of its temperature. Since the excess noise and the resistor noise are independent quantities, add the two powers (or the mean of the squared voltages or currents). The noise figure F acknowledges the addition and, by convention, assumes that the input resistor is at 290 K (or 17 C).

$$F = 1 + \frac{T_A}{290} \tag{16.6.7}$$

An example should help to illustrate the procedure.

Example

Suppose that we had an amplifier of 40-dB gain, a noise figure of 13 dB, an input and output impedance of 50Ω, and a bandwidth of 500 MHz. From the noise-figure specifications, $F = 20$ or the equivalent temperature of $19 \times 290\ \text{K} = 5510\ \text{K} = T_A$. The total *output* noise power in the 500-MHz bandwidth is 10^4 (i.e., the gain) times the total input noise power.

$$\langle P \rangle_{\text{out.}} = Gk(T_A + T_R)\Delta\nu = 0.4\ \mu\text{W}$$

The rms value of the AC voltage across the 50 Ω output is 4.47 mV. If we cooled the input resistor to liquid-helium temperatures (4 K), we change T_R but do not affect T_A. Thus we make a minimal improvement.

It should be clear from this example that the noise temperature T_A bears little, if any, relation to the ambient temperature of the amplifier. T_A is defined to correctly predict the noise contributed by the amplifier, and nothing more.

16.6.4 Background Radiation

Background radiation is probably the most obvious source of noise in an optical system and one that can be reduced with minimum effort. The background is the stray optical photons impinging on the detector. At the very minimum, any detector will be subjected to the blackbody radiation of the background at a finite temperature T_{BG}. We again invoke Planck's law* to estimate this unwanted optical radiation impinging on the detector.

$$\left\langle P_{\text{opt.}} \right\rangle_{\text{BG}} = \int_{\Delta\nu_{\text{opt.}}} \frac{1}{4} I(\nu)\, d\nu A_{\text{det.}} \frac{d\Omega}{4\pi} \tag{16.6.8}$$

where $I(\nu)$ equals the intensity per frequency interval $d\nu$ and is found from (7.2.10) (repeated here for convenience):

$$\frac{c\rho(\nu)}{\eta_g} = I(\nu) = \frac{8\pi n^2}{\lambda_0^2} \frac{h\nu}{\exp(h\nu/kT) - 1} \tag{7.2.10}$$

$\Delta\nu_{\text{opt.}}$ represent the optical bandwidth of the detector, A is the area, and $d\Omega/4\pi$ is specified by the field of view (FOV).

If the signal has a specific sense of polarization we can eliminate one half of the power represented by (16.6.8) by eliminating the orthogonal polarization. It is also desirable to minimize the optical bandpass $\Delta\nu_{\text{opt.}}$ and to restrict the field of view. But in doing so, it is important to recognize that the optical elements used to accomplish these purposes can also radiate blackbody radiation at their own temperature. For instance, if we used an absorptive polarizer at 300 K to eliminate the vertically polarized noise from a 200 K background source, the noise level would increase rather than decrease. This is because any component—the polarizer in this case—is equally good as a radiator as it is as an absorber (Kirchoff's law).

*The factor $\frac{1}{4}$ is used when the *isotropic* blackbody flux given by (7.2.10) is used.

16.7 LIMITS OF DETECTION SYSTEMS

Now that we have identified the major sources of noise in an optical detection system and have become familiar with the different types of detectors, it is time to combine the two to ascertain the limit and to establish realistic signal-to-noise ratios. As such, this section is more like a sequence of examples of previous concepts rather than a section in which new ground is broken. However, these examples point out the extreme importance of high quantum efficiency and low noise in the first stage of amplification in obtaining the highest signal-to-noise ratio.

The major new topic is that of optical heterodyning. With this technique, we can approach the quantum limit of detection given by (16.6.2), even without the benefit of extreme low-noise amplifiers, such as the secondary emission amplifier in a photomultiplier.

16.7.1 Video Detection of Photons

The most straightforward system for detection is shown in Fig. 16.12, where a *p-i-n* photodiode is being irradiated by an optical signal P_s and its output is being fed into the video amplifier chain with finite noise temperatures T_{A1} and T_{A2}. The analysis of the signal-to-noise ratio at the output of this system follows from a straightforward application of the previous sections.

The desired signal power is the square of the signal current times the input resistance, and this power is then amplified by a factor G_1G_2.

The *output* noise consists of a sum of three terms: (1) the second stage contributes $G_2kT_{A2}\Delta\nu$, (2) the noise from the input resistor at T_R and the first amplifier contributes $G_1G_2k(T_R + T_{A1})\Delta\nu$, and (3) the shot noise produced by the current in the detector and subsequently amplified by G_1G_2 contributes a power of $G_1G_2(2eI_{DC}\Delta\nu)R$. Thus S/N is given by

$$\frac{S}{N} = \frac{G_1G_2[e\eta(P_{\text{opt.}}/h\nu)]^2R}{k(G_2T_{A2} + G_1G_2T_{A1} + G_1G_2T_R)\Delta\nu + G_1G_2(2eI_{DC}\Delta\nu)R} \tag{16.7.1a}$$

or

$$\frac{S}{N} = \frac{[e\eta(P_{\text{opt.}}/h\nu)]^2R}{k[(T_{A2}/G_1) + T_{A1} + T_R]\Delta\nu + 2eI_{DC}\Delta\nu R} \tag{16.7.1b}$$

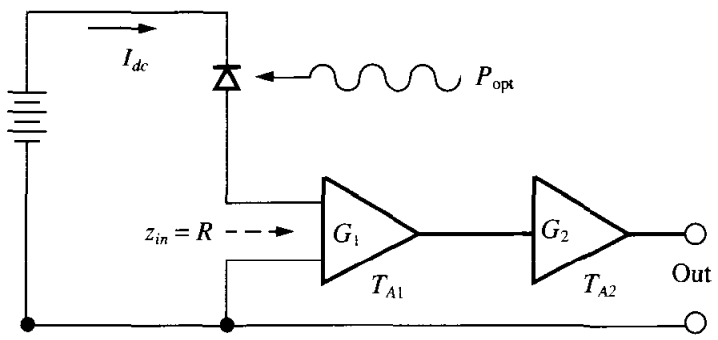


FIGURE 16.12. Photodiode video detection circuit.

This last form is equivalent to summing all noise powers at the input of the first stage and then considering all amplifiers as noiseless. This is the conventional procedure and will be followed hereinafter.

Equation (16.7.1b) also emphasizes the importance of a low noise/high gain front end. If G_1 is large, then the second amplifier can be quite noisy and affect the S/N ratio hardly at all. Obviously, we want to keep the excess noise contributed by the first stage as small as possible.

In any case, we obtain the *limiting performance* when the shot-noise term (due to the detection of the signal) overwhelms the thermal noise. The minimum value of the current through the detector is that caused by the optical power. When a detector system is dominated by shot noise, the system is said to be at the "quantum limit." (Obviously, we do not increase the DC current arbitrarily just to obtain a shot-noise-dominated system.) This fact explains why a photomultiplier and an avalanche photodiode are the premium detectors for video detection and why the heterodyne system reaches that limit without the benefit of an apparent amplifier.

Example: Noise in a Photomultiplier

Consider first the photomultiplier as described in Fig. 16.2. The incident photons create a cathode current of $e\eta(P_{\text{opt}}/h\nu) = I_k$, which is subsequently amplified by the N stages of secondary emission or $G = \delta^N$. By the same token, the shot-noise current is amplified by the same value. Even though the secondary emission amplifier can be considered perfect insofar as thermal noise is concerned, the fact is that the secondary electron emission is a statistical phenomenon with considerable variance about the mean value of δ. Thus the emission from each dynode is also an independent statistical process and therefore creates its own brand of shot noise. The total squared noise current out of the photomultiplier is

$$\overline{i_n^2} = 2e\Delta\nu(\delta^{2N} I_k + \delta^{2N-2} I_1 + \cdots + I_N) \tag{16.7.2a}$$

where $I_1, I_2, \ldots, I_N$ are the currents emitted by the various dynodes. Now the ratio of the dynode emission currents is the mean value of the secondary emission ratio $\delta = I_N/I_{(N-1)} = I_1/I_k$. Hence, (16.7.2a) can be simplified and summed:

$$\overline{i_n^2} = 2eG^2 I_k \Delta\nu(1 + \delta^{-1} + \delta^{-2} + \cdots + \delta^{-N}) \tag{16.7.2b}$$

$$= 2eG^2 I_k \Delta\nu \frac{1 - \delta^{-(N+1)}}{1 - \delta^{-1}} \simeq 2eG^2 I_k \Delta\nu \frac{1}{1 - \delta^{-1}} \tag{16.7.2c}$$

for δ a reasonable number ($2 \rightarrow 4$) and N large.

Now we add the thermal noise contributed by the load resistor or conventional amplifiers. The signal-to-noise ratio is

$$\frac{S}{N} = \frac{[Ge\eta(P_{\text{opt.}}/h\nu)]^2 R_L}{k(T_R + T_A)\Delta\nu + 2eG^2 I_k \Delta\nu[1/(1 - \delta^{-1})]R_L} \tag{16.7.3}$$

where: $$I_k = \frac{e\eta P}{h\nu} + I_{\text{dark}}$$

For anything reasonable, the second term in the denominator overwhelms the thermal noise, and the photomultiplier is entirely dominated by shot noise. If we cool the photocathode so as to eliminate thermionic emission and neglect other extraneous electron emission causes,

$I_{\text{dark}} \doteq 0$ and we come close to the quantum limit rather easily.

$$\frac{S}{N} = \eta \left(\frac{P_{\text{opt}}}{n\nu} \right) \frac{1}{2\Delta\nu} (1 - \delta^{-1}) \tag{16.7.4}$$

We are able to do this by virtue of the secondary emission amplifier.

Example: Detection with an Avalanche Photodiode

A similar advantage is present in an avalanche photodiode, although we cannot, in practice, achieve as high a current gain as can be used in a photomultiplier. Typical avalanche gains of 30 to 100 are used, and this is sufficient for the system performance to be greatly improved over that with a simple photodiode.

In an avalanche photodiode, the carrier production rate $\eta(P/h\nu)$ is increased by a factor of M because of the ionization by the drifting electrons and holes. Hence, the photocurrent and the dark current are enhanced by the same value:

$$i = Me\left[\eta \left(\frac{P}{h\nu} \right) + g_{\text{th}}\right] \tag{16.7.5}$$

Each creation of a new carrier by the avalanche process is a statistical process, one whose average rate (M) can be measured and thus used. However, we cannot predict the occurrence of each individual event, and thus the avalanche multiplication contributes to the shot noise.

At first, one would hope for the same situation as found in a photomultiplier with its "noise-free" gain provided by the secondary emission amplifier (which is an avalanche of a sort). Unfortunately, noise power from an avalanche diode increases as M^n, where the exponent n is between 2 and 3.

$$p_n(\text{shot noise}) = M^n \left\{ 2e\left[\eta \left(\frac{P}{h\nu} \right) + g_{\text{th}}\right]\Delta\nu \right\} R \tag{16.7.6}$$

where g_{th} is due to the thermal generation of carriers, and the term in brackets is the optical generation rate. Both processes create shot noise, and each must be multiplied by M^n.

The fact that the avalanche diode creates excess noise in the multiplication process (i.e., $n > 2$ in the equation above) means that we cannot use an arbitrarily large multiplication factor. This follows directly from the expression for S/N for an avalanche photodiode and video amplifier combination.

$$\frac{S}{N} = \frac{M^2[e\eta(P/h\nu)]^2 R}{k(T_R + T_A)\Delta\nu + M^n \left\{2e^2[\eta(P/h\nu) + g_{\text{th}}]\Delta\nu\right\} R} \tag{16.7.7}$$

If the internal generation rate g_{th} and the optical power are very small, then the thermal noise dominates and an increase in M from 1 helps considerably. However, if M is too large, then the shot-noise term dominates and the signal-to-noise ratio degrades for increasing M.

$$\left.\frac{S}{N}\right|_{M \text{ large}} \sim \frac{1}{M^{n-2}}$$

Obviously, there is an optimum value of M for maximum S/N, an obvious problem for students.

Even though we cannot go to the extreme gain as with a photomultiplier, the multiplication does alleviate the necessity of conventional amplifiers with their own brand of noise. Furthermore, high-quantum-efficiency avalanche photodiodes can be constructed from various combinations of semiconducting materials for different regions.

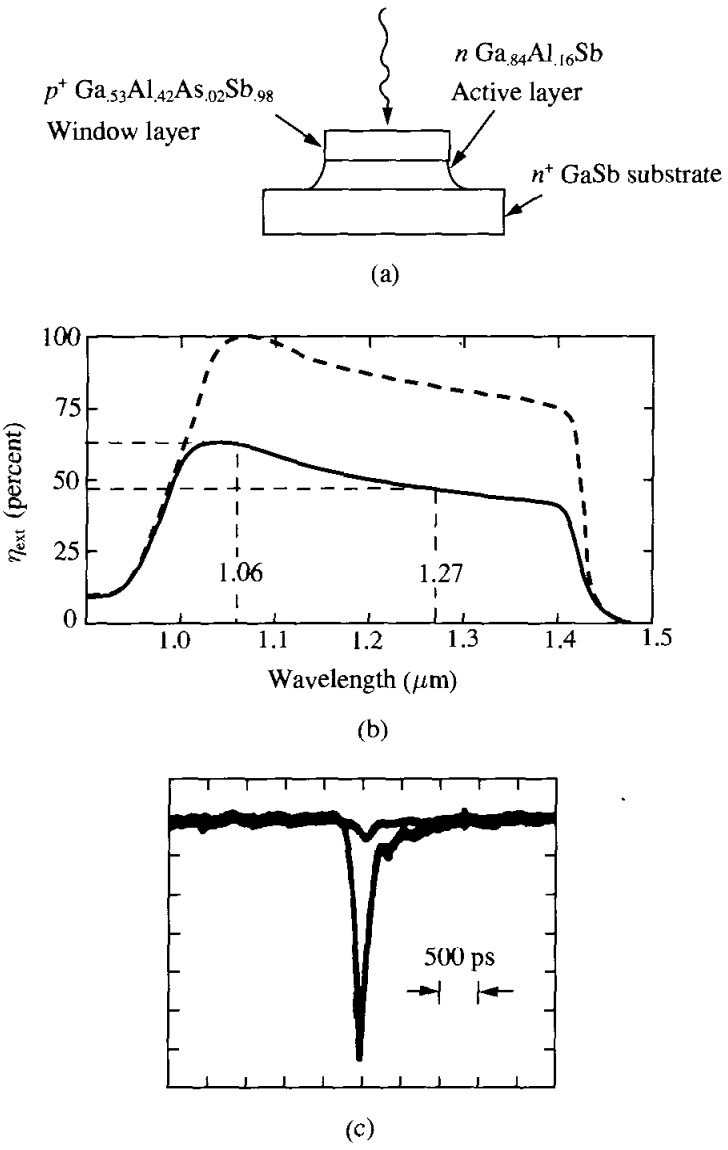


FIGURE 16.13. Details of a GaAlAsSb avalanche photodiode (APD). (a) Structure of a 1.0- to 1.4 μm GaAlAsSb avalanche photodetector. (b) Experimental photoresponse of the GaAlAsSb APD shown in (a) at low bias. The dashed curve shows the projected quantum efficiency for a detector with a suitable antireflection coating. (c) Pulse response to a mode-locked Nd-YAG laser. Lower trace $M = 17$, upper trace $M = 1$. The estimated photodiode response time is 120 ps (FWHM). (After "High Sensitivity Optical Receivers for 1.0–1.4 μm Fiber Optic Systems," Louis R. Tomasetta, H. David Law, Richard C. Eden, Ira Deyhimy, and Kenicht Nakano, *IEEE J. Quantum Electron. QE-14*, No. 11, 800–804, 1978.)

For instance, Fig. 16.13 shows the construction details and measured quantum efficiency of a GaAlAsSb avalanche photodiode that has its peak response in the wavelength region 1.0 to 1.4 μm. If you will recall, this is precisely the wavelength region in which the absorption loss and material dispersion of fiberoptic cable is at a minimum. Hence, many of the future fiber communication systems will utilize wavelengths there and utilize detectors such as those illustrated in Fig. 16.13. Since the quantum efficiency of any photocathode in that region is extremely bad (if not zero), the avalanche photodiode is a clear choice for this application.

Although the prior discussion acknowledged the existence of the excess noise, it did not address the question of why $n > 2$ in (16.7.7). To answer that requires us to delve more deeply into the multiplication itself. It would lead us far astray to give a complete theory of this excess noise, so we will have to be content with a gross physical picture.

The excess noise can be attributed to the fact that the multiplication process is the result of the ionization (or avalanche) by *both* carriers; that is, electrons and holes. If only

one carrier contributed to the multiplication, there would be noise caused by the statistical nature of the avalanche, as in a photomultiplier, but n would equal 2 and there would be no excess noise. But, alas, both do contribute and our hopes are dashed.

To appreciate the implications, consider the case where an electron-hole pair is generated by the photons in the high-field region of the p-n junction diode of Fig. 16.5 and assume that the electron can make two ionizations before being collected by the n contact. This would be true if the field were high enough, and the path length long enough, so that the electron could gain enough energy to create another electron-hole pair as it drifted. Because of the specification that two ionization events can occur, the second electron can also gain enough energy to create another electron-hole pair, as the first generation can create still another pair. Thus we have four electron-hole pairs in response to the absorption of one photon, and $M = 4$. If that were the end of it and the electrons were collected by ohmic contact (where they recombine) on the n terminal, and the hole collected by the p contact, we would have noise-free multiplication.

Unfortunately, that is not the end of it. We have ignored the holes. While the initial hole produced by the photon may be collected by the p terminal without causing any additional ionization, the other three holes are created much farther away from the collecting terminal. They can gain energy from the field and cause ionization in the same manner as the electrons.

If the probability of one hole making an ionization is greater than one third, that of three holes is greater than 1. Thus there is greater than unity probability that a new electron-hole pair will be created in the area near the n region but still in the high-field. To follow that electron, we jump back three paragraphs and start all over. (Obviously, we are in an infinite DO loop.)

What is the result? The device is broken down with the external current infinite (or limited by the external circuit), and the device is useless as a detector. Obviously, we cannot use a value of $M = \infty$. But the point is that any small fluctuation in these probabilities causes a greatly enhanced fluctuation in M. That is the origin of the excess noise and $n > 2$.

It is interesting to note that the avalanche process (and breakdown) is identical to that occurring in a gas discharge. Indeed, the origin avalanche photodiode was a gas-filled phototube. Although some of the details change for gas discharges, the end result is the same, a device passing current limited by the external circuit.

16.7.2 Heterodyne System

The heterodyne detector system is the optical analog of the very common radio receiver, and its schematic is shown in Fig. 16.14. Its operation is trivial to understand, but the realization of the principle requires considerable effort and care. We combine a weak signal at an optical frequency ω_1 with a strong local oscillator at another optical frequency ω_2 to obtain a "beat" or IF frequency $(\omega_2 - \omega_1)$, which is then used for communication or control. Obviously, $\omega_2 - \omega_1$ must be in the passband of the video circuits for the beat note to perform this function. Hence, this receiver has an extremely narrow optical bandpass. This has many advantages from a detection standpoint, but it places rather stringent requirements on the frequency stability of the source and local oscillator. The critical issues to ensure success are (1) to have a detector whose output current is proportional to the optical power

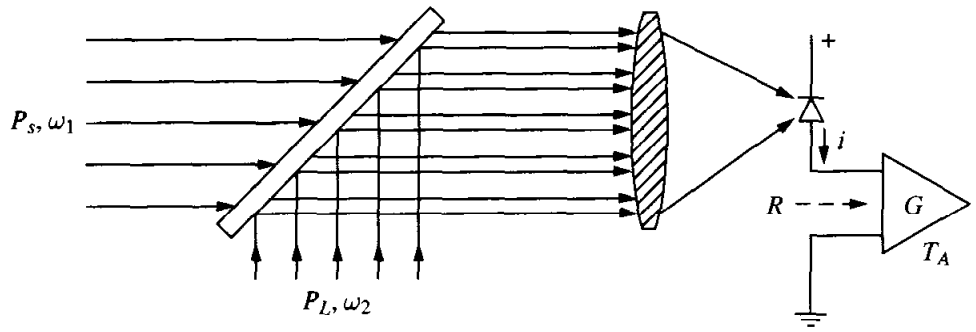


FIGURE 16.14. Optical heterodyne system.

(fortunately, all quantum detectors fall into this category) and (2) to align the signal and local oscillator waves to be collinear and coincident on the detector (a not so trivial, but not impossible task).

To analyze the situation, we revert back to a classic description of the optical field and assume that the output current is proportional to the square of the *total* optical electric field of the waves absorbed by the detector.

$$i = e\eta_{qe}\frac{1}{h\nu}\left[\left(\frac{\mathbf{E}_T \cdot \mathbf{E}_T}{\eta_0}\right)A\right] \tag{16.7.8}$$

One might tend to assume that this classic procedure ignores the "photon" characteristics of the optical signal, and therefore one could not hope for an accurate description of the process in the limit of a few photons. The fact is that this procedure will predict the correct answer, even in this limit, because we *define* the classic field so that the quantity in the brackets is equal to the *photon* power.*

Equation (16.7.8) also relates the peak optical field to the average or DC current in addition to relating it to the photon power.

$$I_{\mathrm{DC}} = \left\langle e\eta_{qe}\frac{1}{h\nu}\frac{E^2}{\eta_0}A\right\rangle = e\eta_{qe}\frac{E^2}{2\eta_0}A = e\eta_{qe}\frac{\langle P_{\mathrm{opt}}\rangle}{h\nu} \tag{16.7.9a}$$

Therefore

$$E^2 = \frac{2\eta_0}{A}\langle P_{\mathrm{opt}}\rangle \tag{16.7.9b}$$

where A is the area of the optical beam and $\eta_0 = 377\ \Omega$. In the heterodyne system shown in Fig. 16.14, the total electric field has components associated with the signal and the local oscillator. We express the field amplitudes in terms of the powers of each wave.

$$E_T = \left(\frac{2\eta_0}{A}\right)^{1/2}\left(P_s^{1/2}\cos\omega_1 t + P_L^{1/2}\cos\omega_2 t\right) \tag{16.7.10}$$

*An excellent, clear, readable, and enjoyable discussion of some of the issues involved in invoking the photon nature of the electromagnetic waves can be found in Scully and Sargent [8].

Thus the current is given by (16.7.8) with the field given by (16.7.10).

$$i = \frac{2e\eta_{qe}}{h\nu}\left(P_s \cos^2 \omega_1 t + 2P_s^{1/2}P_L^{1/2} \cos \omega_1 t \cos \omega_2 t + P_L \cos^2 \omega_2 t\right) \quad (16.7.11a)$$

or

$$i = e\eta_{qe}\frac{1}{h\nu}\Bigg[\underbrace{P_s + P_L}_{\text{DC terms}} + \underbrace{2P_s^{1/2}P_L^{1/2}\cos(\omega_2 - \omega_1)t}_{\text{intermediate frequency (IF)}} \quad (16.7.11b)$$

$$+ \underbrace{P_s \cos 2\omega_1 t + P_L \cos 2\omega_2 t + 2P_s^{1/2}P_L^{1/2}\cos(\omega_2 + \omega_1)t}_{\text{optical frequencies}}\Bigg]$$

The last line of (16.7.11b) represents current at optical frequencies, a result of our classic analysis. These would not be present if a full-blown quantum description had been used, but, in any case, the low-frequency detector circuit would not respond anyway, so we drop them.

The first two terms represent a DC current and are overwhelmingly dominated by the much stronger local oscillator terms. This DC current causes "shot" noise, as before. The IF or "beat" frequency term contains the information about the presence or absence of the signal wave. Thus the video signal current can be expressed as

$$i_s = \frac{2e\eta_{qe}}{h\nu} P_s^{1/2}P_L^{1/2} \cos(\omega_2 - \omega_1)t \quad (16.7.12)$$

where P_s is the peak value of the modulated wave used for communication.

We can make the signal current arbitrarily large by using a very large local oscillator power. This is how the heterodyne system obtains its gain. The penalty to be paid is that the DC current becomes arbitrarily large also:

$$I_{DC} = e\eta_{qe}\frac{P_L}{h\nu} \quad (16.7.13)$$

and thus the shot noise also increases.

Now it is just a matter of elementary computation to compute the S/N ratio of the detector system shown in Fig.16.14. For a 100% modulated wave, corresponding to the "on-off" code analyzed in Sec. 16.6.1, the average signal power is one half of the peak signal power, and thus

$$\frac{S}{N} = \frac{\frac{1}{2}i_s^2 R_L \cdot \frac{1}{2}}{k(T_A + T_R)\Delta\nu + 2e[e\eta_{qe}(P_L/h\nu)]\Delta\nu} \quad (16.7.14)$$

By making the local oscillator power sufficiently strong, the shot noise dominates the thermal noise of the amplifier and we obtain the same quantum limit specified by (16.6.2).

$$\frac{S}{N} = \eta_{qe}\frac{\langle P_s \rangle}{h\nu}\frac{1}{2\Delta\nu} \quad (16.7.15)$$

Note that the heterodyne system provides (thermal) "noise-free" gain in much the same manner as the secondary emission amplifier does for the photomultiplier. Thus we

should not be surprised at the same quantum limit. But note that this "noise-free" gain is not restricted to wavelengths where the vacuum photoelectric effect is applicable. Any square-law detector and any strong local oscillator can be used (indeed, the common household AM/FM radio operates on this principle). (It does not have the advantage of using quantum detectors. Nevertheless, it is incredibly sensitive.)

PROBLEMS

16.1. **(a)** What is the current density emitted by a photocathode with $\eta_{qe} = 0.2$ caused by an optical power of 1 kW/cm^2 at $\lambda = 8000$ Å?

(b) Assume constant quantum efficient *and* constant power. Plot the photocathode current as a function of wavelength between 2000 Å and 1.0 μm.

(c) What should the anode-to-cathode voltage be to ensure that space-charge effects are not important. (*Hint*: Evaluate the Child–Langmuir limit for the current; see Problem 17.2.)

(d) If the $A - K$ spacing were 2 mm for a vacuum photodiode, what would be the transit time of a photoelectron? (Use $V_{A-K} = 2000$ V.)

16.2. We usually approximate the current associated with the transport of charge from the cathode to the anode of a photodiode as being a delta function, thus producing a "white" shot-noise spectra. Evaluate this approximation for a vacuum photocathode with $d = 2$ mm and $V_{A-K} = 2000$ V.

16.3. **(a)** Consider a surface charge $-\rho_s$ located at point x between two conducting planes at $x = 0$ and $x = d$ that are connected together by an external circuit. At this surface layer of charge moves with velocity $+v$ toward $x = d$, show that the current flow in the external circuit is given by $(\rho_s A)(v/d)$. (Hint: Use Gauss's law to find the time rate of change of the induced charge on the contacts at $x = 0, d$.)

(b) Plot the wave shape of the current from a vacuum diode and contrast it to that produced by a semiconductor diode. Assume equal charge transported in the same time for both cases.

(c) Now let an electron-hole pair be generated at some arbitrary point x. Let the electrons travel to $x = d$, the holes to $x = 0$. What is the current? [Find the induced charge by superposition of the solution found in (a)].

16.4. Suppose that N electron-hole pairs were created in the center of the intrinsic region of a p-i-n diode. Assume that bias voltage across this region of width w is V_0 and that the mobility of the electron is twice that of the hole [(typical of silicon; see Streetman [1], p. 87). Plot the output current as a function of time.

16.5. Consider the simple picture of a mode-locked laser at 6328 Å in which $n = 9$ equal-amplitude modes are centered about ν_0. The average power is 27 mW, and the mode spacing is 125 MHz. The output of this laser is detected with a vacuum photodiode with a quantum efficiency of $\eta = 0.15$. Neglect space-charge effects but do not assume ideal circuitry.

(a) What is the distance between the mirrors of this laser?

(b) If the load for the diode consists of 1-kΩ resistor shunted by $C = 0.01\ \mu$F, what is the output voltage?

(c) If $R_L = 50\ \Omega$ and $C = 10$ pF, use (9.5.5b), Fig. 9.17, and the material of this chapter to predict the display on an oscilloscope.

16.6. There is a considerable difference in the "$c/2d$ beats" between the various modes of a laser, depending on whether the system is mode-locked or not. Assume the system described in Problem 16.5 and $R_L = 50\ \Omega$ and $C_x = 20$ pF for the radio receiver used to measure these beats. (Remember that mode 9 can beat with 8, which can beat with 7, and so on, each giving a $c/2d$ beat note.) Derive an expression for this beat-note amplitude, assuming:

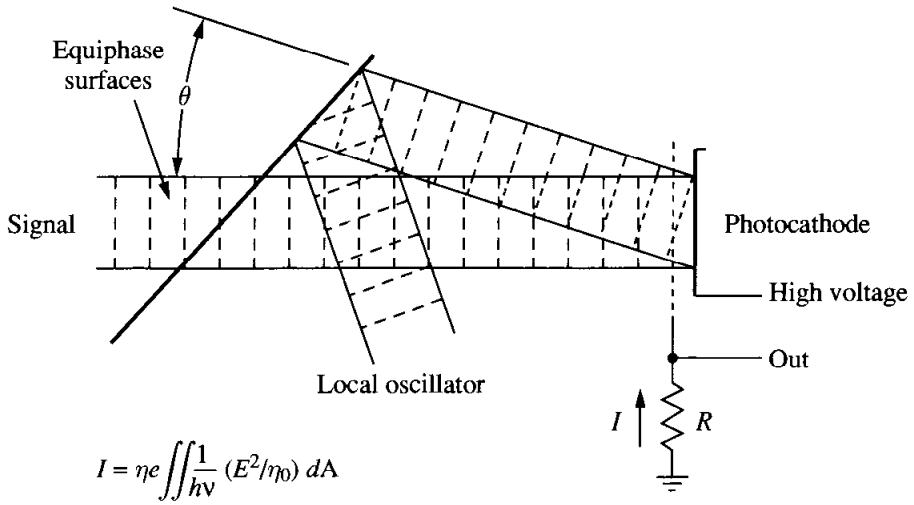
(a) The laser is mode-locked, and ideal circuitry (i.e., $C_x = 0$).

(b) Mode-locked with $R_L = 50\ \Omega$ and $C_s = 20$ pF.

(c) Assume ideal circuitry ($C_s = 0$) but that the phase of each mode changes slowly with time in a random fashion.

16.7. The purpose of this problem is to analyze the degradation in signal-to-noise ratio in a heterodyne system when the incoming beams are *not* perfectly aligned.

(a) Consider the case shown in the accompanying diagram, where the beam splitter is misaligned. We assume that the operator has some intelligence, and thus the beams will overlap at the detector. Assume that the detector is a vacuum photodiode whose photocathode emits a current density (amperes/area) in response to an optical power density (watts/area) and that the optical beams are limited-extent uniform plane waves. Plot the degradation in the signal-to-noise ratio as a function of misalignment angle θ.



(b) Suppose that the surface of the detector (say, a vacuum photocathode) is tilted with respect to the direction of the incoming beams. Show that the misalignment of the detector does *not* affect the S/N ratio.

(c) Suppose that the *centers* of the beams are collinear and coincident on the detector but that the *size* of the local oscillator beam is different from that of

the signal beam. Plot the degradation of the signal-to-noise ratio as a function of the ratio w_{LO}/w_s.

(d) Using the results found in (a) to (c), *discuss* other real-life effects that might degrade the signal-to-noise ratio from the ideal situation. For instance, we know that the beam from the local oscillator will be a Gaussian with a finite radius of curvature of the phase front, which will not be matched to the incoming signal. Is that a serious problem?

16.8. Suppose that we follow the avalanche of electron-hole pairs in a region of a semiconductor w units wide. Start with one pair at some point x between 0 and w and assume an equal probability P of an electron or a hole crossing the region w, creating a new pair. Show that the multiplication factor is $M = 1/(1 - P)$.

16.9. Consider a multimode laser with nine equal amplitude modes similar to that used in Sec. 9.5.2.

$$e(t) = \sum E_0 \cos[(\omega + n\omega_c)t + \phi_n] \qquad -4 < n < 4$$

where

ω_0 = center frequency of the laser

$\omega_c = 2\pi \cdot (c/2d)$, the cavity mode spacing

ϕ_n = phase of the n^{th} cavity mode

This field is incident onto a quantum detector, and the amplitude of the beat note at a frequency of $c/2d$ is measured in addition to the normal quantities. Assume the following numerical values: $\eta_{qe} = 0.25$; $\lambda = 6328$ Å, $[(E_0^2/2\eta_0) \cdot \text{area}] = 1$ mW, and $c/2d = 100$ MHz.

(a) If all of the phases are equal, the laser is mode-locked. Compute

(1) The time averaged current

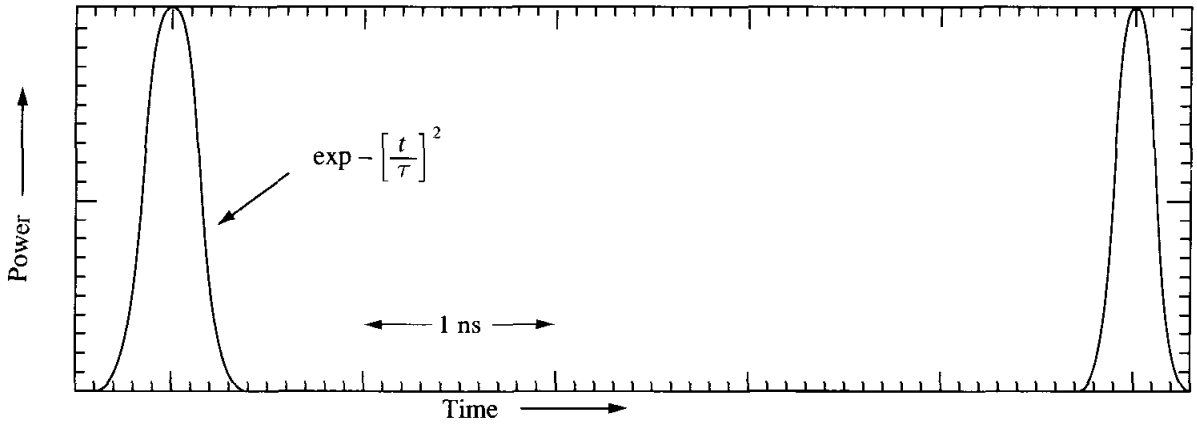
(2) The current as a function of time (identify the peak value and the FWHM)

(3) The amplitude of the $c/2d$ beat note

(b) Now let the phases be distributed according to what follows: $\phi_n/2\pi = 0.2$, 0.7, 0.9, 0.0, 0.2, 0.6, 0.3, 0.5, and 0.9 for n running from -4 to $+4$ (a random number sequence chosen from the Social Security numbers of nine class members). Compute the beat note amplitude at a frequency of $c/2d$ and compare with (a, 3) above.

(c) Use the comparison indicated in (b) to indicate the amplitude of the beat note if the number of modes were large and the phases varied slowly with time (with respect to each other) in a random and uncorrelated fashion.

16.10. The temporal output of a mode-locked laser at $\lambda_0 = 5145$ Å producing an average power of 100 mW is shown on the diagram below.



Note:

$$\int_0^\infty e^{-ax^2}\,dx = \tfrac{1}{2}\sqrt{\frac{\pi}{a}}$$

Assume that $\tau = 0.25$ ns.

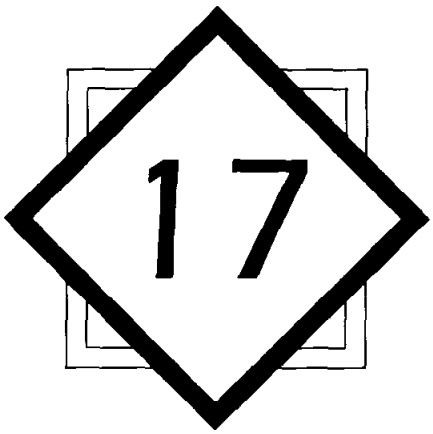
(a) How long is this laser?

(b) If this output is detected with a reversed biased *p-i-n* diode with a quantum efficiency of 20% driving a 100 ohm resistor, what is the output voltage as a function of time? (This amounts to relabeling the vertical axis in terms of volts. Please indicate the peak value.)

(c) How much shunt capacitance can be tolerated before the detected waveform is seriously degraded? (You may use any reasonable criterion for this degradation as long as you specify your choice.)

REFERENCES AND SUGGESTED READINGS

1. B. G. Streetman, *Solid State Electronic Devices*, 2nd ed., Ed. Nick Holonyak, Jr. (Englewood Cliffs, N.J.: Prentice Hall, 1980).
2. A. van der Ziel, *Solid State Physical Electronics*, Ed. Nick Holonyak, Jr. (Englewood Cliffs, N.J.: Prentice Hall, 1968).
3. Jacques I. Pankove, *Optical Processes in Semiconductors*, Ed. Nick Holonyak, Jr. (Englewood Cliffs, N.J.: Prentice Hall, 1971).
4. A. Yariv, *Optical Electronics*, 2nd ed. (New York: Holt, Rinehart and Winston, 1971).
5. *RCA Electro-Optics Handbook*, Technical Series EOH-11, Copyright 1974 by the RCA Corporation.
6. James F. Gibbons, *Semiconductor Electronics* (New York: McGraw–Hill, 1966).
7. Willis W. Harman, *Electronic Motion* (New York: McGraw–Hill, 1953).
8. M. O. Scully and M. Sargent III, "The Concept of the Photon," *Phys. Today*, 38–47, Mar. 1972.
9. G. Margaritondo, "100 Years of Photoemission," *Physics Today 41*, 66–72, 1988.
10. J. R. Barry and E. A. Lee, "Performance of Coherent Optical Receivers," *Proc. IEEE 78*, 1369–1394, 1990.



Gas-Discharge Phenomena

17.1 INTRODUCTION

Much of the qualitative understanding of the gas discharge was obtained in the early 1900s (or even earlier). Indeed, much of today's "modern physics" has its roots in the explanation of one phenomenon or another associated with a gas discharge.* As such, a gas discharge, be it a lamp or a laser, is the "oldest" modern electronic device.

There are many exotic and complicated reasons for the performance of a gas laser, but the prosaic ones are important also, and are sometimes forgotten. A gas is an atomic system in a chaotic state. Consequently, there is very little that humans can do to a gas that has not happened to it before and that will not automatically cure itself if the excitation is removed. (The obvious exception is the electrical initiation of a chemical reaction, in which a new compound is formed such as in the HF laser.) Owing to its tolerance of such mistreatment, we can inject large amounts of power into a gas (with only minimal misgivings) and expect large amounts of optical power to be generated.

Indeed, the simple fluorescent lamp is one of the most efficient light sources available: 70% of the electrical power (i.e., $V \cdot I$) is converted to optical power at 253.7 nm.

*For instance, the whole sequence of events starting with Rydberg's formula, leading to Bohr's theory of the hydrogen atom, and ultimately to Schrödinger's and Heisenberg's quantum mechanics was aimed directly at explaining the spectra emitted by a gas discharge.

Unfortunately, we have to convert that radiation to the visible by utilizing a phosphor. Even so, the overall efficiency is 25% to 30%, better than the best laser. Sodium-vapor lamps and mercury-arc lamps are even more efficient in the production of visible radiation.

One last major advantage of a gas over a solid-state or semiconductor laser lies in its ability to be scaled to large volumes and literally kilograms of active material without straining the budget or manufacturing technology. For instance, a cylinder of CO_2 with $A = 100\ \text{cm}^2$ ($\sim$ 4 in. square) and $l = 100$ cm long at 50 atm contains about 1 kg of an active laser medium at a cost of about 25 cents. Although dry ice is quite cheap, and thus the total cost is ridiculously low, even the cost of more exotic gases is far below that of their solid-state counterparts.

While this example points out the simplicity of scaling, it also points out the inherent problems associated with gas lasers. They tend to be big and require even larger power supplies operating at voltages and currents that are not always compatible with modern solid-state technology.

This chapter attempts to give a primer on gas-discharge theory so as to develop an appreciation of why some gas lasers work so well in spite of the *apparent* nonselective and deceptively simple means of pumping.

We approach a gas discharge from a phenomenological viewpoint and draw on experiment wherever possible. This enables us to go straight to an explanation with a minimal amount of theory. Our goal is to maximize the understanding of the physics for a minimum investment in arithmetic.

First, the terminal characteristics of a simple DC discharge will be described. Unfortunately, we will have to wait until the end of the chapter to give a detailed explanation of the rather strange V–I characteristics. However, it will be made clear that a discharge is *not* a simple positive resistor. If anything it is a negative resistance device that must be ballasted by external means. Failure to do so can result in disastrous consequences (to the power supply, not the gas).

Given these terminal characteristics, we then look more closely at the regions of the discharge. Because of Kirchhoff's current law and some rather obvious physical facts, we find distinct regions of a discharge: the cathode region, consisting of a cathode "dark" space, where a disproportional amount of voltage is dropped; the negative glow, where no voltage is dropped; and the positive column, which occupies the major portion of the inner electrode distance. Even though the positive column is the only *nonessential* part of a gas discharge, 95% of all gas lasers use this region to excite the gas. (The other 5% use the negative glow.)

These simple experimental facts and elementary physics will lead us to some very important conclusions: namely, that a gas discharge is very nearly space charge neutral, and that electrons *and* ions must be created at exactly the same rate as they are lost. This is called the ionization balance condition and is necessary for the *existence* of a discharge.

Once the existence is established (using proof by intimidation), we then look at how the power enters the electron gas and is transferred to the neutrals, exciting the quantum levels of interest. At this point we have to delve deeper into the microscopic domain to introduce a collision cross section for various processes.

Armed with this understanding, we should be able to do a respectable job of predicting the excitation rate of the quantum levels.

17.2 TERMINAL CHARACTERISTICS

A gas discharge is a very simple device: two electrodes separated by a distance l driven by a power supply with a series ballast resistor in the manner shown in Fig. 17.1(a). For the purposes of illustrating some general characteristics, we consider a simple experiment on a very common discharge tube, a fluorescent lamp,* the results of which are plotted in Fig. 17.1(b).

There are many obvious features to be noted:

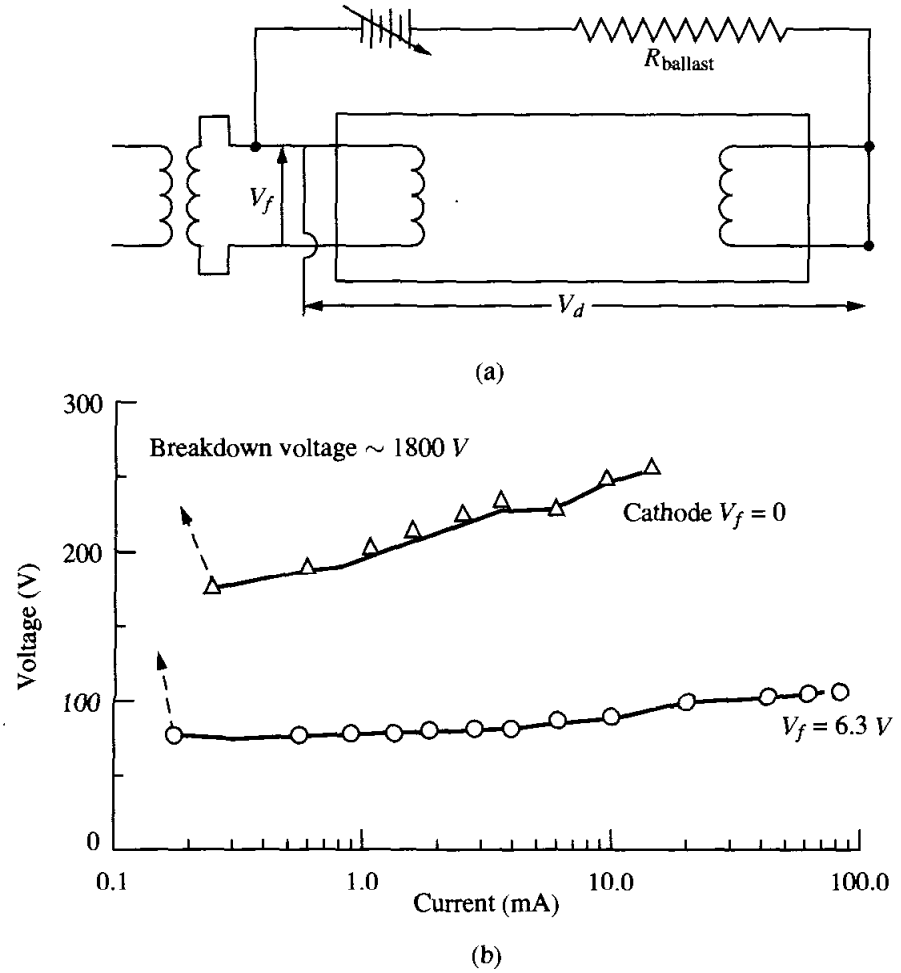


FIGURE 17.1. Experiment on a fluorescent lamp: (a) circuit; (b) data results.

*It is not our purpose here to give a complete explanation of this lamp. Rather, we wish to emphasize some of the "strange" features of any discharge.

a. By no stretch of the imagination can we classify this device by a simple resistance. The voltage changes by only 20 V when the current is varied by three decades. Thus it is most logical to consider a discharge as a current-controlled device.

b. The voltage required to initiate the discharge is very high (1800 V) compared to its normal operating voltage of 80 to 100 V.

c. Whether the cathode can or cannot emit electrons by thermionic emission has a marked effect on the terminal voltage. As we will see next, the excess voltage required for the cold cathode case is dropped across a very small region adjacent to the cathode.

Although these features are peculiar to a common fluorescent lamp, the same general characteristics are present even in the most exotic gas-discharge lasers.

17.3 SPATIAL CHARACTERISTICS

If we were to measure the potential as a function of the distance between the cathode and the anode, data similar to those shown in Fig. 17.2(b) would be obtained and we would find a strong correlation between the "dark" and "bright" regions of the discharge. The potential rises very sharply near the cathode, remains more or less constant for a short distance, and rises more or less uniformly throughout the rest of the discharge length. These regions are called the cathode "dark" space,* the negative glow, and the positive column, for traditional reasons.

We can start to understand this figure and the reasons for these regions by adding some "obvious" physics in succeeding graphs. For instance, we *know* that the total current is constant independent of z, as shown in Fig. 17.2(c). We would guess that most of the current is carried by the mobile electrons, although some could be carried by the massive positively charged ions. Whatever the fraction of the current carried by each type of charge, *their sum must be a constant.*

This simple fact explains the cathode regions and also dependence of the terminal voltage on the emission characteristics of the cathode. If the cathode does not emit very many electrons, then the ions must carry the lion's share of the current in the CDS, as shown in Fig. 17.2(c). But, because of their large mass, this requires a large field, and thus a large cathode fall voltage (typically 100 to 400 V) is required to accelerate the ions to sufficient velocity to carry the current demanded by the circuit. If the cathode emits copious quantities of electrons through thermionic emission, then the ions do not have to carry so much of the current, and the cathode fall voltage drops to roughly the ionization potential of the gas.

The negative glow is caused by those electrons that have gained a kinetic energy corresponding to a significant fraction of the cathode fall voltage. The energetic electrons slow down in the negative glow by exciting and ionizing collisions, and thus this region of a discharge is an "electron beam" produced plasma.

*It is not dark. Rather, the visible emission is much weaker there than in the negative glow and positive column.

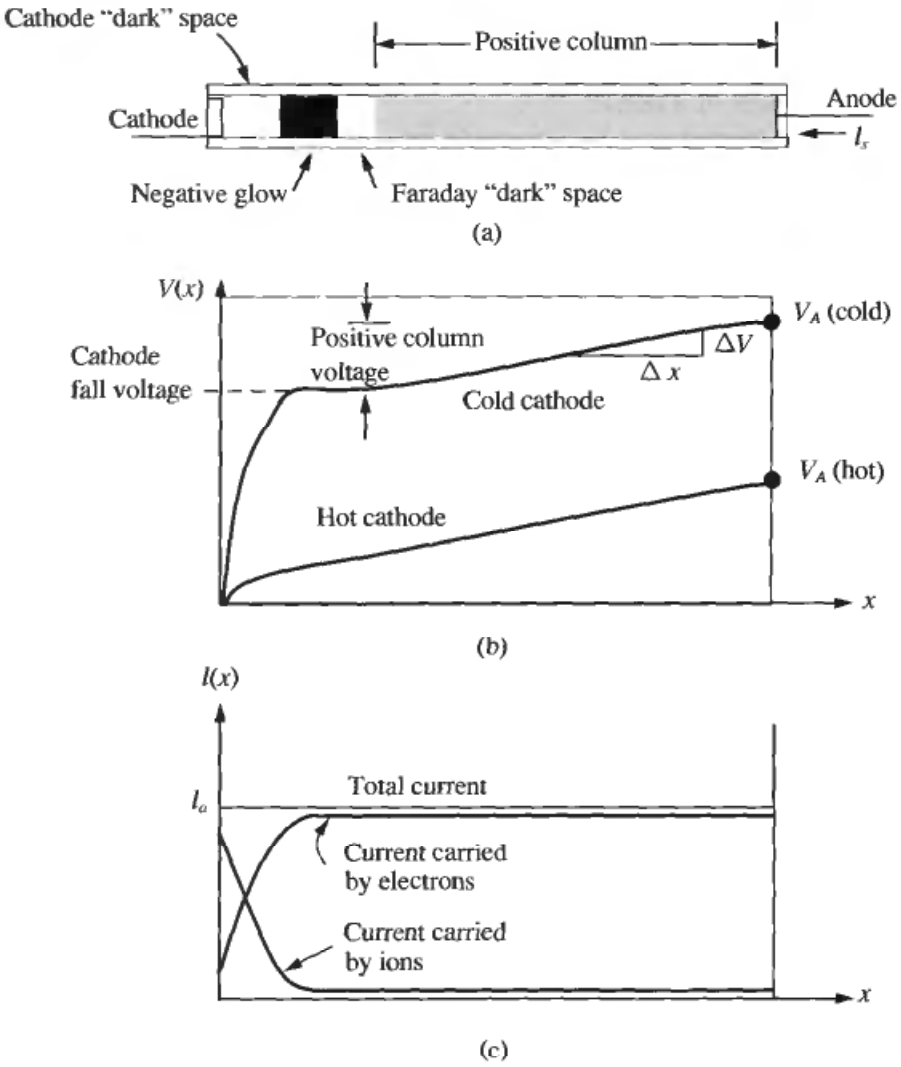


FIGURE 17.2. Regions of a gas discharge.

In the positive column, the potential varies linearly with distance, and thus the field E is a constant. The fact that the field is a constant can be utilized in Gauss's law to specify the most characteristic feature of plasma physics: *space-charge neutrality*.* Thus, while we have free charged carriers, the number density of electrons is compensated, almost perfectly, by an equal number of positive ions.

The electrons, being much lighter, respond to the presence of the electric field in a much more vigorous fashion than do the more sluggish ions. Consequently, almost all of the electrical power enters through the electrons, to be apportioned by them to the other constituents of the system. It is these other constituents, the excited neutral atoms and molecules (and sometimes excited ions), which produce the coherent radiation of the laser.

*Only near surfaces such as the anode, cathode, or walls do space-charge effects play a role. For our purposes here, we ignore these "sheath" regions.

Thus it is most important that we pay close attention to the electron gas and learn how it is produced, how it gains energy from the field, and how it then transfers this energy to quantum states of a laser.

17.4 ELECTRON GAS

17.4.1 Background

In a typical gas laser, the electron density ranges from 10^{10} to 10^{13} cm^{-3}, typical of a low-pressure discharge, to 10^{15} to 10^{17} cm^{-3} for a high-pressure heavily pumped laser.* Obviously, we cannot afford the computational effort to follow the trajectory of each individual electron as it gains energy from the field and loses it on collisions with the more numerous neutral atoms or molecules. Thus a statistical procedure is in order where we treat the electrons as a minority gas interacting primarily with the heavy neutral atoms or molecules.

17.4.2 "Average" or "Typical" Electron

Let us follow the trajectory of a typical electron in the time interval between collisions (i.e., immediately after emerging from a collision and before it makes another, such as that shown in Fig. 17.3). In that time interval, the electron experiences only the force of the applied electric field according to Newton's laws:

$$\left.\begin{aligned} m\frac{d\mathbf{v}}{dt} &= -e\mathbf{E} \\ \text{or} \qquad \mathbf{v} = \mathbf{v}_i - \frac{e}{m}\mathbf{E}t &\triangleq \mathbf{v}_i + \mathbf{v}_{\text{ord}} \end{aligned}\right\} \tag{17.4.1a}$$

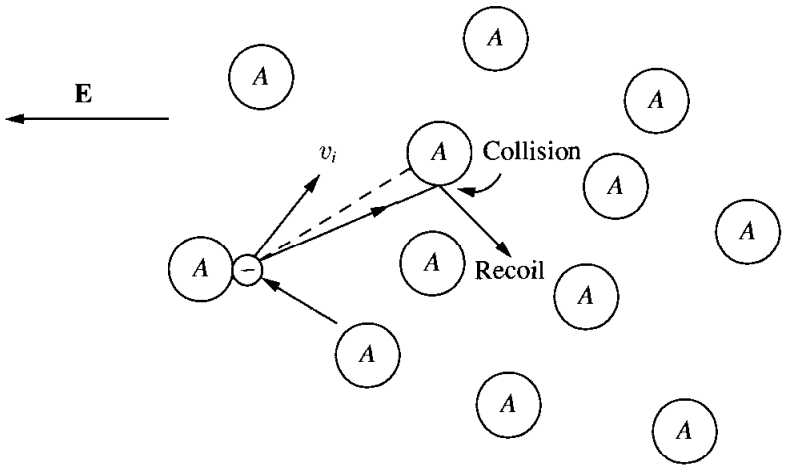


FIGURE 17.3. Trajectory of a typical electron.

*In spite of these large numbers, the degree of ionization of a gas laser seldom exceeds 0.1%.

where $\mathbf{v}_i$ is the electron's initial thermal velocity, which has, of course, any orientation whatsoever with respect to the "ordered" value specified by the electric field. For n_e electrons, the electrical current is

$$\mathbf{i} = \sum_{j=1}^{n_e} (-e)\mathbf{v}_j \tag{17.4.1b}$$

where each $\mathbf{v}_j$ has the format of (17.4.1a).

If we consider these equations and Fig. 17.3 for a moment, then the following comments are at least palatable, if not obvious:

1. Once the electron collides with an atom, A, the recoil becomes a new "initial" velocity to be reapplied in (17.4.1).
2. The electrons collide with the atoms by virtue of their random "initial" or "thermal" velocity. As we will show shortly, (17.4.13c), the "ordered" velocity is much smaller than the rms value of the random velocity; hence, the collision rate does not depend on the field explicitly.
3. This "initial" random velocity is, for the most part, in any of 4π directions, and the collision probability is more or less independent of this direction.
4. Collisions destroy any memory of its prior trajectory and condition. Thus the ordered momentum gained from the field is converted to disordered thermal motion.
5. For n_e large in (17.4.1b) the vector sum of the initial velocities averages to zero, with a net current being conducted by the "drift" motion of the charges in the time between collisions.

Since the gain in momentum of each electron is interrupted by collisions, we modify the momentum balance equation to account for this scattering:

$$m\frac{d\mathbf{w}_d}{dt} = -e\mathbf{E} - m\mathbf{w}_d\nu_c \tag{17.4.2}$$

where ν_c is the collision frequency for momentum scattering and the $\mathbf{w}_d$ (rather than v) is used for the mean drift motion of the electrons. For steady (DC) fields, one can ignore the inertial term and obtain the drift velocity of the "average" electron.

$$\mathbf{w}_d = -\frac{e\mathbf{E}}{m\nu_c} = -\frac{e\tau}{m}\mathbf{E} \tag{17.4.3}$$

with $\tau = 1/\nu_c$ being the mean free time. The drift velocity per unit electric field is the mobility

$$\mu_e = -\frac{w_d}{E} = \frac{e}{m\nu_c} = \frac{e\tau}{m} \tag{17.4.4}$$

Now the collision rate of the electron in a gas depends on the density of scattering atoms, the "size" or cross section of the atom (σ_c), and the relative velocity between the atom and electron. This velocity is almost entirely caused by the random thermal speed of the electron v_{th}.

$$\nu_c = N_A \sigma_c v_{\text{th}} \tag{17.4.5}$$

The net electrical current carried by this drift is

$$\mathbf{J} = n_e e \mathbf{w}_d \tag{17.4.6}$$

By combining (17.4.3) with (17.4.6), we obtain the conductivity in terms of average electron behavior

$$\sigma = \frac{\mathbf{J}}{\mathbf{E}} = \frac{n_e e^2}{m\nu_c} \tag{17.4.7}$$

The fact that the electron gas carries current means that the *electric field* transfers energy *to* the *electron gas* with a rate

$$p_{\text{el}}(\text{W/vol}) = \mathbf{E} \cdot \mathbf{J} = \frac{n_e e^2}{m\nu_c} E^2 \tag{17.4.8}$$

Note that the statement of (17.4.2) has apparently broken the log-jam of mathematical equations, but now they are coming fast and furious. Let us stop here and recapitulate our ideas before proceeding further with the question of the fate of this power.

Let us back up to (17.4.1a) and square it to obtain the kinetic energy of the electron:

$$\tfrac{1}{2} m\mathbf{v} \cdot \mathbf{v} = \tfrac{1}{2} m \left(|\mathbf{v}_i|^2 + 2\mathbf{v}_i \cdot \mathbf{v}_{\text{ord}} + |\mathbf{v}_{\text{ord}}|^2 \right) \tag{17.4.9}$$

Now $\mathbf{v}_i$ is randomly oriented with respect to the ordered velocity and thus the second term averages out to zero for a large number of electrons. We attempt to describe the behavior of this large number by an average electron by considering the fate of the energies represented by the first and last terms of (17.4.9). This attempt is sketched in Fig. 17.4.

Thus the electron gains energy from the field between collisions, loses a very small fraction of its total kinetic energy in an elastic collision (nothing is perfectly elastic), and accumulates this energy until it can make an *inelastic* collision in which the *internal* quantum energy of the atom increases, thereby decreasing the kinetic energy by the corresponding difference. It is this last process that leads to the excited states for a gas laser.

With this picture in mind we can describe the rate at which the mean kinetic energy w_e of the electron gas is changing with time. It is increasing because of the power from

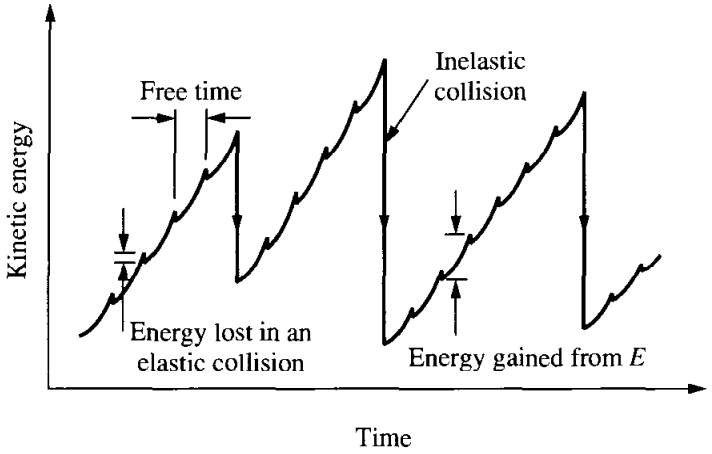


FIGURE 17.4. Possible energy history of an electron.

the electric field and decreasing in small steps by elastic collisions and in large steps by inelastic collisions.

$$\frac{dw_e}{dt} = p_{\text{el}} - \underbrace{\nu_c \delta n_e \left(\tfrac{3}{2}\epsilon_k - \tfrac{3}{2}\epsilon_A\right)}_{\text{gas heating}} - \sum_j \underbrace{n_e \nu_{\text{inel}} \Delta W_j}_{\text{excitation}} \tag{17.4.10}$$

where

w_e = density of electrons times the average energy of the typical electron.
$w_e = n_e(\frac{3}{2}\epsilon_k)$
ϵ_k = a characteristic energy of the electrons (kT_e)
ϵ_A = a characteristic energy of the atoms (kT_A)
δ = fraction of the excess energy lost per elastic collision = $2m/M$
ν_{inel} = inelastic collision rate
ΔW_j = energy lost in an inelastic collision

Equation (17.4.10) contains much more in the way of definitions than science. Indeed, the various terms are introduced to aid in the physical reasoning. The first term is the power entering the electron gas from the field, and the last two specify how and why the power leaves the electron gas. As Fig. 17.4 implies, each electron loses a small fraction of its excess kinetic energy, even in an elastic collision. Since there are n_e electrons with an average collision rate of ν_c, the power lost to the neutrals in the form of gas heating is the product of the density of electrons, the collision rate ν_c, and the difference between the characteristic energies of the electrons and neutrals.*

The gas excitation term follows from a similar line of reasoning. There are n_e electrons, with each electron making on the average ν_{inel} excitation collisions per second, with each event costing an energy ΔW_j. Thus the last term of (17.4.10) represents the following excitation event

$$e(\text{KE}) + \text{A} \rightarrow \text{A}^*(\Delta W) + e(\text{KE} - \Delta W) \tag{17.4.11}$$

and A* may be a rotational, vibrational, or electronic state of an atom or molecule.

The characteristic energies, ϵ_k, can be related to the temperature of the various gases by

$$\epsilon \Rightarrow kT \tag{17.4.12}$$

This procedure implies that the various gases have a Maxwellian velocity distribution, an excellent approximation for the atoms but a rather poor one for the electrons, as will be seen later.

Some very important trends associated with all gas discharges can be shown by combining (17.4.8) and (17.4.10) under some very simplifying assumptions. Let us ignore the inelastic term in (17.4.10) and thus condemn the conclusions to be strictly applicable

*If the electrons are isothermal with the neutrals, no energy is interchanged, since just as many electrons are heated by neutrals as neutrals are by electrons.

to an inefficient laser or light source, since, by definition, an insignificant amount of power is used to excite the quantum levels. By doing so, we strip away unnecessary mathematics and obtain a very important result. For a steady-state system with $dW_e/dt = 0$,

$$0 = \frac{n_e e^2}{m\nu_c} E^2 - n_e \nu_c \delta \tfrac{3}{2} (\epsilon_k - \epsilon_A) \tag{17.4.13a}$$

or

$$\frac{3}{2}(\epsilon_k - \epsilon_A) = \frac{2}{\delta}\frac{1}{2} m \left(\frac{eE}{m\nu_c}\right)^2 \tag{17.4.13b}$$

$$= \frac{2}{\delta}\left(\frac{1}{2} m w_d^2\right) \tag{17.4.13c}$$

(17.4.13c) provides a promised connection between the mean energy of the electrons and the kinetic energy associated with the drift motion of the swarm. Note that δ, a very small number, appears in the denominator. Thus the random kinetic energy of the electron, $\frac{3}{2}\epsilon_k$, is much greater than the drift energy $\frac{1}{2} m w_d^2$ imparted to the electrons by the field. We can state these conclusions in reverse order; namely, (17.4.13c) states that the drift motion is a very small fraction of the random kinetic energy.

The form of (17.4.13b) demonstrates another quite general trend associated with gas discharges if the explicit formula for ν_c (17.4.5) is used.

$$\frac{3}{2}(\epsilon_k - \epsilon_A) = \frac{2}{\delta}\frac{1}{2} m \left(\frac{e}{m\sigma_A v_{\text{th}}}\right)^2 \left(\frac{E}{N}\right)^2 \tag{17.4.13d}$$

We could approximate the thermal velocity by $\frac{1}{2} m v_{\text{th}}^2 = \frac{3}{2}\epsilon_k$ and find an explicit expression for ϵ_k as a function of E/N. Frankly, the arithmetic mess is not worth it but the following trend should be clear.

The characteristic energy, ϵ_k, (or temperature) of the electron gas is a monotonically increasing function of the ratio of electric field to neutral gas density:

$$\epsilon_k = f\left(\frac{E}{N}\right) \tag{17.4.14}$$

This trend is perfectly general in spite of the approximations made.

It is interesting to apply this trend to the data shown in Fig. 17.1. Since the lamp voltage was a constant (more or less), we now know that the electron "temperature" is independent of current. (In fact, if we had measured the positive-column voltage only, we would have found that it *decreased* with current and thus the electron "temperature" does also.) This should not be construed to imply that we do not increase the power to the discharge with increasing current. We do. However, the average energy of electron gas does not change to any extent.

Even though the picture of an "average" electron is simple and appealing, it has some basic conceptual difficulties. The "average" electron does not have the energy to make the inelastic collisions essential for the excitation of the electronic quantum levels of interest

in a laser. The CO_2 laser is an exception, and this fact alone accounts for the inherent efficiency (more about this later). It is the "exceptional" electron that performs the function of exciting the mercury in the fluorescent lamp, and the amazing part is that 60% to 70% of the electrical power is utilized by these "exceptional" electrons. The following example can be utilized to emphasize this conceptual difficulty as well as to give a feeling for the numbers involved in the mathematics given above.

Example: The Fluorescent Lamp of Fig. 17.1

This lamp is 48 in. long and filled with about 3 torr of argon with a small drop of mercury. The vapor pressure of mercury at normal operating temperature is about 30 mtorr. (The energy-level diagram for mercury is given in Fig. 15.1.) Since there is so little mercury and so much argon, the elastic collision rate is almost entirely controlled by the argon.

Let us consider the case when the cathode is a good thermionic emitter so that the cathode fall can be estimated to be about 15 V. Then E/N of the positive column can be found from the measured voltage at 10 mA, length, and pressure specifications (see Fig. 17.1): $E = (V - V_k)/L$, $V = 90$ V, $l = 48 \times 2.54$ cm, therefore $E = 0.62$ V/cm $= 62$ V/m. $p = 3$ torr; therefore, $N_A = 3 \times 3.54 \times 10^{16}$ cm^{-3} and $E/N = 5.79 \times 10^{-18}$ V-cm^2. To compute the drift velocity by (17.4.3), we need an estimate of the elastic collision rate ν_c, which requires, in turn, an estimate of the characteristic energy of the electrons. (This is typical of gas-discharge calculations. We need the final answer before we can start.) Let us anticipate that the characteristic energy ϵ_k will be close to 1.0 eV and then force the calculations to yield a self-consistent picture. With this initial guess, the thermal velocity is found from $\frac{1}{2}mv_{th}^2 = \frac{3}{2}e(\epsilon_k)$ (ϵ_k in volts). Therefore $v_{th} = 7.26 \times 10^8$ cm/sec.

The elastic collision cross section in argon is a strong function of the electron energy, as exemplified by the values shown in Table 17.1. For our *rough* calculations here, we pick the value of the cross section at the characteristic energy and find $\sigma_c = 1.5 \times 10^{-16}$ cm^2 and then use the thermal velocity computed previously to estimate $\nu_c = N_A \sigma v_{th}$.

[Eq. (17.4.5) $= 1.16 \times 10^9$ collisions/sec; (i.e., roughly 1 collision per ns)]

Thus an estimate for the drift velocity is given by (17.4.3):

$$w_d = \frac{eE}{m\nu_c} = 9.3 \times 10^3 \text{ m/sec} = 9.3 \times 10^5 \text{ cm/sec}$$

As promised, the drift is much smaller than the thermal velocity and the drift energy is small indeed.

TABLE 17.1

Total Scattering Cross Section in Argon			
ϵ (V)	σ_c ($\times 10^{-16}$ cm^2)	ϵ (V)	σ ($\times 10^{-16}$ cm^2)
0.107	1.27	0.693	0.745
0.198	0.36	1.01	1.49
0.257	0.155	1.51	2.33
0.295	0.161	2.09	3.18
0.465	0.31	2.48	4.06

From Kieffer (ref. 1).

$$\tfrac{1}{2} m w_d^2 = 2.47 \times 10^{-4} \text{ eV}$$

If every collision were elastic, then δ in (17.4.10) would be $2m/M$, with M being the mass of the argon atom (40 amu). But we know that the fluorescent lamp does work (it does produce light), and thus there must be many inelastic collisions. To avoid evaluating the last term of (17.4.10), we pick an effective δ of five times the classical value and use the simplified form (17.4.13c). (NOTE: This is equivalent to assuming that 80% of the electrical power is transferred from the electrons to the neutrals by processes other than elastic collisions.)

$$\frac{3}{2}(\epsilon_k - \epsilon_A) = \frac{2}{\delta}\left(\tfrac{1}{2} m w_d^2\right) \qquad \text{with} \quad \delta \sim 5 \times \frac{2m}{M} = 1.36 \times 10^{-4}$$

or

$$\epsilon_k = 2.44 \text{ eV}$$

Obviously this is not consistent with our initial estimate of $\epsilon_k = 1.0$ eV. Hence, we revise our estimate to 2.09 eV and repeat the previous five calculations. Table 17.2 shows how one rapidly converges to the self-consistent *mathematical* solution for this lamp. Table 17.2 was worked out to a far greater precision than the accuracy of the theory warrants. As such, Table 17.2 should be used solely as an example for the physical thought involved and as an example of how we force mathematics to conform to the physics.

As we shall see subsequently, using (17.4.3) through (17.4.14) is a gross simplification, and thus the exact numerical answers are in error. However, the trend expressed in Table 17.2 is correct; namely, the drift velocity is much smaller than the random thermal velocity, the drift energy ϵ_d is much smaller than the characteristic energy ϵ_k, and the collisions, even in this low-pressure gas, are frequent indeed.

Nevertheless, the numerical values given here are typical, in spite of the gross oversimplification in theory. These numbers serve to point out the inherent weakness of the average or "typical" electron approach.

For instance, Table 17.2 indicates that a characteristic energy of $\sim$ 1.25 eV will yield a self-consistent set of parameters. However, the energy of the first excited quantum state in argon, 11.54 eV, is large compared to this quantity, and thus almost no excitation is seen in the majority gas. The resonant 6^3P_1 state of mercury is 4.9 eV, and in spite of its minority

TABLE 17.2 Calculations on the Lamp Assuming That $\delta = 5 \times 2m/M = 1.36 \times 10^{-4}$

ϵ_k (V)	v_{th} ($\times 10^5$ m/s)	σ_c ($\times 10^{16}$ cm^2)	ν_c ($\times 10^9$ s^{-1})	w_d ($\times 10^3$ m/s)	ϵ_d ($\times 10^{-4}$ V)	ϵ_k (V)
1.0	7.26	1.5	1.16	9.3	2.47	2.47
2.09	10.49	3.18	3.54	3.04	0.26	0.26
1.5	8.9	2.3	2.17	4.98	0.71	0.71
1.4	8.6	2.14	1.96	5.52	0.87	0.87
1.2	7.95	1.81	1.53	7.07	1.42	1.42
1.3	8.28	1.99	1.75	6.18	1.09	1.09
1.26	8.14	1.91	1.65	6.53	1.22	1.22
1.25	**8.11**	**1.89**	**1.63**	**6.62**	**1.25**	**1.25**

status, considerable excitation is seen. Indeed, it is the radiation from this state at 254 nm that is responsible for the excitation of the phosphor.

But even this energy is large compared to the characteristic energy of the electrons. Thus we are forced to conclude that the average electron does *not* excite the quantum levels of interest.

17.4.3 Electron Distribution Function

The example of the preceding section points out the inherent weakness of the "average" electron approach: Phenomena such as electronic excitation and/or ionization require electron energies far in excess of the average of the electron gas. To compute these quantities we must specify the fraction of the electron number density that has sufficient energy to perform the various excitations of interest. That specification is the electron distribution function.

For instance, all electrons can be influenced by the electric field, and thus all electrons participate in the conduction of current. All electrons undergo elastic collisions, and thus all contribute to gas heating. But only a fraction have enough energy to excite a vibrational level, a smaller fraction can excite an electronic level in the neutral gas, and a much smaller fraction can ionize the atoms.

The defining equation for the electron distribution function is

$$\frac{dn_e}{n_e} = f(v_x, v_y, v_z)\, dv_x\, dv_y\, dv_z \tag{17.4.15}$$

Thus f is the density of the electrons in the three-coordinate velocity space. Since the electrons have to be somewhere in this space, we also have a normalization condition:

$$\iiint_{-\infty}^{+\infty} f(v_x, v_y, v_z)\, dv_x\, dv_y\, dv_z = 1 \tag{17.4.16}$$

Fig. 17.5(a) shows the geometrical interpretation of the distribution function and also serves to illustrate the conversion from velocity coordinates (v_x, v_y, v_z) to that speed, that is, v *without* a subscript, or kinetic energy, ϵ.

$$\tfrac{1}{2} m v^2 = \tfrac{1}{2} m (v_x^2 + v_y^2 + v_z^2) = \epsilon \tag{17.4.17}$$

Thus, if $f(v_x, v_y, v_z)$ is specified, one goes through the mathematical gyration of converting the Cartesian velocity coordinates to spherical ones: the speed and the two angles.

As we would guess, the distribution function is almost (but not quite) spherically symmetric, and thus we need only specify its value in terms of one variable, either speed or kinetic energy. This enables us to plot the density of electrons per unit of speed or energy as shown in Fig. 17.5(b) and (c). As shown in Fig. 17.5(c), it is convenient to define a different quantity, F (of energy ϵ), to describe the fraction of electrons within an energy interval $d\epsilon$ at ϵ.

$$\frac{dn_e}{n_e} \triangleq F(\epsilon) d\epsilon \triangleq 4\pi v^2 f(v)\, dv \triangleq f(v_x, v_y, v_z)\, dv_x\, dv_y\, dv_z \tag{17.4.18}$$

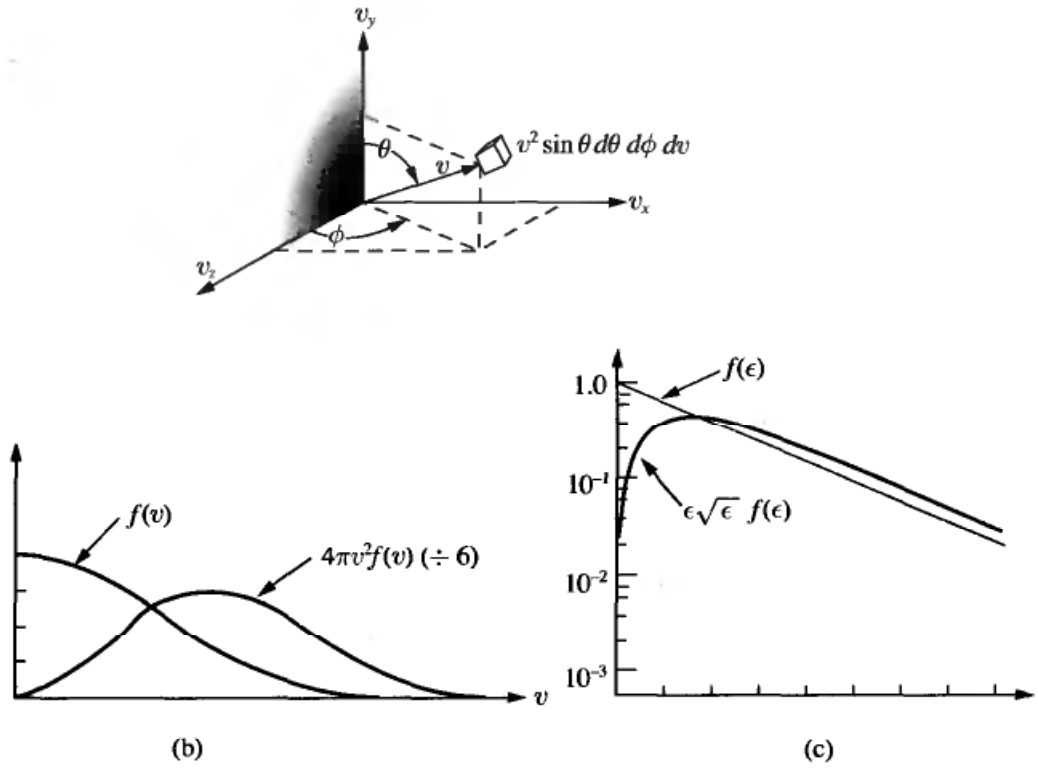


FIGURE 17.5. Electron distribution function: (a) the geometry and relation between the speed and the Cartesian velocities, (b) a plot of $f(v)$ and $4\pi v^2 f(v)$ *assuming* spherical symmetry, and (c) a plot of $f(\epsilon)$ and $\epsilon^{1/2} f(\epsilon)$ *assuming* spherical symmetry.

where the relations between v, v_x, v_y, v_z, and ϵ are specified by (17.4.17). This assumption of spherical symmetry will return to haunt us in the next section.

To be specific and to illustrate the various methods of describing the distribution function, let us assume a spherically symmetric Maxwell–Boltzmann distribution of electron velocities:

$$f_0(v_x, v_y, v_z) = A \exp\left[-\frac{1}{2}\frac{m(v_x^2 + v_y^2 + v_z^2)}{kT_e}\right] \tag{17.4.19a}$$

where

kT_e = a characteristic energy of the electron gas = ϵ_k

T_e = electron "temperature"

The normalization condition, (17.4.16), evaluates the constant A, and the final form of f_0 is

$$f_0(v_x, v_y, v_z) = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left[-\frac{1}{2}\frac{m(v_x^2 + v_y^2 + v_z^2)}{kT_e}\right] \tag{17.4.19b}$$

We now use the string of equalities expressed by (17.4.17) and (17.4.18) to determine

$$F(\epsilon) = 4\pi v^2 f_0(v)(dv/d\epsilon)$$

with

$$v^2 = v_x^2 + v_y^2 + v_z^2 \qquad v = \left(\frac{2\epsilon}{m}\right)^{1/2} \qquad \frac{dv}{d\epsilon} = \frac{1}{2}\left(\frac{2}{m}\right)^{1/2} \epsilon^{1/2}$$

or

$$F(\epsilon) = \left(\frac{1}{kT_e}\right) \frac{2}{\pi^{1/2}} \left(\frac{\epsilon}{kT_e}\right)^{1/2} \exp\left(-\frac{\epsilon}{\epsilon_k}\right) \qquad \epsilon_k = kT_e \tag{17.4.20}$$

The factors preceding the exponential term come from such mundane mathematical gyrations as the normalization and the transformation from (v_x, v_y, v_z) coordinates to speed or energy coordinates. Therefore terms such as those will be present in all energy distribution functions. However, the last factor is the controlling term and is called the *spherically symmetric* part of the distribution function, which will be denoted by the special symbol f_0.

Quite often, only this spherically symmetric part of the distribution function is given: It may be 1 at zero energy as it is here, its amplitude may be specified in terms of a particular author's favorite units, or it may be expressed in arbitrary units. Only its relative shape is important, because we can always retreat to (17.4.19a) and redo the next two equations. Some authors prefer to plot only f_0 in the manner indicated in Fig. 17.5(c) (i.e., log f_0 versus ϵ). The reason should be clear. If one has a simple exponential term for this spherically symmetric part, the graph will be a straight line. If it is not a Maxwellian, it will be obvious at a glance.

Most electron energy distribution functions are not given by a simple Maxwellian, and thus the word "temperature" has no meaning whatsoever. However, the characteristic energy, ϵ_k, which takes on the value of kT_e for a Maxwellian, continues to have validity and, most important, can be measured as a function of E/N.

17.4.4 Computation of Rates

The determination of the spherically symmetric part of the distribution function is beyond the scope of this book. We will assume that someone else has done those calculations and that they are available. What can we do?

Everything that is desired to compute the excitation and transport processes in a laser (or a lamp) can be found from f_0 provided, of course, that the cross section for a particular process is known. (However, we must assume that these data are known, for without them, f_0 cannot be found.) The purpose of this section is to provide a prescription for this computation.

For any process r that depends primarily on the random motion of the electrons, the average of the process is computed by first identifying that rate for a single electron at a specific energy (or velocity), multiplying by the fraction of electrons that have this energy

[or $F(\epsilon)\,d\epsilon$], and summing over all energies (or velocities).

$$\langle r\rangle = \int_0^\infty r(\epsilon)F(\epsilon)\,d\epsilon \tag{17.4.21a}$$

or

$$\langle r\rangle = \iiint_{-\infty}^{+\infty} r(v)f(v_x, v_y, v_z)\,dv_x\,dv_y\,dv_z \tag{17.4.21b}$$

For instance, let the rate be the elastic collision frequency, ν_c, which played such an important role in Sec. 17.4.2. Then the microscopic rate $r = \nu_c = N_A\sigma_{\text{el}}(\epsilon)v$, with $v = (2\epsilon/m)^{1/2}$. If we assume a Maxwellian distribution and assume that σ is a constant, independent of energy, then all integrations can be carried out in terms of elementary functions:

$$\langle \nu_c\rangle = \int_0^\infty \underbrace{N_A\sigma_{\text{el}}\left(\frac{2\epsilon}{m}\right)^{1/2}}_{\nu_c} \underbrace{\frac{2}{\sqrt{\pi}}\left(\frac{\epsilon}{kT_e}\right)^{1/2}\exp\left(-\frac{\epsilon}{kT_e}\right)d\left(\frac{\epsilon}{kT_e}\right)}_{F(\epsilon)d\epsilon} \tag{17.4.22a}$$

$$= N_A\sigma_{\text{el}}\left(\frac{8kT_e}{\pi m}\right)^{1/2}\int_0^\infty xe^{-x}\,dx \tag{17.4.22b}$$

$$= N_A\sigma_{\text{el}}\left(\frac{8kT_e}{\pi m}\right)^{1/2} = N_A\sigma_{\text{el}}\langle\bar{v}\rangle \tag{17.4.22c}$$

where $\langle\bar{v}\rangle$ is the mean speed of the assumed Maxwellian distribution, $(8kT_e/\pi m)^{1/2}$. Although the answer suffers from all the assumptions made, it is reasonably close to the correct answer.

If we request an inelastic collision rate, ν_{inel}, the format provided by (17.4.12) is correct, but the limits of integration must be changed. For instance, it would be silly to consider the rate of 1-eV electrons exciting a 5-eV electronic level. That rate is 0. Thus, given the energy level ϵ_j above the ground state, we find

$$\langle \nu_{\text{inel}}\rangle = N_A\sigma_{\text{inel}}\left(\frac{8kT_e}{\pi m}\right)^{1/2}\int_{\epsilon_j}^\infty \left(\frac{\epsilon}{kT_e}\right)\exp\left(-\frac{\epsilon}{kT}\right)d\left(\frac{\epsilon}{kT_e}\right)$$

$$= N_A\sigma_{\text{inel}}\langle\bar{v}\rangle\left(1+\frac{\epsilon_j}{kT_e}\right)\exp\left(-\frac{\epsilon_j}{kT_e}\right) \tag{17.4.23}$$

where it has been assumed that σ_{inel} is a constant for $\epsilon > \epsilon_j$ and is 0 otherwise. That assumption and that of a Maxwellian distribution were made for the sole purpose of illustrating the procedure. In all but contrived problems, these integrals must be evaluated numerically.

Note that the probability of an inelastic collision decreases as the exponential factor, $\exp(-\epsilon_j/kT)$. Even though the inelastic cross section may be comparable to the elastic one, the rates of the two types of collisions are considerably different. For instance, we found that the elastic rate was 1.63×10^9 sec^{-1} in Table 17.2. Using a constant cross section of 1.89×10^{-16} cm^2 above a threshold of 11.54 eV for excitation of argon atoms, a characteristic energy of 1.25 eV, we obtain $\nu_{\text{inel}} = 1.63 \times 10^6$ sec^{-1}, far smaller than the elastic value.

If we attempt to use (17.4.21b) to predict a transport quantity, say, the drift velocity, we run into a serious problem of our own making—all answers come out to be identically zero! The reason is that current is a vector quantity—a drift in a specific direction with respect to the field. In mathematical terms, (17.4.21) would tell us to multiply an odd function, the drift, by a spherically symmetric velocity distribution, an even function, and sum over velocities, a procedure that guarantees a zero result.

17.4.5 Computation of a Flux

The assumption responsible for the null result is that the distribution function is spherically symmetric, whereas, in fact, it is not. In the presence of an electric field that drives the current, the distribution function is skewed in the direction of the drift motion of the electrons. We can obtain a very good estimate of this anisotropy by using some rather transparent reasoning and some elementary vector analysis.

For a steady state situation, the net change in the distribution function (whatever it may be) with respect to time from *all* causes is, by definition, zero. Thus the electric field accelerates the electrons in a definite direction and therefore drives the distribution function away from spherical symmetry, whereas collisions scatter the directed velocity and thus try to return the system to the isotropic state. The time rate of change of this interplay is zero:

$$\mathbf{a} \cdot \nabla_v f(v) + \nu_c(f - f_0) = 0 \tag{17.4.24a}$$

or

$$f = f_0 - \frac{\mathbf{a} \cdot \nabla_v f}{\nu_c} \tag{17.4.24b}$$

$$= f_0 + \frac{e\mathbf{E}}{m\nu_c} \cdot \nabla_v f \tag{17.4.24c}$$

where f_0 is the spherically symmetric part of the distribution function and $\mathbf{a}$ is the acceleration from the electric field ($\mathbf{a} = -e\mathbf{E}/m$ as usual). Note that the format of the second term of (17.4.24a) implies that any deviation from spherical symmetry would disappear as $\exp(-\nu_c t)$ if the electric field were suddenly removed from the plasma. This is precisely what our intuition would suggest.

Now it is a matter of patience and fortitude with vector analysis to obtain an expression for the electric current. For an electric field in the z direction, we must evaluate

$$J_z = -ne \iiint v_z f \, dv_x \, dv_y \, dv_z \tag{17.4.25a}$$

with f provided by (17.4.24b), and since we assume that the anisotropy is small, it is permissible to replace f in (17.4.24b) by f_0 in the gradient term. It is convenient and conventional to use the electron speed v [the radial velocity coordinate of Fig. 17.5(a)] rather than the Cartesian coordinate velocities v_x, v_y, v_z in the development. Thus

$$v_z = v \cos\theta$$

$$\nabla_v f = \frac{\partial f}{\partial v}\mathbf{a}_v + \frac{1}{v}\frac{\partial f}{\partial \theta}\mathbf{a}_\theta + \frac{1}{v \sin\theta}\frac{\partial f}{\partial \phi}\mathbf{a}_\phi$$

$$\mathbf{a}_v \cdot \mathbf{a}_z = \cos\theta \qquad \mathbf{a}_z \cdot \mathbf{a}_\theta = -\sin\theta \qquad \mathbf{a}_z \cdot \mathbf{a}_\phi = 0$$

$$dv_x\, dv_y\, dv_z = v^2 \sin\theta\, d\theta\, d\phi\, dv$$

where $\mathbf{E} = E_z \mathbf{a}_z$. After integrating with respect to θ and ϕ, the following formula is obtained for the electrical current in terms of the isotropic part of the distribution function:

$$J_z = -\frac{4\pi}{3}\frac{ne^2}{m}\int_0^\infty \frac{E}{\nu_c(v)} v^3 \frac{\partial f_0}{\partial v}\, dv \tag{17.4.25b}$$

It is interesting to note that *if* the isotropic part of the electron distribution function is given by a Maxwellian (17.4.19b) and *if* the collision frequency ν_c is independent of electron speed (a constant mean free time), then the electrical conductivity, J/E, found from (17.4.25b) is identical to the simple formula found earlier, (17.4.7). Seldom are the "if's" satisfied. Hence, the more exact theory must be used for detailed calculation. However, the simple formula (17.4.7) can be used when only rough ($\sim$ 2 times) numbers are desired.

Quite often the drift velocity has been *measured* as a function of E/N, and thus the *correct* answer is available without any effort on our part. Of course, the answer will probably be presented in a graphical format rather than by an elementary function, but it is rigorously correct. An example of this approach will be presented in Sec. 17.6.5.

17.5 IONIZATION BALANCE

For any of the foregoing theories to be useful or pertinent, we must have the electrons. Such a mundane statement is very important, because the ionization balance condition controls the E/N of a discharge, and E/N in turn controls the performance of the laser.

To appreciate the implications of such a trivial and transparent statement, consider the continuity equation for electrons:

$$\frac{dn_e}{dt} = \text{production of new electrons (and ions) by those present in a discharge}$$
$$+ \text{production of new electrons by an externally controlled source (e.g., an electron beam, UV ionization, x-rays)}$$
$$- \text{losses of electrons by various processes in the plasma} \tag{17.5.1}$$

To make the problem simple and transparent, we first ignore the external source and pick a simple lifetime description for the electron losses; that is, the loss is proportional to n_e divided by some lifetime τ_e (which is presumed to be known). Now the first term in (17.5.1) is nothing more than an inelastic collision of an electron with a neutral atom producing an ion and a second electron. Thus the production of new electrons is the product of n_e and the ionization frequency of the electron gas:

$$\frac{dn_e}{dt} = +\nu_i n_e - \frac{n_e}{\tau_e} \tag{17.5.2}$$

If we take this equation literally, it "proves" that a discharge cannot exist. For if n_e is zero, dn_e/dt is also, and thus n_e will never change from zero. However, the equation is too simple. The external source is never exactly zero, and thus there are always a few electrons around to start the process. Obviously, ν_i must be larger than $1/\tau_e$ for the electron density to grow to a respectable value from the initial value.

Having talked ourselves out of one difficulty, in the starting of a discharge, we back into another. Why and how is an equilibrium between production and loss ever established? Although we might first guess that the assumptions for (17.5.2) are again at fault, the answer usually lies with a much more practical and mundane consideration. Our power supply is finite!

As more and more electrons are produced, a corresponding increase in current is passed by the discharge, the available voltage across the discharge tube drops, which lowers the E/N, which, in turn, lowers ν_i until an equality is established between the production and loss terms of (17.5.2). This self-limiting process is sketched in Fig. 17.6. At point A, the voltage across the lamp is large, E/N is large, and $dn_e/dt > 0$. Therefore the current increases. At C, the reverse is true and the current decreases. At B, we have a balance.

Since an ionization event is just a special type of inelastic collision event, we can use (17.4.23b) to proceed further in our analysis of the DC discharge provided that we set ϵ_j equal to the ionization energy of the gas. We set $dn/dt = 0$ in (17.5.2) and find a single equation for the characteristic electron energy.

$$N_A \sigma_i \left(\frac{8kT_e}{\pi m}\right)^{1/2} \left(1 + \frac{\epsilon_j}{kT_e}\right) \exp\left(-\frac{\epsilon_j}{kT_e}\right) = \frac{1}{\tau_e} \tag{17.5.3}$$

Although not a simple task, (17.5.3) could be solved for this characteristic energy. The many assumptions involved make the effort somewhat futile. But the trends expressed here are most important and hold true even with the most esoteric type of discharge.

For a given gas (which specifies the ionization energy ϵ_i and σ_i) at a specified pressure (which specifies N_A) and loss rate (which may be much more complex than the simple lifetime model used here), there is a unique value of the characteristic energy which satisfies the ionization balance equation. Inasmuch as ϵ_k and E/N are related [see (17.4.14)] then

There is a unique value of E/N for a self-sustained steady state discharge.

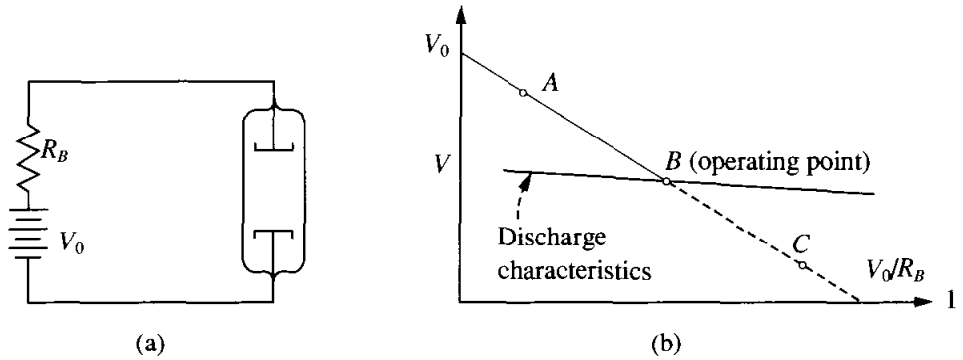


FIGURE 17.6. (a) Circuit. (b) Approach to equilibrium for a DC discharge.

Thus a solution to an equation such as (17.5.1) combined with one similar to (17.4.14) yields the maximum, minimum, and *only* value of E/N permitted for a self-sustained operation. Note that current or electron density does not enter in this balance, and thus E/N is independent of current. This fact explains the flat V–I characteristic of the fluorescent lamp of Fig. 17.1. Some of these conclusions change slightly for an electron-loss process, which has a stronger electron density dependence, but it is only a minor change.

If an external source, such as an electron beam, is used to provide additional ionization, the E/N can be lowered. What happens is that the external source makes up for the decrease of internal ionization due to the lower E/N. This role of the external source enables one to obtain the best E/N ratio for very efficient excitation of levels of the CO_2 laser.

17.6 EXAMPLE OF GAS-DISCHARGE EXCITATION OF A CO_2 LASER

17.6.1 Preliminary Information

Let us close this chapter with a practical example of laser excitation by a gas discharge, that of a high-pressure CO_2 laser in the transverse electric atmospheric (TEA) configuration shown in Fig. 17.7.

We will concentrate on the gas-discharge problem with the goal of determining what fraction of the electrical power drawn from the power supply is used to create the upper (001) state of the CO_2 molecule and how the rest of the power is apportioned to the other processes. Fortunately, this problem has been addressed with detailed experiments and theoretical calculations by the members of the Westinghouse Research Laboratory,* with the results being published in two classic papers (see Refs. 2 and 3). The purpose of this section is to understand and appreciate the significance of this work. We will *not* address the excited-state kinetics and optical coupling problem.

17.6.2 Experimental Detail and Results

The geometry shown in Fig. 17.7 is typical of those used to excite high-power lasers. The experiment consisted of charging the capacitor to a sufficiently high voltage, closing the switch (a triggered spark gap), and measuring the current and voltage.

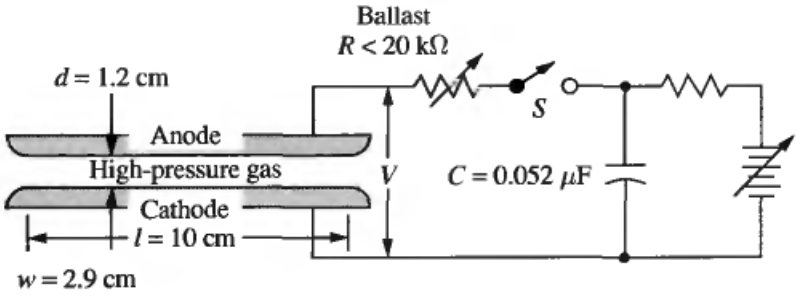


FIGURE 17.7. Geometry of a small high-pressure CO_2 laser discharge (as described in L. J. Denes and J. J. Lowke, *Appl. Phys. Lett. 23*, 130, Aug. 1973.)

*Many other workers and groups have addressed this problem. However, these two papers [2, 3] are unique in the sense that they are readily available, easily read, and complete in every sense of the word, and relate theory to experiment directly.

When the switch is closed, the full capacitor voltage appears across A-K. However, after a very short time interval of $\sim$ 1 to 2 ns, the ionization in the gap increases, which allows the current to increase, which, in turn, lowers the voltage across the gap due to the IR drop across the series resistance and the unavoidable $L(di/dt)$ drop in the wiring inductance. Eventually (in about 20 ns), an equilibrium is established in the manner described in the preceding section (see Fig. 17.5): The current is limited such that the A-K voltage plus IR drop equals the capacitor voltage. From that time on, one can consider the experiment as a DC discharge.

If this type of procedure is repeated for various values of series resistance and total pressure, the electrical V–I characteristics of the discharge can be determined. A collection of such data is shown in Fig. 17.8, where three different mixtures are used in combination with three different total pressures.

Note that the V–I terminal characteristics of this exotic laser mixture are similar to those of the lowly fluorescent lamp described in Fig. 17.1, with the maintaining voltage (Ed) being independent of current (JA) over three to four orders of magnitude, in agreement with the ideas expressed by (17.6.1). Note also that the E/N depends quite critically on the mixture involved, with heavy helium concentrations requiring the lowest E/N. Thus, by (17.4.13d), we should expect the lowest average energy ϵ_k of the electron gas. This, as we shall see, results in the highest efficiency of excitation of the upper laser level.

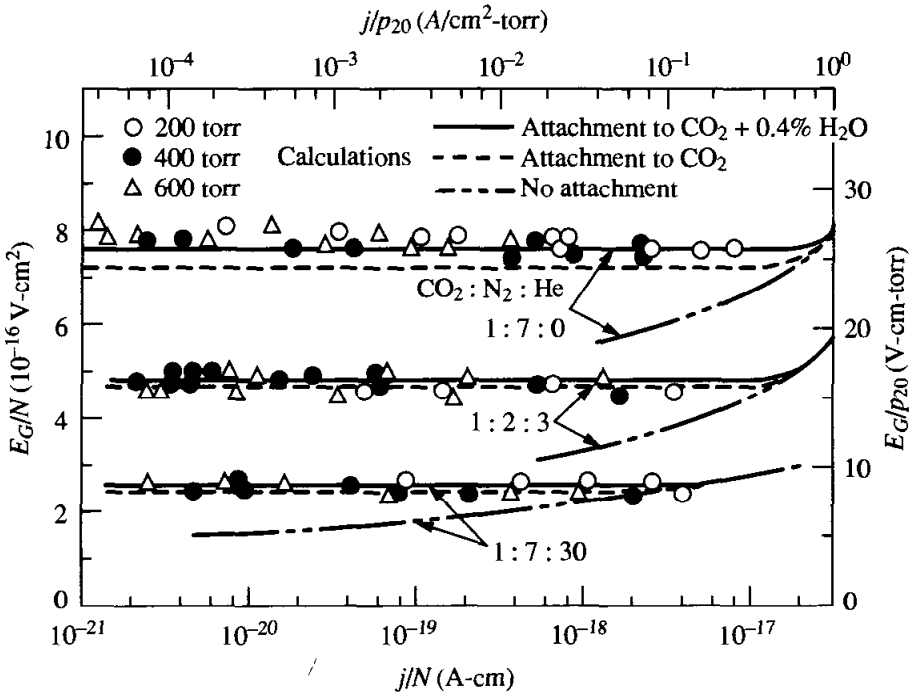


FIGURE 17.8. Measured and calculated V–I characteristics for three $CO_2/N_2/He$ laser mixtures. Experimental values at three values of pressure are indicated by data points. Curves represent results of numerical calculations with varying conditions of attachment. (Reprinted from Fig. 2 of Denes and Lowke [2] with permission of L. J. Denes and J. J. Lowke and the American Physical Society.)

These are the experimental facts of life. The purpose of the following sections is to obtain a correlation between the theory and experiment.

17.6.3 Theoretical Calculations

As has been implied many times throughout this chapter, gas-discharge *calculations* are extremely messy and time consuming, even though the theory of the calculations is straightforward. In all but contrived problems in textbooks, we must resort to the computer to obtain realistic results in a finite length of time. Fortunately, Lowke et al. [3] have published the results of their calculations, which are directly applicable to the experiment described. The solid curves in Fig. 17.8 are one result of these calculations.

Our purpose here is to show how one arrives at these predictions and, as a bonus, to determine what fraction of the electrical power is used to excite the laser.

We will examine a specific case:

1. Total gas pressure = 400 torr.
2. Mixture of 1:2:3 parts of CO_2:N_2:He
3. Concentration of various constituent gases
   - **a.** $[CO_2] = 2.36 \times 10^{18}\ \text{cm}^{-3}$
   - **b.** $[N_2] = 4.72 \times 10^{18}\ \text{cm}^{-3}$
   - **c.** $[\text{He}] = 7.08 \times 10^{18}\ \text{cm}^{-3}$
   - **d.** $[N_T] = \sum = 1.42 \times 10^{19}\ \text{cm}^{-3}$

It is important to realize that a considerable body of knowledge must be accumulated about the electron collision processes before one can even begin. For instance, if a discharge is to exist, then some sort of ionization collision must take place. That collision may be with CO_2, N_2, or helium, and to compute that rate, we must multiply the rate $\sigma_i v_e$ by the distribution function and sum over all velocities in accordance with (17.4.21). That procedure sounds simple enough except for the fact that the distribution function of the electrons is also unknown at this stage.

Thus the first step is to compute a distribution function $F(\epsilon)$, which requires the knowledge of all important inelastic and elastic collision processes. To appreciate why this knowledge is necessary, let us reconsider Fig. 17.4. The electrons move up the energy scale because of the presence of the electric field at a more or less smooth monotonic rate but take large discontinuous steps downward because of the inelastic collisions. Thus, if an inelastic collision is very probable within a given energy interval, there is a rapid depletion of electrons out of that range, and the equilibrium population of free electrons with those energies will be small. For instance, there are two very important inelastic processes that are vitally important to the excitation of the upper CO_2 laser level:

$$e + CO_2(000) \rightarrow CO_2(001) \tag{17.6.1a}$$

$$e + N_2(v = 0) \rightarrow N_2(v = n), \qquad n = 1 \text{ to } 8$$
$$\hookrightarrow + CO_2(000) \rightarrow CO_2(001) \tag{17.6.1b}$$

Lowke et al. [3] account for 31 distinct inelastic processes in the computation of $F(\epsilon)$. The details of that computation are beyond the scope of this book, but the results are very pertinent and can be used with minimal effort on our part.

Typical distribution functions for various values of E/N for the 1:2:3 mixture are shown in Fig. 17.9.* As we would expect, there are more electrons at higher energy for higher values of E/N. The curve labeled $E/N = 3 \times 10^{-16}$ (V-cm^2) illustrates a very important point made earlier, that the distribution function is *not* a Maxwellian. Indeed, the dashed curve is a Maxwellian, with the same characteristic energy ϵ_k or "electron temperature."

There is a simple reason for the fact that the actual distribution has a deficiency of electrons in the 2- to 5-eV range. This is because of the very large resonant cross section for vibrational pumping of N_2 (17.6.1b). Fig. 10.32 illustrates that cross section.

Once the distribution function is determined, we can compute the electrical transport quantities in the straightforward but tedious manner indicated by (17.4.25). Obviously, this has to be done numerically, and the typical results are plotted in Fig. 17.10, which gives the drift velocity w as a function of the parameter E/N. (Note that if we blindly trusted the simple average electron approach, the curves would be a straight line. The fact that they are

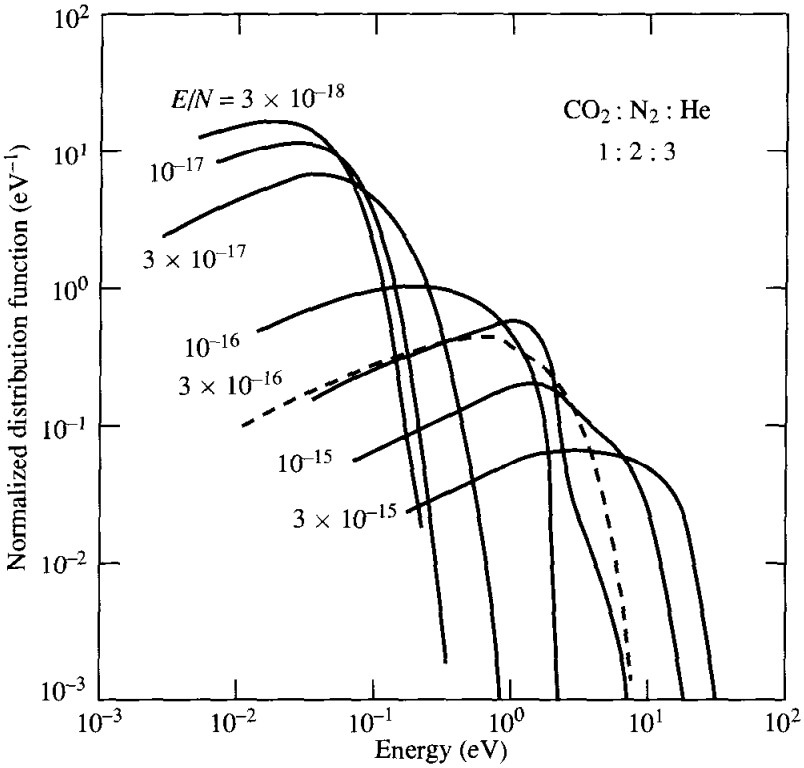


FIGURE 17.9. Derived distribution functions for CO_2:N_2:He = 1 : 2 : 3 for various values of E/N (V-cm^2). (Reprinted from Fig. 4 of Lowke et al. [3] with permission of J. J. Lowke, A. V. Phelps, and B. W. Irwin and the American Physical Society.)

*The distribution function is $F(u) = u^{1/2} f(u)$, with u being the electron energy (in eV) and normalized such that $\int_0^\infty u^{1/2} f(u)\, du = 1$.

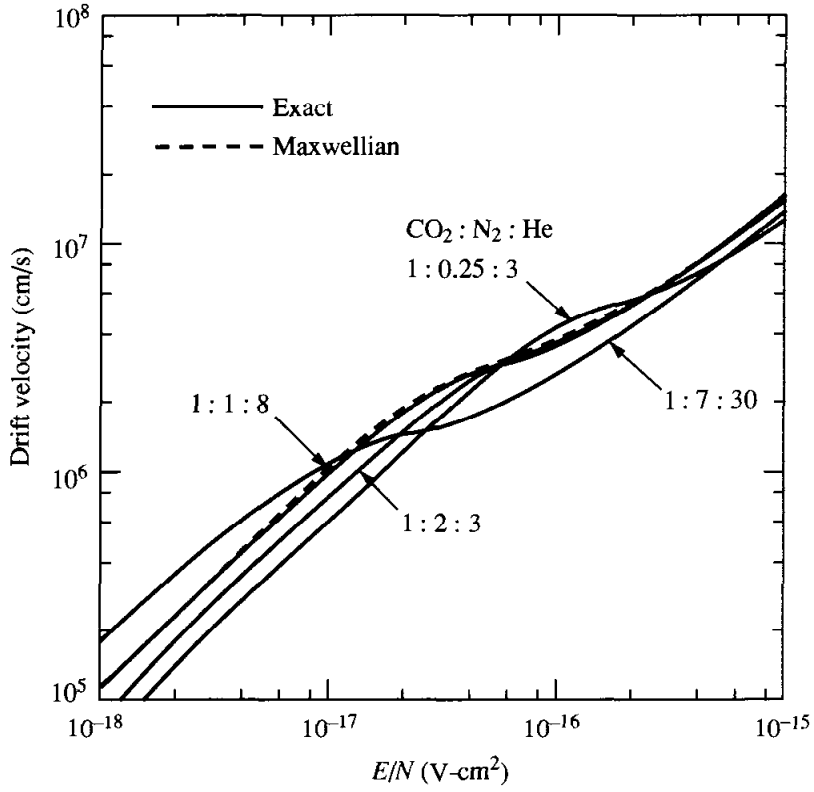


FIGURE 17.10. Calculated drift velocities of electrons for various gas mixtures of CO_2, N_2, and He. (Reprinted from Fig. 6 of Lowke et al. (Ref. 3) with permission of J. J. Lowke, A. V. Phelps, and B. W. Irwin and the American Physical Society.)

not is due to the presence of the large inelastic cross sections mentioned earlier and their influence on the distribution function.)

In fact, one should approach such a graph with a proper appreciation of prior history. The drift velocity is a measurable quantity (see, e.g., Huxley and Crompton [4]), and it is from such measurements that the cross sections used in this calculation were inferred. Thus we should treat Fig. 17.10 as more than a mere calculation, almost as an experimental fact.

One also obtains the transverse diffusion coefficient in such measurements, and thus the ratio of it to the electron mobility, $w_d/E = \mu$, can be considered as an experimental number to be accorded the same respect as the drift velocity.

This ratio is plotted in Fig. 17.11. The importance of this quantity can be appreciated if one recalls the Einstein relation

$$\frac{D_T}{\mu} = \frac{kT_e}{e} \qquad \text{(for a Maxwellian)} \tag{17.6.2a}$$

a result valid only for a Maxwellian distribution. The fact that the true distribution is not Maxwellian implies that the words "electron temperature" are meaningless. However, the ratio D_T/μ still has the dimensions of energy (in eV) and is a general prescription for the characteristic energy of the electron gas.

$$\epsilon_k = \frac{D_T}{\mu} \qquad \text{(definition of "characteristic energy")} \tag{17.6.2b}$$

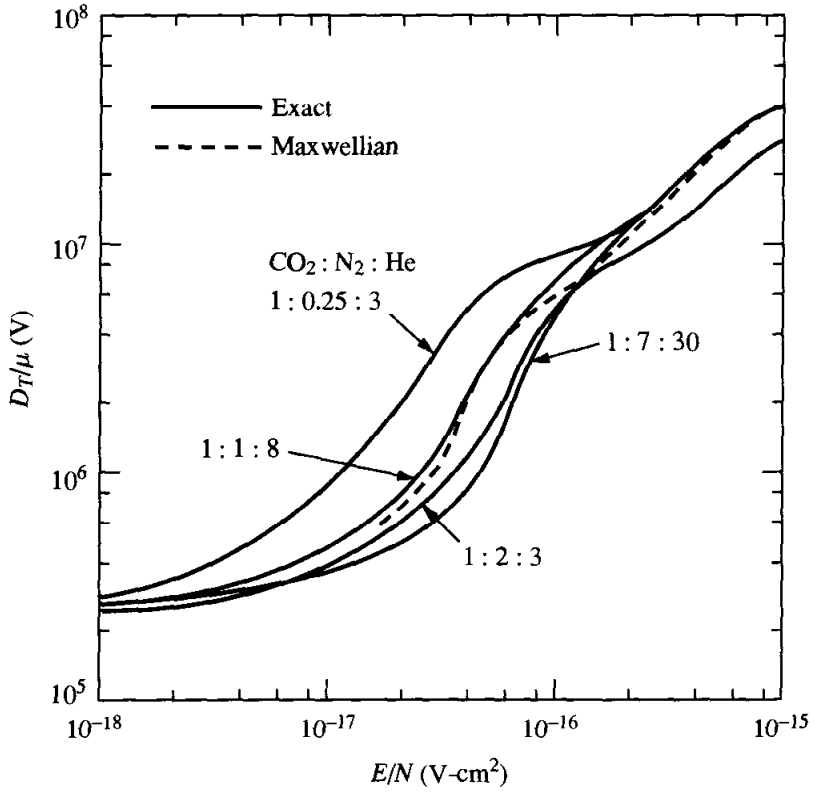


FIGURE 17.11. Calculated values of D_T/μ for various gas mixtures of CO_2, N_2, and He. (Reprinted from Fig. 7 of Lowke et al. [3] with permission of J. J. Lowke, A. V. Phelps, and B. W. Irwin and the American Physical Society.)

17.6.4 Correlation Between Experiment and Theory

This subsection shows how the theoretical calculations of Sec. 17.6.3 can be correlated with the experimental results given in Sec. 17.6.2.

The fact that a discharge exists in quasi-CW fashion means that there must be an exact balance between the electron production and loss.

$$e + M \rightarrow M^+ + 2e \qquad \text{(production)} \qquad (17.6.3a)$$

$$e + M' \rightarrow M'' + 0 \cdot e \qquad \text{(loss)} \qquad (17.6.3b)$$

Whereas the ionization process indicated by (17.6.3a) is somewhat obvious for a production mechanism, the loss, (17.6.3b), encompasses a multiplicity of possibilities. We could lose the electron by attachment (creating a sluggish negative ion), the electron could recombine with a positive ion, or the electron (and ion) could disappear from the system by diffusion.

If we assume that attachment and recombination are the dominant loss process, the electron continuity equation becomes

$$0 = \frac{dn_e}{dt} = n_e \alpha w_d - n_e a w_d - \gamma n_e N_i + S_{\text{ext}} \qquad (17.6.4)$$

In (17.6.4), αw_d is the ionization rate associated with (17.6.3a); $a w_d$ is the attachment rate per electron, implied by (17.6.3b); γ is the electron-ion recombination coefficient; and S_{ext} is an ionization process controlled by devices external to the discharge (more about this later).*

*The term α is called the first Townsend ionization coefficient.

Since we are dealing with a plasma, with its requirement of near-space-charge neutrality, we can set $N_i = n_e$ and abbreviate the third term of (17.6.4) as γn_e^2. Let us restrict our attention for the moment to the case of low electron densities so that γn_e^2 can be neglected and also assume that $S_{\text{ext}} = 0$. Thus (17.6.4) reduces to a trivial-appearing equation:

$$\alpha = a \qquad \text{or} \qquad \frac{\alpha}{N} = \frac{a}{N} \tag{17.6.5}$$

The problem hidden in this simple equation is that both α and a are functions of E/N, and thus an iterative procedure must be followed to obtain a solution.

The information necessary to obtain such a solution is contained in Fig. 17.12, which was calculated by those authors from the computed distribution function and the published cross sections for the various processes. The procedure for obtaining the E/N value for a discharge in a 1:2:3 mixture is illustrated in Table 17.3. A first guess for E/N is 3×10^{-16} V-cm^2; at this value $\alpha/N = 1.42 \times 10^{-22}$ cm^2 and $a/N = 4 \times 10^{-21}$ cm^2 from Fig. 17.12. Obviously, the loss is greater than the production, and hence we need a larger E/N. Table 17.3 illustrates how fast we can converge to a solution. Having obtained our first

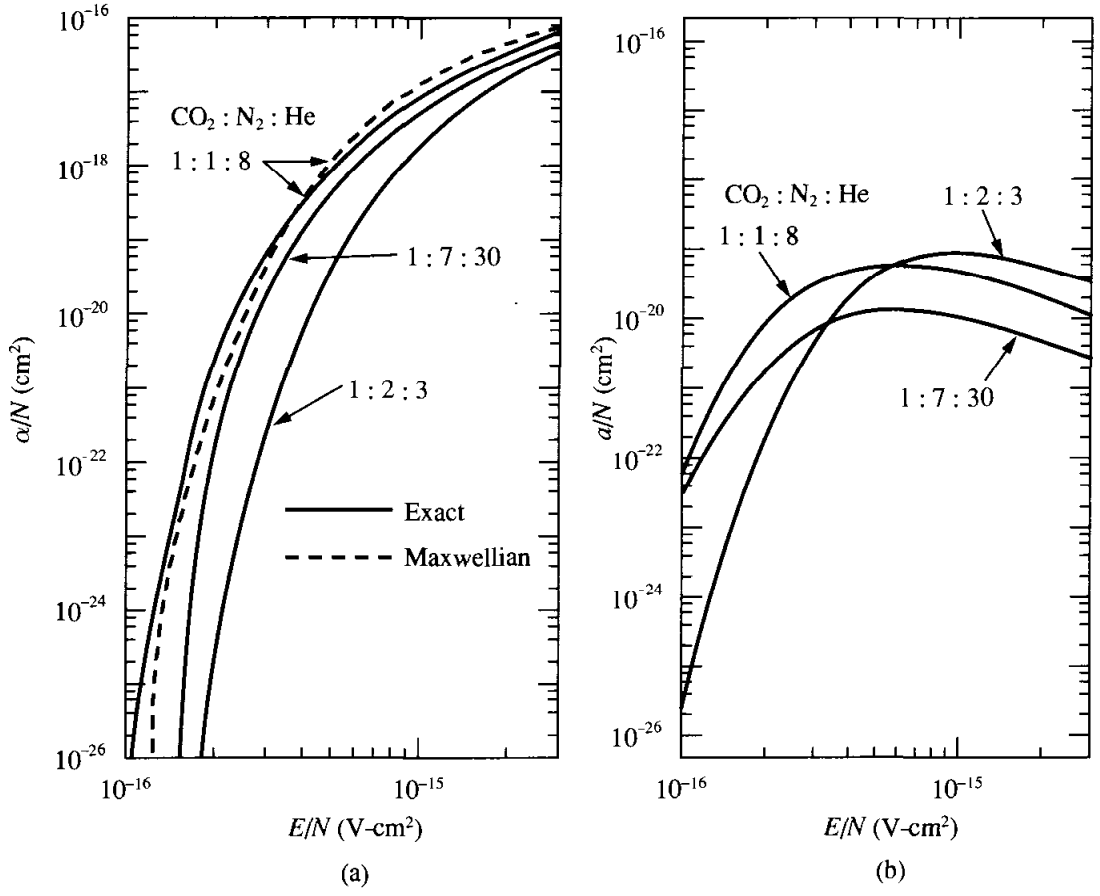


FIGURE 17.12. Calculated values of α/N and a/N for various gas mixtures of CO_2, N_2, and He; α is the ionization coefficient (cm^{-1}) and a is the attachment rate. The dashed curve indicates values obtained assuming a Maxwellian distribution of the 1:1:8 mixture. (Reprinted from Figs. 9 and 10 of Lowke et al. [3] with the permission of J. J. Lowke, A. V. Phelps, and B. W. Irwin and the American Physical Society.)

TABLE 17.3 Solution for E/N in a 1:2:3 Mixture

E/N (V-cm^2)	α/N (cm^2)	a/N (cm^2)	Comments
3×10^{-16}	1.42×10^{-22}	4×10^{-21}	Loss $\gg$ production
5×10^{-16}	3.6×10^{-20}	2.77×10^{-20}	Production > loss
4×10^{-16}	4×10^{-21}	1.7×10^{-20}	Loss > production
4.8×10^{-16}	2.66×10^{-20}	3.05×10^{-20}	Close
4.9×10^{-16}	3.1×10^{-20}	3.1×10^{-20}	Solution

theoretical solution, it is time to compare it to a measurement. If we refer back to Fig. 17.7, where E/N is plotted for this mixture, excellent agreement between theory (solid curves) and experiment (triangles, dots, and circles) is demonstrated.

Now let us examine the implications of the neglect of recombination in (17.6.4). If we identify $J = n_e e w_d$ (as usual) and neglect S_{ext} again, this equation can be cast into the following format:

$$\frac{J}{N}\left(\frac{\alpha}{N} - \frac{a}{N}\right) = \frac{\gamma}{(ew_d)^2}\left(\frac{J}{N}\right)^2 \tag{17.6.6}$$

Thus the solution given by Table 17.3 assumes that the right side of (17.6.6) is much smaller than either term on the left. Obviously, if the current is high enough, this is no longer valid, and (17.6.6) requires that $\alpha/N > a/N$. This is the reason that the E/N required for a self-sustaining discharge increases at the high current points of Fig. 17.7 (γ was assumed to be 10^{-7} cm^3/s for those curves.)

Before we proceed to the laser performance, it is appropriate to translate some of these numbers to a more familiar and practical format. For a pressure of 400 torr, we have $N = 1.42 \times 10^{19}$ cm^{-3}. Hence

$$E = \left(\frac{E}{N}\right) N = 4.9 \times 10^{-16} \times 1.42 \times 10^{19} = 6.96 \text{ kV/cm} \tag{17.6.7a}$$

and the terminal voltage across the electrodes is

$$V = Ed = 6.96 \times 10^3 \text{ V/cm} \times 1.2 \text{ cm} = 8.35 \text{ kV} \tag{17.6.7b}$$

For $J/N \ll 10^{-17}$ A-cm, the right side of (17.6.6) can be neglected, and thus V is independent of I.

$$\left.\begin{aligned} J &< 10^{-17} \text{ A-cm} \cdot (N = 1.42 \times 10^{19} \text{ cm}^{-3}) \\ \text{or} \\ J &< 142 \text{ A/cm}^2 \\ \text{Therefore} \\ I &< 4.03 \text{ kA} \end{aligned}\right\} \tag{17.6.7c}$$

Obviously, it is not a simple task to switch such high currents and handle such high voltages.

Let us arbitrarily pick a current density far below this value so as to have confidence in our prior solution, say, $J = 1$ A/cm^2. Then

$$I = 29 \text{ A} \tag{17.6.7d}$$

$$V = 8.35 \text{ kV} \tag{17.6.7e}$$

and the power delivered to this $1.2 \times 10 \times 2.9$ cm^3 volume is

$$P = 242 \text{ kW} \tag{17.6.7f}$$

or

$$\frac{P}{V} = 6.96 \text{ kW/cm}^3 \tag{17.6.7g}$$

No wonder the laser has to be pulsed.

It is also of interest to compute the plasma parameters, such as the degree of ionization and the characteristic energy of the electrons:

$$n_e = \frac{J}{ew_d} \tag{17.6.8a}$$

We use Fig. 17.9 to find $w_d = 8.9 \times 10^6$ cm/ sec at $E/N = 4.9 \times 10^{-16}$ V-cm^2. Thus

$$n_e = \frac{1 \text{ A/cm}^2}{1.6 \times 10^{-19} \text{ c} \times 8.96 \times 10^6 \text{ cm/ sec}}$$

$$= 7.02 \times 10^{11} \text{ e-i pairs/cm}^3 \tag{17.6.8b}$$

Degree of ionization

$$f = \frac{n_e}{N} = \frac{7.02 \times 10^{11} \text{ cm}^{-3}}{1.42 \times 10^{19} \text{ cm}^{-3}} = 4.9 \times 10^{-8} \tag{17.6.8c}$$

To say that this gas is weakly ionized is an understatement of classic proportions!

The characteristic energy of these 7×10^{11} cm^{-3} electrons in response to the voltage of 8.35×10^3 V can be found from Fig. 17.11. For $E/N = 4.9 \times 10^{-16}$ V-cm^2,

$$\frac{D_T}{\mu} = \epsilon_k = 1.6 \text{ eV}$$

Thus, in spite of the high voltage applied, the characteristic energy is quite low. As we shall see, we would prefer it to be lower still.

17.6.5 Laser-Level Excitation

The preceding section showed that we can predict the electrical characteristics quite accurately. Now we ask the laser question: How much of the 242 kW (17.6.7f) could be extracted as optical power at 10.6 μm? In general, this is a most complex question. But it surely can never be more than that fraction used to excite the upper state: The purpose of this section is to compute that limit.

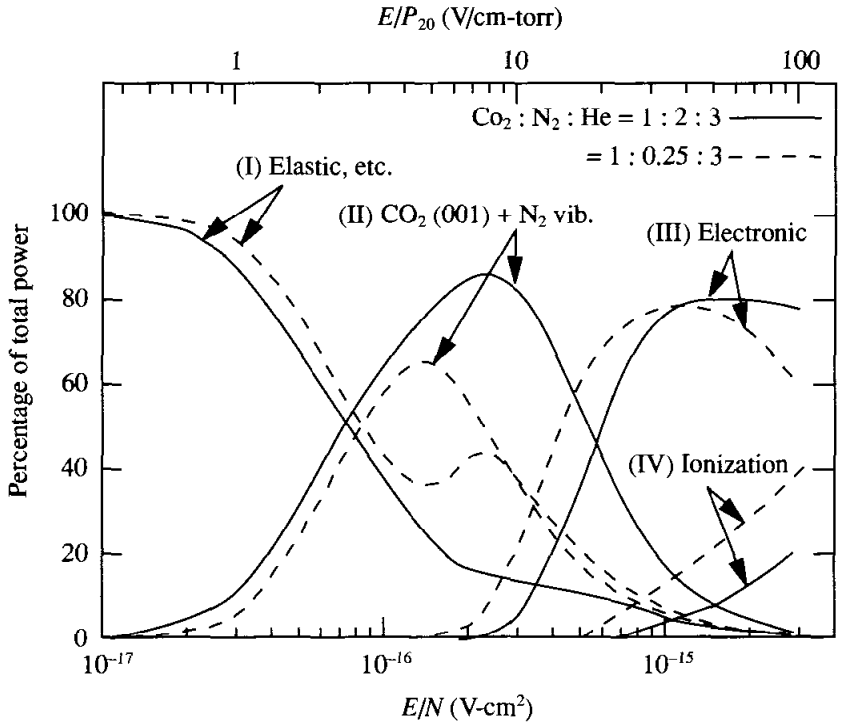


FIGURE 17.13. Percentage of power lost to (I) elastic collisions, rotational excitation of N_2, and excitation of bend and stretch modes of CO_2; (II) CO_2 (001) level and the first eight vibrational levels of N_2; (III) electronic excitation; and (IV) ionization. Increasing the ratio of N_2 to CO_2 increases the predicted efficiency given by (II). (Reprinted from Fig. 12 of Lowke et al. [3] with the permission of J. J. Lowke, A. V. Phelps, and B. W. Irwin and the American Physical Society.)

To compute this quantity, return to the distribution function at the E/N value for the discharge and account for the uses of the energy transferred from the electrons to the neutral atoms. Lowke et al. [3] did this accounting, and the results are reproduced in Fig. 17.13.

They kept four separate accounts, which are plotted in Fig. 17.13. Category I represents the fraction of total power used for not-too-interesting and most probably deleterious purposes, such as gas heating through elastic collisions, exciting the lower laser level, and rotational excitation. Category II is the answer to the question of this subsection. This is the fraction used to excite those quantum states that eventually feed the 001 level of CO_2. Thus the energy lost in creating a vibrationally excited N_2 molecule is included here because it can transfer its energy to CO_2 very efficiently.

Category III is the power used to excite the electronic levels of the molecules. That fraction is useless for the CO_2 laser but generally harmless otherwise. Category IV is the fraction used in ionization and thus in keeping the discharge alive. As the curves show, this fraction is negligible except at high E/N values. Even though minuscule and ignorable insofar as a power balance is concerned, it is absolutely essential for the *existence* of the discharge.

For the value of $E/N = 4.9 \times 10^{-16}$ V-cm^2, Fig. 17.13 predicts the following ratios:

(I) Elastic, gas heating, etc. 14.3%

(II) Upper laser level	54.3%	(17.6.9)
(III) Electronic excitation	31.4%	
(IV) Ionization	Negligible but positive	

Thus if we use 242-kW electric power (17.6.7f) to excite this gas mixture, 54.3% is used to create the upper state. Since the quantum efficiency of this laser is 39.9%, one could hope for an optical generation rate of

$$P_{\text{opt}} = 242\ \text{kW} \times 0.543 \times 0.399 = 52.4\ \text{kW} \tag{17.6.10}$$

or a "wall-plug" efficiency of 21.7%.

Obviously, this is the best one could expect with optimum extraction and perfect excited-state kinetics. It is most impressive that such numbers have been approached.

There are also practical points to be gained from these numbers:

1. Although the wasted fraction 14.3% is small, the power is large: $0.143 \times 242\ \text{kW} = 34\ \text{kW}$. Thus the gas gets hot in a hurry.
2. Even if only 1% of the 52.4 kW of optical power were absorbed in the mirrors, they, too, will become hot and probably be damaged.
3. The power supply to drive this laser is not a simple device.

Again, we conclude, no wonder such lasers are pulsed.

17.7 ELECTRON BEAM SUSTAINED OPERATION

As impressive as the performance figures of the preceding section are, we see from Fig. 17.13 that we can do better. For instance, if E/N were 2.27 V-cm^2, then nearly 85% of the electrical power would go into exciting the upper laser level. Unfortunately, E/N of a *self-sustained discharge* is *not* 2.27×10^{-16} V-cm^2 but is 4.9×10^{-16} V-cm^2, as computed previously. Thus this 85% is a pipe dream unless we do things differently.

This is the role of the external source of ionization, S_{ext}, which until now has been so conveniently ignored. Its function is to create the ionization so that the applied field can transfer energy to the electrons at the optimum rate. We try to have as many electrons as possible in the proper energy interval to excite the upper laser level. This external ionization source can be an x-ray machine, a nuclear reactor, or an energetic electron beam. The latter is the most efficient, convenient, and practical and is therefore the most common method of providing this ionization. Electrons are accelerated to very high voltages—say, 100 kV—so that they can pass straight through the foil used to separate the high-pressure laser discharge cavity from the vacuum required to sustain the high voltage.

Once through the foil, the energetic electrons slow down by making many ionizing and excitation types of collisions. For instance, it takes roughly twice the ionization energy to create an electron-ion pair. But even the maximum ionization potential of any gas—that of helium is 24.5 eV—is insignificant compared to the energy of the electron beam. Thus

each beam electron is capable of producing an enormous number of secondaries, and it does so. This additional controlled source is, in effect, a substitution for the internal ionization one. Hence, the required E/N to effect a balance between production and loss is lowered.

This means that we have to rework the entire problem from the continuity equation onward. We start with the prescribed value of $E/N = 2.27 \times 10^{-16}$ V-cm^2 for optimum laser excitation and find from Fig. 17.12 that the internal production factor is $\alpha/N = 2 \times 10^{-24}$ cm^2, whereas the loss factor is $a/N = 10^{-21}$ cm^2. Thus we now neglect the first term in the continuity equation, (17.6.4), and cast it into the format involving current.

$$\left(\frac{a}{N}\right)\frac{J}{eN} + \frac{\gamma}{(ew_d)^2}\left(\frac{J}{N}\right)^2 = \frac{S_{\text{ext}}}{N^2} \tag{17.6.11}$$

If we demand the same optical power out as before, 52.4 kW, then the amount used to create the upper state is P_{opt} divided by the quantum efficiency, or

$$P_{\text{category II}} = \frac{52.4}{0.399} = 131 \text{ kW} \tag{17.6.12a}$$

This is 85% of the electrical power, or

$$P_{\text{elect}} = \frac{131 \text{ kW}}{0.85} = 154 \text{ kW} \tag{17.6.12b}$$

The voltage across the device is prescribed by the desired E/N.

$$V = \left(\frac{E}{N}\right)Nd$$

$$= 2.27 \times 10^{-16} \times 1.42 \times 10^{19} \times 1.2 = 3.86 \text{ kV} \tag{17.6.12c}$$

To obtain the 154 kW at this voltage means that the electrical currents must increase from the previous case (17.6.7d):

$$I = \frac{P_{\text{elect}}}{V} = \frac{154 \text{ kW}}{3.86 \text{ kV}} = 39.9 \text{ A} \tag{17.6.12d}$$

This translates to a current density of 1.38 A/cm^2 or $J/N = 9.72 \times 10^{-20}$ A-cm. Thus the drift velocity is found from Fig. 17.9. The plasma parameters are

$$w_d = 5.46 \times 10^6 \text{ cm/ sec} \tag{17.6.13}$$

and hence the electron density is given by

$$n_e = \frac{J}{ew_d} = \frac{1.38}{1.6 \times 10^{-19} \times 5.46 \times 10^6}$$

$$= 1.58 \times 10^{12} \text{ e-i pairs/cm}^3 \tag{17.6.14a}$$

This has increased by a factor of 2 over that computed previously in (17.6.8). The characteristic energy is much reduced.

$$\epsilon_k = \frac{D_T}{\mu} \quad \text{(see Fig. 17.11)} = 0.91 \text{ eV} \tag{17.6.14b}$$

All parts of (17.6.12) are the desired result. Now, what must be the strength of this external source of ionization to obtain this result? To answer this, we return to (12.6.11) and use the preceding numbers to compute S_{ext}. For a typical value of $\gamma = 10^{-7}$ cm^3/ sec,

$$S_{ext} = N^2 \left[\left(\frac{a}{N} \right) \frac{J}{N} \frac{1}{e} + \frac{\gamma}{(ew_d)^2} \left(\frac{J}{N} \right)^2 \right] \tag{17.6.11}$$

$$= 3.72 \times 10^{17} \text{ e-i pairs/cm}^3/\text{ sec} \tag{17.6.15}$$

If one is to use an energetic beam—say, 100 kV—to create this ionization, the S_{ext} can be related to the beam current density by

$$S_{ext} = n_b v_b \sum_j N_j \sigma_j = \frac{J_{beam}}{e} \sum_j N_j \sigma_j \tag{17.6.16}$$

$$j = CO_2, N_2, \text{He, the type of gas}$$

where J_{beam} is the injected current density of the high-energy beam, v_b the beam velocity $= (2eV_b/m)^{1/2}$, V_b the beam voltage, N_j the density of the jth type of gas, and σ_j the ionization cross section of that gas at the beam energy (presumed to be 100 kV). Table 17.4 gives the pertinent data for CO_2, N_2, and He.

Equating (17.6.16) to (17.6.15) and using Table 17.4 with the appropriate densities of the gases, we find that

$$J_b = 3 \times 10^{-3} \text{ A/cm}^2 \tag{17.6.17a}$$

or

$$I_b = 88 \text{ mA} \tag{17.6.17b}$$

The beam power is

$$P_{beam} = V_b I_b = 8.8 \text{ kW} \tag{17.6.18}$$

Thus our optimized laser wall plug efficiency is

$$\eta = \frac{P_{opt}}{P_{discharge} + P_{beam}} = \frac{52.4 \text{ kW}}{154 \text{ kW} + 8.8 \text{ kW}} = 32.2\% \tag{17.6.19}$$

TABLE 17.4 Data for Beam Ionization by 100-kV Electrons

Gas	Ionization Potential (eV)	Estimated Cross Section (cm^2)*
CO_2	13.769	3.2×10^{-18}
N_2	15.576	1.9×10^{-18}
He	24.586	4.2×10^{-19}

*Data extrapolated from Kieffer [1].

The examples of these two sections illustrate the potential power of high-pressure gas lasers. The CO_2 system is not the only gas laser capable of producing these extreme power levels; however, it does prove the main points of this chapter:

1. A gas discharge laser is a very simple device, two electrodes with a rather poor-quality vacuum between them. A precise description requires an enormous investment in prior experimental work and definite familiarity with computational techniques.
2. However, the ideas involved are very simple. It is hoped that this chapter will inspire you to examine texts devoted exclusively to gas-discharge theory.
3. If nothing else, you should realize that a gas laser is capable of producing enormous powers—in some cases at respectable efficiencies.

PROBLEMS

17.1. Consider a 1-cm-radius discharge tube filled uniformly with an e-i pair density of 10^{13} cm^{-3}. If all the electrons disappeared and thus uncovered the positive ionic charge, what would be the potential of the wall with respect to the center? (Ans.: 4.52 MV. Obviously this means that the electrons have a vanishingly small probability of disappearing at a different rate than that of the ions.)

17.2. Show that the maximum electrical current density that can be passed by a diode without space-charge neutralization is given by

$$J = \frac{4}{9}\epsilon_0 \left(\frac{2e}{m}\right)^{1/2} \frac{V^{3/2}}{d^2}$$

where the anode-to-cathode spacing is d and the potential difference is V (the Child–Langmuir law).

17.3. Suppose that an electron collides "elastically" with a stationary heavy neutral of mass M. Show that if the scattering is isotropic, the electron loses, on the average, a fraction $2m/M$ of its energy per collision.

17.4. Suppose that one suddenly removed the voltage across the lamp of Table 17.2. Estimate how long it will take the electron energy to decay from 1.25 eV to that of the neutrals (0.026 eV). (Assume elastic losses only.)

17.5. The current represented by the *random* motion of the electrons is huge compared to the circuit current. For instance, consider the parameters of a fluorescent lamp in common use: $I = 0.25$ A, $D = 1$ in.; $N_e = N_i = 10^{12}$ cm^{-3}. Assume that the electron distribution is Maxwellian at $kT/e = 1.4$ eV and show that the random current is given by

$$J_r = \frac{N_e e}{4}\left(\frac{8kTe}{\pi m}\right)^{1/2}$$

Evaluate for the conditions given.

17.6. Suppose that the spherical symmetric part of the distribution function is given by

$$f_0(v) = A/[1 + (v/v_0)^N]$$

where v_0 is some characteristic *speed* and A the value to be determined from the normalization condition. (a) Evaluate A and (b) plot $F(\epsilon)$ as a function of energy.

17.7. Show that if ν_c is independent of velocity and f_0 is Maxwellian, (17.4.7) is the correct value for the conductivity.

17.8. Evaluate the percentage error involved in using (17.4.3) for the drift velocity by evaluating (17.4.25b) for the following dependence of ν_c on velocity.

$$\nu_c(v) = \nu_0 \left(\frac{v}{v_0} \right)^m \qquad m = -3, \cdots, +3$$

17.9. Can you give a simple explanation of why elastic collisions are always much more frequent than inelastic ones?

17.10. Repeat the reasoning discussed in Sec. 17.4.5 for the case where the plasma density is nonuniform. Find an expression analogous to (17.4.25b) for the flux of electrons in the direction of the spatial gradient. That coefficient is called the diffusion coefficient.

17.11. Suppose that the power supply voltage in Fig. 17.1 has an open circuit voltage of 300 V in series with a ballast resistor of 20 kΩ. What is the operating current and lamp voltage for (a) a hot cathode and (b) a cold cathode?

17.12. Use the data of Figs. 17.10 and 17.11 to plot the ratio of the drift energy to the characteristic energy.

17.13. Plot the predicted operating E/N versus electron-beam current for the case discussed in Fig. 17.7. Vary the beam current from 0 to 0.1 A/ cm^2.

Problems on the excitation of the CO_2 laser. (Some of these questions *require* that you read the original articles. It is hoped you will do this in *any* case.)

17.14. Suppose that the ballast resistor in Fig. 17.6 was 100 Ω and the capacitor was charged to 20 kV. Ignore the initial phases of the discharge described in the text taking place for times less than 50 ns. Use the case discussed in the text of a 1:2:3 mixture at 400 torr.

(a) Sketch the time variation of the current through the discharge. (Assume that the triggered spark gap opens when the current drops below 20 A, and neglect the voltage drop across S.) What is the duration of the pulse? (Ans.: $\tau = 9.12\,\mu$s.)

(b) **(1)** How much energy is stored in the capacitor at $t = 0$? (Ans.: 10.4 joules.)

(2) Estimate how much of this energy is transferred to the gas. (Ans.: 4.19 joules.)

(3) How much energy remains stored in the capacitor? (Ans.: 2.79 joules.)

(4) How much energy is lost in the 100-Ω ballast resistor? (Ans.: 3.42 joules.)

(5) What is the peak power dissipated in the ballast? (Ans.: 1.36 MW.)

(c) How much of the energy delivered to the discharge is used

(1) To excite the N_2 vibrational levels and the 001 state of CO_2? (Ans.: 2.28 joules.)

(2) To heat the gas via elastic collision and excitation of the lower laser level? (Ans.: 0.599 joules.)

(d) If we assume that all the energy of (c, 2) goes into random translational motion of the gas molecules, what is the temperature rise at the end of the current pulse? (Ans.: $\Delta T = 58.6$ K)

(e) Assume that an equilibrium is established between the 001 state of CO_2 and the $v = 1$ state of N_2 and the energy of $(c, 1)$ resides in these states. Assume a perfect match in energy between the two states.
  (1) How much is in the CO_2 system? (Ans.: 0.76 joules.)
  (2) How much is in the N_2 system? (Ans.: 1.52 joules.)
  (3) What percentage of the CO_2 and N_2 molecules are in the 001 state or $v = 1$ state? (Ans.: 19.7%.)

(f) The energy stored in the CO_2 system can be extracted at a rate limited only by the photon-buildup time (see Chapter 9 on gain switching), whereas that stored in the N_2 system is limited by the collisional transfer rate between N_2 and CO_2. Assume a gain switched pulse of 3-ns duration and a transfer rate of 10^7 sec^{-1}. Assume further that the lower state decays with a lifetime of 10 ns.
  (1) Estimate the energy in the gain switch pulse. (Ans.: $0.76/2 = 0.38$ joules.)
  (2) Estimate the energy in the long tail representing the transfer from N_2 to CO_2. (Ans.: 1.52 + 0.38 joules = 1.90 joules.)
  (3) Sketch the output of this laser. (Ans.: A peak of 0.127 MW followed by a long tail [more or less triangular] with a slope of 10^7 sec^{-1} containing the 1.9 joules of [f(2)].)

17.15. Suppose the electron gas of a fluorescent lamp was described by a Maxwellian distribution function with a characteristic energy $kT_e/e = 1.5$ eV whereas the Hg^+ has a temperature of 300 K. Any insulated object in contact with this plasma (i.e., a glass wall or a floating metallic probe) must attain a negative potential with respect to the bulk in order that equal numbers of positive and negative charges arrive at the surface. Compute this "floating potential" difference.

17.16. Assume that the spherically symmetric part of the distribution function of the electron gas can be approximated by

$$f_0(v) = A\left[1 - \frac{v}{v_0}\right] \qquad \text{for} \quad v < v_0 \qquad 0 \text{ otherwise}$$

(a) Evaluate the parameter A from the normalization requirement.
(b) Relate the parameter v_0 to the average energy of the electron gas.

Now assume there is an electric field E that gives the electrons energy, and they in turn transfer their excess energy to the neutrals (density of 10^{17} cm^{-3}) by three types of collisions:

Elastic: $\sigma_c = 10^{-16}$ cm^2; $G = 10^{-4}$

Excitation: $\sigma_x = 10^{-17}$ cm^2; threshold energy of 5 eV = energy lost per collision

Ionization: $\sigma_i = 10^{-18}$ cm^2; threshold energy of 10 eV = energy lost per collision

(c) Use (17.4.25) and the definition of current, $J = new_d$ to compute the drift velocity **w** as a function of E/N. Plot these results in the manner of Fig. 17.10.

(d) Use the energy balance equation to relate the (D_T/μ) of the electrons to E/N. Plot the results in the same manner as was used for Fig. 17.11.

(e) Compute the fractional power used for the various processes as a function of E/N and plot the results in the manner used for Fig. 17.13.

REFERENCES AND SUGGESTED READINGS

1. L. J. Kieffer, "A Compilation of Electron Collision Cross Section Data for Modeling Gas Discharge Lasers," JILA Information Center Report No. 13, Sept. 1973.
2. L. J. Denes and J. J. Lowke, "*V–I* Characteristics of Pulsed CO_2 Laser Discharges," *Appl. Phys. Lett. 23*(3), 130–132, Aug. 1973.
3. J. J. Lowke, A. V. Phelps, and B. W. Irwin, "Predicted Electron Transport Coefficients and Operating Characteristics of CO_2-N_2-He Laser Mixtures," *J. Appl. Phys. 44*(10), 4664–4471, 1973.
4. L. G. H. Huxley and R. W. Crompton, *The Diffusion and Drift of Electrons in Gases* (New York: John Wiley & Sons, 1974).
5. W. L. Nighan, "Electron Energy Distribution and Collision Rates in Electrically Excited N_2, CO, and CO_2," *Phys. Rev. A2*, 1989–2000, 1970.

GENERAL GAS DISCHARGE REFERENCES

John F. Waymouth, *Electric Discharge Lamps* (Cambridge, Mass.: M.I.T. Press, 1971).

A. von Engel, *Ionized Gas*, 2nd ed. (London: Oxford-at-the-Clarendon Press, 1965).

James D. Cobine, *Gaseous Conductors* (New York: Dover Publications, 1958).

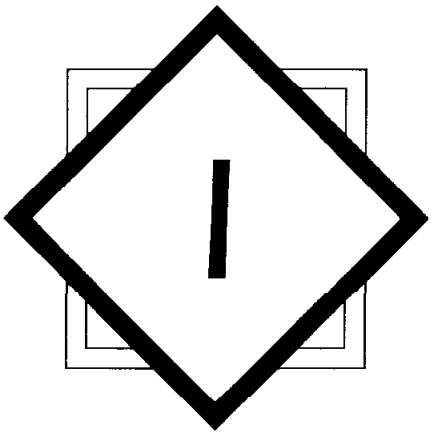
S. C. Brown, *Introduction to Electrical Discharges in Gases* (New York: John Wiley & Sons, 1966).

B. E. Cherrington, *Gaseous Electronics and Gas Lasers* (Oxford: Pergamon, 1979).

G. F. Weston, *Cold Cathode Glow Discharge Tubes* (London: ILIFFE, 1968).

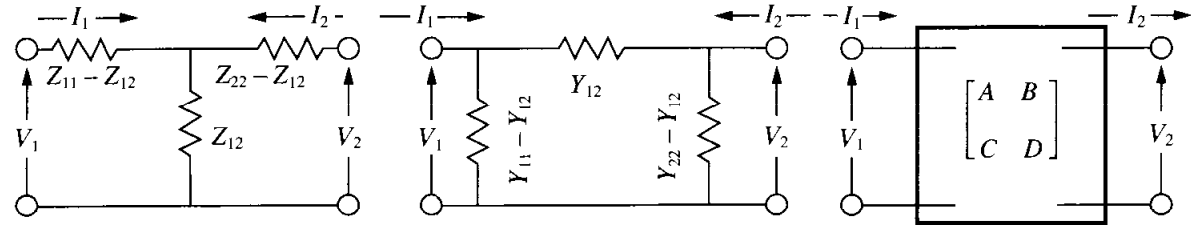
B. M. Smirnov, *Physics of Weakly Ionized Gases*, translated from Russian. (Moscow: Mir, 1981).

B. Chapman, *Glow Discharge Processes* (New York: John Wiley & Sons, 1980).



An Introduction to Scattering Matrices

There are a variety of formalisms used to describe electrical or optical components with a few of the elementary ones indicated in the diagram below with the relationships between the terminals parameters given by the matrices directly below Fig. I.1.



$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \qquad \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \qquad \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_1 \\ I_1 \end{bmatrix} \tag{I.1}$$

FIGURE I.1. Traditional network representation of a two-port network.

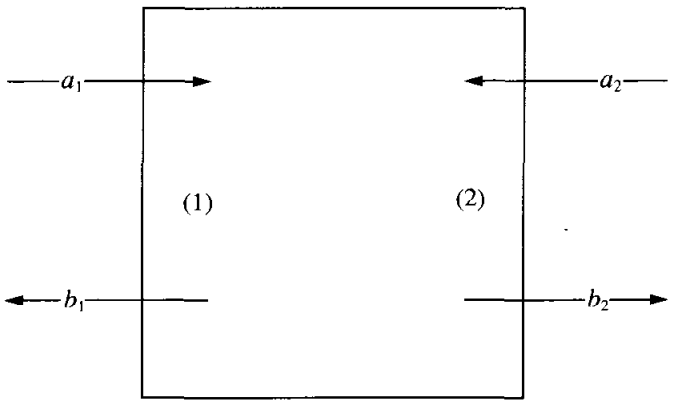
If the network is reciprocal (as the drawing indicates), then $Z_{12} = Z_{21}$, $Y_{12} = Y_{21}$ and $AD - BC = 1$. It is strictly a matter of taste, choice, or experience as to which format to use, because all are perfectly general (or can be made so) and all are interrelated.

One of the most intuitively satisfying formalisms is the "scattering" matrix expressed by the following generic equation:

$$[\text{outgoing waves}] \triangleq [\mathbf{S}]\,[\text{incoming waves}] \tag{I.2}$$

where [S] is the matrix describing that relationship. It is a highly developed and powerful topic in microwave electronics with instruments available to directly measure the elements of [S]. It will handle, with virtually equal ease, a 1-to-n port network, which may or may not be reciprocal. The interested reader can consult the microwave literature for all of the many intricate interrelationships (more or less based upon conservation of energy). Here, let us experience an introduction and two applications of its usefulness in optics.

Consider a two-port system in a situation where there could be two independent but coherent waves incident on the system. We describe these incident waves by the electric fields with a reminder that the magnetic fields are given by $\mathbf{H} = \mathbf{n} \times \mathbf{E}/\eta$ with $\mathbf{n}$ being the direction of propagation ($+z$ for the left incoming wave and $-z$ for the right) and η is the wave impedance appropriate to the two sides. The scattering matrix relates the outgoing waves to the incoming ones. It is traditional and convenient to eliminate all unnecessary constants and abbreviate the incoming waves by [**a**], the outgoing by [**b**], and indicate the port number by a subscript. For the two-port shown



$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{I.3}$$

where we define $a_1 \cdot a_1^*$ as the incoming power (or intensity) and $b_1 \cdot b_1^*$ as the outgoing power (or intensity) at port 1 with corresponding quantities for port 2.

The appeal of this formalism lies in transparent interpretation of the elements of [S] (i.e., S_{ij} represents a complex fraction of the wave incident on the i^{th} port which is "scattered" out of the j^{th} port). For instance, suppose $a_2 = 0$ (i.e., there is no field incident from the right). Then the wave emerging from port 1 is $b_1 = S_{11}a_1$, which immediately assigns an interpretation to S_{11} as being the field reflection coefficient Γ_1 (when there is nothing incident on port 2). The wave emerging out of port 2 is $b_2 = S_{12}a_1$, and thus S_{12} is the field transmission coefficient t_{12} going from 1 to 2. If the incoming wave is from the right, then $S_{22} = \Gamma_2$, the field reflection coefficient as seen from the right, and S_{21} is the field transmission going from 2 to 1, t_{21}, and all with the proviso of conditions of the "thought" or "Gedanken" experiments. These coefficients may be complex, indicating phase shifts relative to a chosen reference plane.

Power conservation indicates some general constraints on the values of the elements. For instance if $a_2 = 0$ (nothing incident from the right) and $a_1 \neq 0$, then $b_2 = S_{12}a_1$ and $b_2 \cdot b_2^* = |S_{12}|^2 a_1 a_1^*$; $b_1 = S_{11}a_1$ and $b_1 \cdot b_1^* = |S_{11}|^2 a_1 a_1^*$.

Since the incident power (or intensity) is $a_1 \cdot a_1^*$, power conservation requires that $|b_1|^2 + |b_2|^2 \leq |a_1|^2$ or

$$\boxed{|S_{12}|^2 + |S_{11}|^2 \leq 1} \tag{I.4a}$$

For $a_1 = 0$, $a_2 \neq 0$, we obtain

$$\boxed{|S_{21}|^2 + |S_{22}|^2 \leq 1} \tag{I.4b}$$

If the scattering element is lossless, then the equality sign applies since power is strictly conserved. If the element is excited from both sides (or ports), then the power emerging from the element must equal the incoming value (for a passive system). Forming the sum $|b_1|^2 + |b_2|^2$ from (I.3) leads to

$$\begin{aligned} |b_1|^2 + |b_2|^2 = (|S_{11}|^2 + |S_{12}|^2)|a_1|^2 + (|S_{22}|^2 + |S_{21}|^2)|a_2|^2 \\ + [S_{11}S_{12}^* + S_{22}^*S_{21}]a_1a_2^* + [S_{11}^*S_{21} + S_{22}S_{21}^*]a_1^*a_2 \end{aligned} \tag{I.5}$$

The left side of (I.5) is the outgoing or scattered power, and since each multiplier of $|a_{1,2}|^2$ is 1 by (I.4), the right side of the top line is the equal incoming power, a fact that forces the second line of (I.5) to be zero. The two terms of the last line are obviously complex conjugates of each other and thus both must be purely imaginary since their sum must be zero. Since the phases of a_1 and a_2 can be chosen arbitrarily (experimentally), we cannot blame the zero sum result on the products $a_1 \cdot a_2^*$ or the complex conjugation of one of them. We must require that each square bracket vanish.

$$\boxed{[S_{11}S_{12}^* + S_{22}^*S_{21}] = 0} \tag{I.6}$$

and thus its complex conjugate—the second square bracket—also vanishes.

Finally, if we invoke reciprocity (and most optical components are dismayingly so), then:*

$$\boxed{S_{12} = S_{21} \quad \text{(by reciprocity)}} \tag{I.7}$$

which merely states that the *field* transmission going from side 1 to 2 is the same as going from 2 to 1.

We are free to assign one of the coefficients to be real (but carry its own sign) and by tradition this is $S_{11} = \Gamma_1$ and by extension $S_{22} = \Gamma_2$ (What this implies is that we choose a reference plane so that the reflected field is either in phase or out of phase with the incident. This corresponds to moving off of the true interface by no more than $\lambda/2$, which is inconsequential especially at optical frequencies. For those with masochistic tendencies, one can use (I.6) and not make this choice.) Such an assignment requires that $S_{12} = S_{21}$ (I.7) and is purely imaginary, $S_{12} = \pm jt$, and (I.6) thus requires that $S_{11} = S_{22} = \Gamma$ with no subscript necessary. The field reflection and transmission coefficients are the same

*A theorem from Onsager states that a static magnetic field is required to obtain a nonreciprocal component at any frequency, optical or otherwise.

when measured from either side. Conservation of power (I.4a) or (I.4b) states the obvious, $|\Gamma|^2 + |t|^2 = 1$ or $R + T = 1$, where $R = |\Gamma|^2$ and $T = |t|^2$.

Thus the scattering matrix for a simple ideal lossless mirror has only one truly independent parameter, Γ, and is given by

$$[\mathbf{S}] = \begin{bmatrix} \Gamma & \pm jt \\ \pm jt & \Gamma \end{bmatrix} \tag{I.8}$$

Some become rather nervous about the presence of j in the matrix for (I.8). There is a very simple physical reason for it. Consider the simplest of all dielectric mirrors consisting of a single layer of length $\lambda/4$ (in the dielectric) suspended in a vacuum. Without solving the problem in detail, we know that the phase of the transmitted wave lags the incident by 90° which is the origin of the j in (I.8). If the thickness is not $\lambda/4$, then Γ is not real, but we can choose the reference planes 1 and 2 to make S_{11} and S_{22} real by adjusting their positions with respect to the actual interface. Most dielectric mirrors will have an odd number of layers so that this gets repeated an odd number of times.

Some also get nervous about the statement $\Gamma_1 = \Gamma_2$, which would state that a mirror could be inserted into a laser cavity backwards without penalty, contrary to everyday experience. Seldom is a mirror a simple two-port network. But if we could obtain a strict two-port lossless mirror, then $\Gamma_1 = \Gamma_2 = \Gamma$ as has been derived. Most practical mirrors have the coating on one side of a substrate, which has a wedge in the manner shown (greatly exaggerated) in Fig. I.2.

For such a mirror, $S_{11} \neq S_{22}$, because the system is at least a four-port network as sketched in Fig. I.2. The scattering matrix formalism can describe it, but the tedium of the arithmetic is discouraging, but not nearly as bad as attempting to modify the $[Z]$, $[Y]$, or $[ABCD]$ formalisms.

A very simple but practical four-port whose scattering matrix can be stated by inspection is a thin beam splitter (with a 50:50 power division) shown in Fig. I.3.

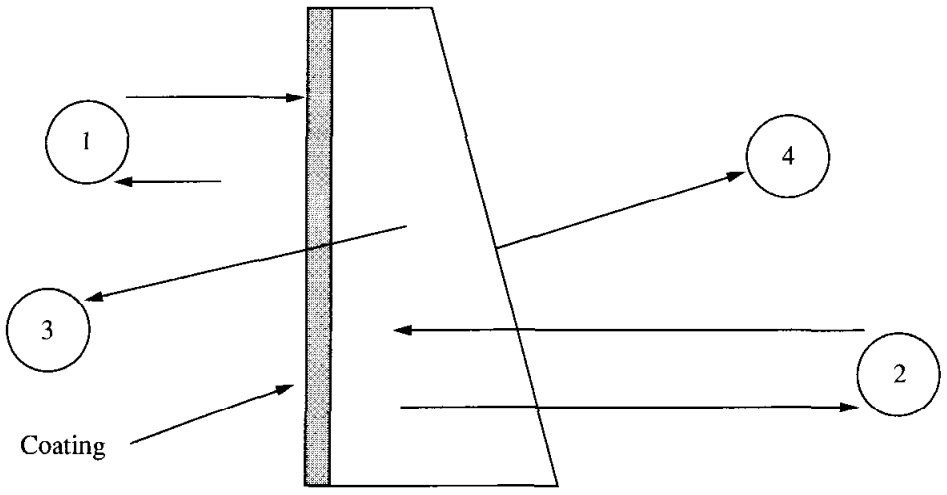


FIGURE I.2. A coated mirror on a wedged substrate. Such a mirror is often treated as if it were a two-port, but because of the extra losses and coupling to the other ports, $\Gamma_1 \neq \Gamma_2$.

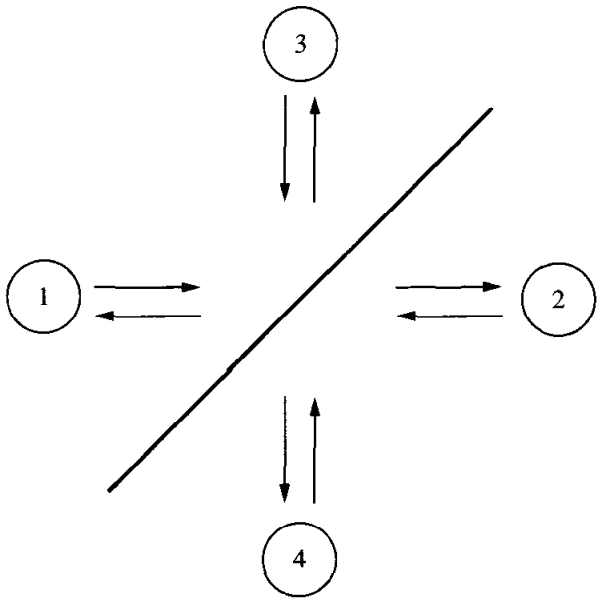
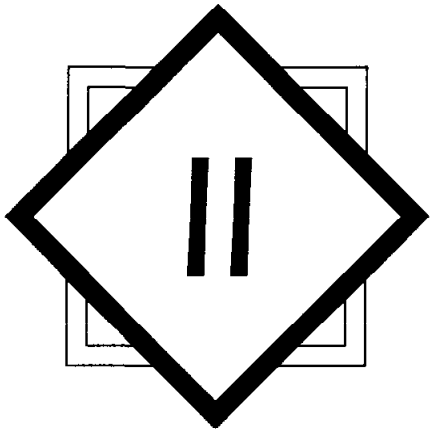


FIGURE I.3. A beam splitter.

When a scattered signal arises because of reflection from the splitter, the coefficient is $1/\sqrt{2}$, and if by transmission, the coefficient is $\pm j 1/\sqrt{2}$. If there is not a direct "ricochet" path, then the coefficient is zero. Thus the scattering matrix is

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 0 & \pm j\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \pm j\frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & 0 & \pm j\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \pm j\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \tag{I.9}$$



Detailed Balancing or Microscopic Reversibility

The purpose of this appendix is to develop theory to enable us to compute the "reverse" reaction rate given that the "forward" rate has been measured. We applied this principle in Chapter 7 when the Einstein coefficient for absorption B_{12} was shown to be directly related to that for stimulated emission by (7.3.10a)

$$g_2 B_{21} = g_1 B_{12}$$

Consequently, a measurement of one quantity—say, the absorption cross section—also determines the stimulated emission cross section.

However, the principle of detailed balancing can be applied to inverse processes involving radiation, electron collisions, chemical and nuclear reactions, or combinations. Therefore it is an extremely important principle of physics with many ramifications and important applications.

Let us consider an inelastic collision of an electron with neutral atoms in its ground state creating an excited state, say, the upper laser level.

$$e(\epsilon_i) + N \rightarrow N^*(E_2) + e(\epsilon_i - E_2) \tag{II.1}$$

Obviously, this can happen only if the initial energy of the electron, ϵ_i, exceeds E_2. Although this process is most desirable in a laser, the first idea to be emphasized is that there is *always* the *reverse* process that is deleterious to the operation. Thus an electron of energy ϵ' could collide with the atom in the excited state, $N^*(E_2)$, and deactivate it back down to the ground level:

$$e(\epsilon') + N^*(E_2) \rightarrow N + e(\epsilon' + E_2) \tag{II.2}$$

Note that any electron of any energy ϵ' is allowed to undergo collisions of the second kind (II.2), but only a select group with $\epsilon' > E_2$ have enough energy to participate in (II.1).

The foregoing equations represent reverse processes, and it must be emphasized that detailed balancing applies to that situation *only*. For instance, the same electron participating in (II.2) could also deactivate the atom from state 2 to another state 1. Detailed balancing provides no information about this alternative route.

The key is to recognize that the cross section (for whatever process) is an integral part of the atom. Even though the environment can have a major, if not dominant, effect on the ultimate behavior of an atom, the cross section for a particular process is independent of that environment. Consequently, we can imagine a hypothetical thermodynamic equilibrium situation involving these same atoms N, electrons, a radiation field (and anything else).

The word "equilibrium" implies that all quantities are independent of time. Thus the electron density, the ground-state density, the excited-state population, and the radiation flux must all be constant. It follows, then, that the net rate of production of state 2 by all causes—namely, by the absorption of a photon of energy corresponding to the $2 \rightarrow 0$ transition, or by the process of (II.1), or by whatever—must exactly balance the decay rate of state 2, that is, by spontaneous emission, by stimulated emission from $2 \rightarrow 0$, or by the process in (II.2). An equilibrium situation demands this equality between the rates but does not dictate which is most important.

A thermodynamic equilibrium demands that each interaction, taken *by itself* and *in the microscopic limit*, must yield the appropriate thermodynamic population. It is in the balancing of the rates in detail and at the microscopic limit that gives this principle its name: detailed balancing or microscopic reversibility.

Once we say that the system is in thermodynamic equilibrium, we know everything about it. For instance, we know the radiant energy density per frequency interval around $h\nu$:

$$\rho(\nu) = \frac{8\pi n^2 n_g h\nu^3}{c^3} \frac{1}{\exp(h\nu/kT) - 1} \tag{II.3}$$

and the ratio of the populations of states 2 and 0:

$$\frac{N_2}{N_0} = \frac{g_2}{g_0} \exp\left(-\frac{E_2}{kT}\right) \tag{II.4}$$

and the electron energy distribution function:

$$f(\epsilon)d\epsilon = \left(\frac{4}{\pi}\right)^{1/2} \left(\frac{\epsilon}{kT}\right)^{1/2} \exp\left(-\frac{\epsilon}{kT}\right) d\left(\frac{\epsilon}{kT}\right) \tag{II.5}$$

all being specified by a common thermodynamic temperature T.

Let us apply the principle to the electron collision sequence and show that we can generate the cross section for process 2 given the energy dependence of σ for 1. Figure II.1 shows the *detail* of the balance required by thermodynamic equilibrium. Obviously, there are some electrons with enough energy to excite N_0 to N_2, and all electrons can de-excite N_2 to N_0. As shown in Fig. II.1(a), electrons at $\epsilon = \epsilon_i$ make an inelastic collision reappear at $\epsilon_i - E_2$, whereas the superelastic processes push the group upward.

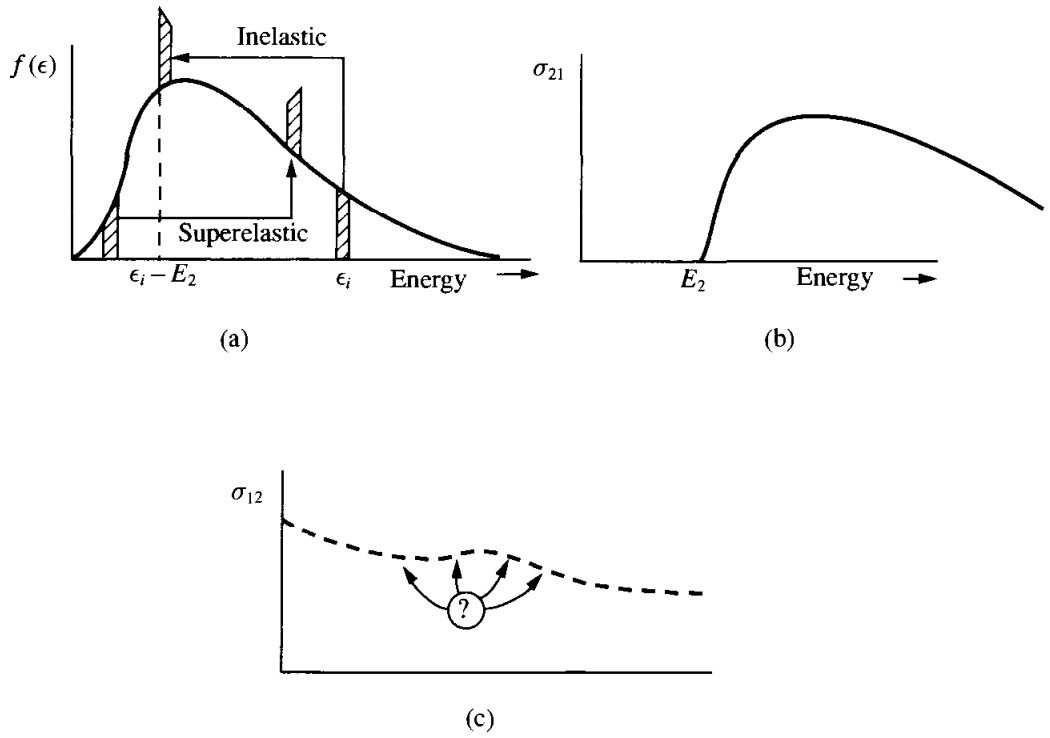


FIGURE II.1. (a) Electron energy distribution; (b) inelastic process (II.1); (c) superelastic process (I.2).

Equilibrium demands that the net rate of change of electrons in *any* energy interval be zero, and thermodynamic equilibrium demands that the electron distribution be given by (II.5). Thus we must balance, in detail, the rate *out* of a given energy interval by the rate *into* that same interval.

Consider the inelastic collisions by electrons in a small energy interval around ϵ_i. The total collision rate caused by that body of electrons is

$$dR_{02}(\epsilon_i) = \left[n \left(\frac{4}{\pi} \right)^{1/2} \left(\frac{\epsilon_i}{kT} \right)^{1/2} \exp\left(-\frac{\epsilon_i}{kT} \right) d\left(\frac{\epsilon}{kt} \right) \right] \cdot N_0 \cdot \sigma_{02}(\epsilon_i) \cdot \left(\frac{2\epsilon_i}{m} \right)^{1/2} \tag{II.6}$$

number of electrons at $\epsilon = \epsilon_i$

number of targets

cross section for inelastic process

speed of an electron with energy ϵ_i

This differential rate must be balanced by electrons leaving a correspondingly small interval around $\epsilon_i - E_2$ by the superelastic process:

$$dR_{20} = \left[n \left(\frac{4}{\pi} \right)^{1/2} \left(\frac{\epsilon_i - E_2}{kT} \right)^{1/2} \exp\left(- \frac{\epsilon_i - E_2}{kT} \right) d\left(\frac{\epsilon}{kT} \right) \right]$$
$$\times N_2 \sigma_{20}(\epsilon_i - E_2) \left[\frac{2(\epsilon_i - E_2)}{m} \right]^{1/2} \tag{II.7}$$

As shown in Fig. II.1(c), σ_{20} is unknown, but we know it exists. Detailed balancing requires that the rates given by (II.6) and (II.7) be equal, which allows one to solve for σ_{20}. After canceling common factors and combining a few terms, we have

$$\epsilon_i \left[\exp\left(\frac{\epsilon_i}{kT} \right) \right] N_0 \sigma_{02}(\epsilon_i) = (\epsilon_i - E_2) \left[\exp\left(- \frac{\epsilon_i - E_2}{kT} \right) \right] N_2 \sigma_{20}(\epsilon_i - E_2) \tag{II.8}$$

Now (II.4) is used to express N_2 in terms of N_0.

$$\epsilon_i \exp\left(- \frac{\epsilon_i}{kT} \right) N_0 \sigma_{02}(\epsilon_i)$$
$$= (\epsilon_i - E_2) \exp\left(- \frac{\epsilon_i - E_2}{kT} \right) \frac{g_2}{g_0} N_0 \exp\left(- \frac{E_2}{kT} \right) \sigma_{20}(\epsilon_i - E_2)$$

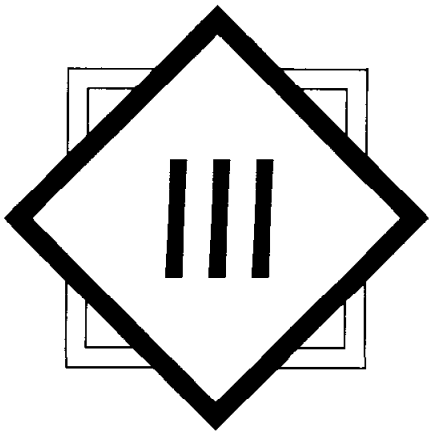
Canceling common factors of N_0 and $\exp(-\epsilon_i / kT)$ leads to the desired relationship.

$$\sigma_{20}(\epsilon_i - E_2) = \frac{g_0}{g_2} \frac{\epsilon_i}{\epsilon_i - E_2} \sigma_{02}(\epsilon_i) \tag{II.9}$$

If we substitute $\epsilon = \epsilon_i - E_2$, the equation looks better.

$$\sigma_{20}(\epsilon) = \frac{g_0}{g_2} \frac{\epsilon + E_2}{\epsilon} \sigma_{02}(\epsilon + E_2) \tag{II.10}$$

Thus, knowing the cross section for the inelastic process, we can compute the superelastic cross section. Although we have derived this by considering a hypothetical situation of the thermodynamic equilibrium, this cross section is a part of the atom, and atoms have been "found" in nonequilibrium situations (most of the time). We can thus use this value for σ_{02} for these more interesting situations.



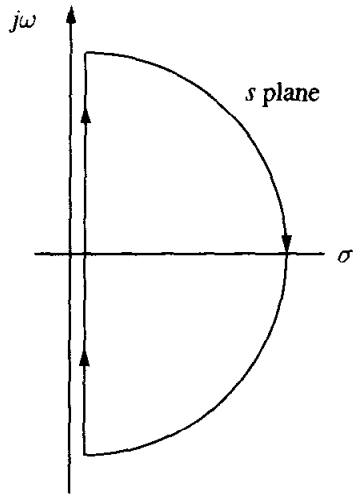
The Kramers–Kronig Relations

INTRODUCTION

The Kramers–Kronig relations relate the real and imaginary parts of a physical quantity of a generalized immittance function such as the resistance and reactance of a network, the gain (or loss) and the phase shift, or the real and imaginary parts of the dielectric constant. The significance of the relationships is that if you specify (or measure) *one* of these quantities as a function of frequency, then the other is also known (to within an additive constant term). Of course, some of the integrals may have to be evaluated numerically, but that is not a serious problem.

You are probably very familiar with the application of this theory in feedback amplifier design. We measure the gain (in dB) as a function of frequency and the result plotted as a function of the logarithm of frequency. From such "corner" or Bode plots, the phase shift can be determined. This is based on the theory to follow. The same theory is applicable to the optical domain, although its use requires considerably more sophistication.

Consider a function $F(s)$ regular in the right half of the s-plane. All physical functions fall into this class in order to obey causality. Thus $F(s)$ can be an impedance function or the complex propagation constant of an electromagnetic wave.



$$\oint_c F(s)\,ds = 0 \tag{III.1}$$

Along the real frequency axis, $s = j\omega$ and

$$F(j\omega) = u(\omega) + jv(\omega) \tag{III.2}$$

where $u(\omega)$ and $v(\omega)$ are the real and imaginary parts.

We should recognize that any physical quantity $F(j\omega)$ must satisfy the following:

$$u(\omega) = u(-\omega) \tag{III.3}$$

and

$$v(\omega) = -v(-\omega) \tag{III.4}$$

This statement can be derived from the following: suppose that B, a real physical quantity, is related to A by the function $F(s)$. For points along the imaginary axis, $s = j\omega$, we have

$$B_{\text{complex}} = F(j\omega)A_{\text{complex}}e^{j\omega t} \tag{III.5}$$

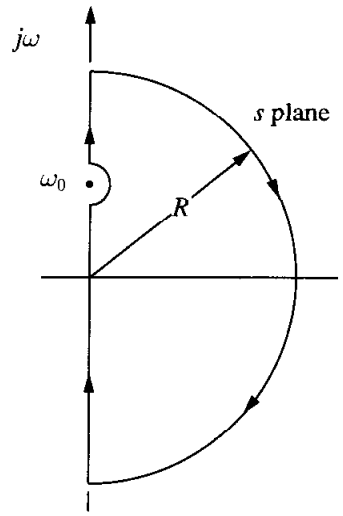
To construct the real quantity B, we add the complex conjugate:

$$B_{\text{real}} = F(j\omega)Ae^{j\omega t} + F^*(j\omega)A^*e^{-j\omega t}$$

Thus B is real if $F^*(j\omega) = F(-j\omega)$, which in turn implies (III.3) and (III.4).

Now consider the integral along the path indicated on the next page:

$$\oint \frac{F(s)\,ds}{s - j\omega_0} = 1 \tag{III.6}$$



For any respectable function in the real physical world, $F(s) = 0$ as $|s| \to \infty$. Therefore there is no contribution along the closure line ($R \to \infty$). The function inside the integral is analytic along and within the curve shown. Hence, $I = 0$. Since we are going halfway around the pole at $s = j\omega_0$, we obtain the following:

$$0 = \int_{-j\infty}^{j\omega_0^-} \frac{F(j\omega)\,d(j\omega)}{j(\omega - \omega_0)} + j\pi F(j\omega_0) + \int_{j\omega_0^+}^{j\infty} \frac{F(j\omega)\,d(j\omega)}{j(\omega - \omega_0)}$$

Collecting real and imaginary parts, we obtain

$$\pi v(\omega_0) = \int_{-\infty}^{\omega_0^-} \frac{u(\omega)}{\omega - \omega_0}\,d\omega + \int_{\omega_0^+}^{\infty} \frac{u(\omega)}{\omega - \omega_0}\,d\omega$$

$$= \int_{-\infty}^{+\infty} \frac{u(\omega)}{\omega - \omega_0}\,d\omega \tag{III.7}$$

$$-\pi u(\omega_0) = \int_{-\infty}^{+\infty} \frac{v(\omega)}{\omega - \omega_0}\,d\omega \tag{III.8}$$

Equations (III.7) and (III.8) are one form of the Kramers–Kronig relations. Now let us use the symmetry properties of $u(\omega)$ and $v(\omega)$, that is, (III.3) and (III.4).

$$\pi v(\omega_0) = \int_{-\infty}^{0} \frac{u(\omega)}{\omega - \omega_0}\,d\omega + \int_{0}^{+\infty} \frac{u(\omega)\,d\omega}{\omega - \omega_0}$$

Since the variable of integration can be anything we choose, substitute $-x$ for ω in the first integral and $+x$ for ω in the second.

$$\pi v(\omega_0) = \int_{-\infty}^{0} \frac{u(-x)\,d(-x)}{-x - \omega_0} + \int_{0}^{\infty} \frac{u(+x)\,d(+x)}{x - \omega_0}$$

$$= \int_{0}^{\infty} u(x)\left(\frac{1}{x - \omega_0} - \frac{1}{x + \omega_0}\right) dx$$

Collecting terms yields

$$v(\omega_0) = \frac{2\omega_0}{\pi} \int_0^\infty \frac{u(x)}{x^2 - \omega_0^2}\, dx \tag{III.9}$$

Manipulation of (III.8) proceeds in the same fashion:

$$\begin{aligned} -\pi u(\omega_0) &= \int_{-\infty}^0 \frac{v(\omega)\, d\omega}{\omega - \omega_0} + \int_0^\infty \frac{v(\omega)\, d\omega}{\omega - \omega_0} \\ &= -\int_0^{+\infty} \frac{v(-x)\, d(-x)}{-x - \omega_0} + \int_0^{+\infty} \frac{v(+x)\, d(+x)}{x - \omega} \\ &= \int_0^\infty v(x) \left(\frac{1}{x + \omega_0} + \frac{1}{x - \omega_0} \right) dx \\ u(\omega_0) &= -\frac{2}{\pi} \int_0^\infty \frac{x v(x)\, dx}{x^2 - \omega_0^2} \end{aligned} \tag{III.10}$$

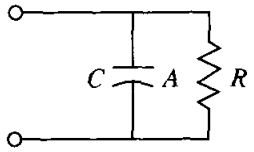
Equations (III.9) and (III.10) are one of the forms quoted for the Kramers–Kronig relations. In all but the simplest cases, we must resort to numerical integration to utilize these relations, but the fact that they exist is important in itself.

Example

One of the few cases that can be handled by analytic techniques is that of the parallel combination of a resistor and a capacitor. The driving-point impedance is given by

$$Z(j\omega) = \frac{R(1/j\omega C)}{R + (1/j\omega C)}$$

$$R(\omega) + jX(\omega) = \frac{R}{1 + \omega^2\tau^2} - \frac{j\omega\tau R}{1 + \omega^2\tau^2}$$



Thus a suitable candidate for $u(\omega)$ of (III.9) is $R(\omega)$, and $v(\omega)$ is identified with $X(\omega)$. Let us be "blind" to the equation above and assume that we know $R(\omega)$ only.

$$X(\omega_0) = \frac{2\omega_0}{\pi} \int_0^\infty \frac{R}{1 + x^2\tau^2} \frac{1}{x^2 - \omega_0^2}\, dx$$

Use a partial-fraction expansion for the integrand:

$$\frac{R}{1 + x^2\tau^2} \frac{1}{x^2 - \omega_0^2} = \frac{A}{x + \omega_0} + \frac{B}{x - \omega_0} + \frac{C}{x\tau + j} + \frac{D}{x\tau - j}$$

where

$$A = \frac{R}{1 + \omega_0^2\tau^2} \frac{1}{(-2\omega_0)} \qquad C = \frac{R}{2j} \frac{\tau^2}{1 + \omega_0^2\tau^2}$$

$$B = \frac{R}{1 + \omega_0^2\tau^2} \frac{1}{2\omega_0} = -A \qquad D = -\frac{R}{2j} \frac{\tau^2}{1 + \omega_0^2\tau^2} = -C$$

Now we can integrate this expansion term by term:

$$\int_0^\infty \frac{A\,dx}{x+\omega_0} + \int_0^\infty \frac{-A\,dx}{x-\omega_0} = -2\omega_0 \int_0^\infty \frac{dx}{x^2-\omega_0^2}$$

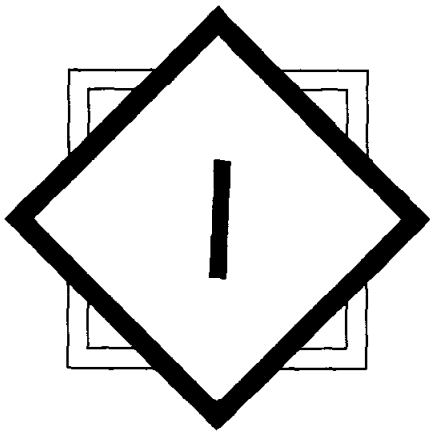
$$= 2\ln\left|\frac{\omega_0 - x}{\omega_0 + x}\right|\Bigg|_0^\infty = 0$$

$$\int_0^\infty \frac{dx}{x^2\tau^2+1} = \frac{1}{\tau^2}\int \frac{dx}{x^2+(1/\tau^2)} = \frac{\pi}{2\tau}$$

Therefore

$$X(\omega) = \frac{2\omega_0 R}{\pi(1+\omega_0^2\tau^2)}\left(\tau^2\frac{\pi}{2\tau}\right) = \frac{\omega_0\tau R}{1+\omega_0^2\tau^2}$$

Thus after considerable work, we obtain the expected result.



Index

A_{21}, Einstein's coefficient for spontaneous emission, 178
  classical value, 597
  derivation from Schrödinger's equation, 624
  relation to B_{21}, B_{12}, 182
  relation to the oscillator strength, 594
ABCD law: definition, 76
  applied to free space, 76
  applied to a graded index fiber, 100
  applied to a thin lens, 77
  applied to stable cavities, 135
ABCD matrix: definition, 36
  of a curved mirror, 39
  of a dielectric discontinuity, 56
  of a gas lens, 54
  of a GRIN fiber, 53
  of a homogeneous length, 37
  of a periodic sequence of lens, 40
  of a tapered mirror (*see also* prob. 3.19), 562–63
  of a thin lens, 38
Absorption
  cross-section, 189
  Einstein coefficient for, B_{12}, 179
  in GaAs, 461
  oscillator strength for, 594
  relation to gain coefficient, 189
  in a semiconductor, 452
  3-photon absorption, 645
Active cavity, 157, 240
Active mirrors, 528
Additive mode locking, 321
$Al_xGa_{1-x}As$ lasers
  heterobarriers in, 476
  material dependence on x, 469
Alexandrite lasers, 394–96
Al_2O_3 (*see* ruby or Ti:sapphire lasers)
Algebraic form of Maxwell's equations, 10
α, symbol for distributed loss, 208
  or Townsend ionization coefficient, 753
α_e, correction to rotational constant, 685
Amplification coefficient (*see* gain coefficient)
Amplified spontaneous emission
  effect on Q switching, 295
  limitation on gain, 237
Amplitude oscillations, in gain switch lasers, 282

Angular momentum J
relation to degeneracies g_1, g_2, 181
relation to L, S, 682
Angular spread
of a Gaussian beam, 70
in terms of integral equation, 573
Anisotropic media, 16–19
ordinary, extraordinary waves, 18, 19
wave propagation in uniaxial crystal, 16
Anomalous dispersion
in a fiber, 114
near an atomic resonance, 595
Antiguiding (*see* prob. 12.11), 579
Anti-Stokes emission, 652
Ar^+Cl^-, Ar^+F^- lasers (*see also* excimer lasers), 412–14
Area, theorem, 670
of a pulse, 665
Argon
elastic scattering cross-section, 739
ion laser, 403–5
ionization potential, 414
metastable levels, 414
Arrays of lasers, 568–74
Astigmatism, 49
astigmatic distance in gain guided lasers, 582
from Brewster angle windows, 50
in cavities, 51
in gain guided lasers, 514
from spherical mirrors, 50
Atomic response to external field
classical analysis, 591
density matrix, 628
Schrödinger's description, 624–27
Atomic transitions
in argon ion laser, 403
in Cr^{3+} (ruby laser), 352
in Erbium, 373
isotope shift
in hydrogen (deuterium), 224
in mercury, 695
in neon, 223
in Nd:glass, 370
in Nd:YAG, 360
in Neon, 399
Attachment
in CO_2 laser discharges, 754
in excimer discharges, 416
Attenuation coefficient (*see also* absorption coefficient)
evanescent decay of fields, 93
relation to the atomic susceptibility, 593
Avalanche photodiode, 721
Average electron, 734

B_{21} (B_{12}), Einstein coefficient for stimulated emission (or absorption), 179–80
relation to A_{21}, 182
from Schrödinger's equation, 623
Background radiation, 717
Bandgap
discontinuity in heterojunctions, 476
engineering, 473
of various semiconductors, 444
$B_{e,v}$ rotational constants, 685
Beam slicer, 538
Beams
guided, 85
Hermite-Gaussian (*see also* Gaussian beams), 63–79
Beat note, 727
Bernard and Duraffourg condition for gain, 454
Bistable cavity, 608
Blackbody radiation
Planck's law, 173–78
background radiation, 717
Rayleigh–Jean's limit, 177
Bleaching (*see* saturation)
Bohr orbits (relation to spectroscopic notation), 682
Boltzmann distribution and relations, 177
distribution with a band, 376
for rotational state on a molecule, 686
for thermal population, 235
Bound electron, 591
Boundary value problems in electromagnetics, elementary, 20
Bragg reflectors (*see* distributed feedback)
Bragg wavelength (*see also* distributed feedback), 518, 528
Branching ratio, 262
Brewster's angle, 21–23
astigmatism from, 50
prisms to compensate for dispersion, 321
in semiconductor cavity, 508
Broadening
homogeneous (lifetime, collision, natural), 191
inhomogeneous (Doppler, isotope, phonon), 196
Built-in potential, 703

c axis of ruby, 351
CO_2 lasers, 405–11
discharge excitation, 748–58
electron beam excitation, 758–61
physical data, 408
vibration structure, 406
CO lasers
rotational data, 686
vibrational data, 687

VR spectra on 7-6 band, 689
Cathode dark space (CDS), 732
Cavities
confocal geometry, 43
folded, 61
stable, 42
unstable, 44, 534–42
Cavity
definition of a mode, 133
linewidth
relation to Q, F, 150
relation to τ_p, 153
Causality, 774
Characteristic energy of electrons
for any distribution, 752
in CO_2 mixtures, 749
for a Maxwellian, 737
relation to E/N, 738
Child–Langmuir law, 761
Chirp
due to dispersion in fibers, 113
phase modulation in modelocking, 607
in semiconductor lasers, 494
in soliton propagation, 117
Chirped grating, 533
Chromium
as a dopant in Al_2O_3, 353
as a dopant in chrysoberyl ($BeAl_2O_4$), 392
Classical model of an atom, 591
Collision broadening, 193
Coherence, 23–31
Collisions
effect on density matrix, 632
of electrons, 744
inelastic, superelastic, 735, 772
line broadening, 193
Commutator notation, 634
Complex beam parameter $q(z)$ relation to, 66
cavities with tapered mirrors, 565
differential equation for, 66
radius of curvature, 68
spot-size, 68
stable cavity parameters, 135
transformation (*see also* $ABCD$ law), 76
Complex dielectric constant
in an active medium, 511
relation to the susceptibility, 592
Conductivity, electrical
free electron
average electron approach, 736
in terms of the distribution function, 746
relation to photon lifetime, 603
Confocal cavity
stable
by $ABCD$ law, 44
by integral equation, 550
unstable
analysis, 540
burn pattern, 542
Confocal parameter, 70
Constitutive relations, 9, 592
Continuous lens, 51
Coupled mode analysis, 520
Coupling coefficient (κ), 520
Coupling, electromagnetic
coupling loss, 290
efficiency (η_{cpl}), 261
Q-switch laser, 290–94
optimum, for a ring laser, 267
CPM (colliding pulse modelocked laser), 317
Critical fluorescence power
for an excimer, 415
for ruby, 357
Current amplification
in avalanche photodiodes, 707
in photomultipliers, 700
Current spreading in semiconductor lasers, 510

***D*, delay time dispersion**, 111
Damping constant (of a driven oscillator), 591
Debye (a measure of dipole moment):
($1\ D = 3.33 \times 10^{-30}$ c-m), 626
Degeneracy
of atomic levels ($2J + 1$), 181
of cavity resonant frequencies, 155
δ
fraction of excess energy lost per collision, 738
secondary emission coefficient, 700
Density of electrons/holes
relation to Fermi level, 450
Density of electron/hole states, 445
area density of, 472
joint or reduced, 455
for a quantum well, finite barrier, 481
for a quantum well, infinite barrier, 472
Density matrix
definition (Sec. 14.5), 627
equation of motion for, 634
vector form of, 668
Density, optical, 28
Detachment (in excimers), 416
Detailed balancing, 770
Detectors, quantum, 698–707
Detuning (*see also* mode pulling)
from passive cavity resonance, 605

Diatomic molecules
rotational, vibration, electronic transitions, 685–91
Dielectric constant
complex, 82
relation to susceptibility, 10
Difference equation (for a ray), 41
Diffraction losses, 155
hole coupling, 157
in a planar mirror system, 549
in stable cavities, 157
in unstable cavities, 541
Diffusion length, 466
Diffusion, transverse, 752
Dipole antenna, 596
Dipole moment
relation to gain coefficient, 642
for unpolarized light, 623
Direct bandgap semiconductors, 443
Discharge pumping
of CO_2 lasers, 748
of excimer lasers, 416
Dispersion
D, delay time dispersion, 111
in fibers, 106–8
modal, 99
normal, anomalous, 114
in prisms, 321
Displacement current, 9
for uniaxial crystal, 17
Dissociation energy, 684
Distributed feedback laser, 517–31
Bragg reflector for VCSEL, 484
Distribution function, 741
Divergence, transverse (∇_t), 64
Donut mode, 81, 567
Doppler effect
line shape for, 198
saturation in, 232
Drift velocity, 735
Driven oscillator, 591
Dye lasers, 386
absorption/emission, 390
cavities for, 395
cw configuration, 390
molecular model, 389
Dynodes, 700

E-beam
excitation of excimer lasers, 415
in FEL lasers, 417
sustained operation of CO_2 laser, 758
Effective focal length (*see also* astigmatic cavities), 50
Effective index
in fibers/heterojunctions, 92, 107, 516
in uniaxial crystal, 19
Effective mass, 444
Efficiency (*see also* quantum, coupling and pumping efficiencies)
factors in, 261
stimulated emission, 292
EH, HE modes in step index fibers, 91
Einstein coefficients
A_{21}, B_{21}, B_{12} as defined by rate equations, 179
from Schrödinger's equation, 617, 621–24
Einstein relation, 752
Elastic collision frequency, 735, 744
Electric current
average electron approach, 735
from distribution function, 746
Electrical length of a cavity, 145
Electron affinity (of halogens), 414
Electron gas, 734
Electron lifetime, 746
Electron mobility
average electron approach, 735
in CO_2 mixtures, 752
Electron oscillator model, 591
Electron temperature
in a fluorescent lamp, 740
in He/Ne laser, 399
relation to E/N, 738, 752
Emission cross-section
relation to absorption, 382
relation to fluorescence, 383
E/N, characteristic parameter of a discharge, 738
of CO_2 laser mixtures, 749
of He/Ne laser, 399
Energy balance equation, 737
Energy density, blackbody, 178
Energy extraction, by a pulse, 315
Envelope (of the field)
in mode locking, 305
of a pulse, 109
in self-induced transparency, 666
Equivalent lens waveguide, 39
Erbium doped fiber amplifier (EDFA), 373
data for AlP/SiO_2 fibers, 375
energy levels, 373
Excimer lasers, 411–17
Expansion angle, of a Gaussian beam, 70
Extraction efficiency
in Q switching, 292
in a ring laser, 267
Extraordinary wave, 18

$f\#$ (f-number) of a lens, 78
Fabry–Perot cavity
  quality factor, Q, finesse, F, and linewidth, $\Delta\nu$, 151
  transmission, 149
  tuning by gas pressure, 166
Facet reflectivity, 440, 509
FEL (free electron laser), 417
  wavelength of, 422
Fermat principle, 56
Fermi level, 449
  (*see also* quasi-Fermi levels), 450
Fibers
  graded index, 96
  step index
  wave equation (TE, TM modes), 90
  waveguide model, 90
  zigzag analysis, 87
Finesse (Sec. 6.3)
  definition, 151
    relation to Q and τ_p, 153
  of a Fabry–Perot cavity, 151
Fill factor, f, 602
Filter, quarter wave, 525
$F(J)$, rotational energy, 685
Flashlamp
  efficiency, 357
  spectral emission from Xe or Kr lamp, 357
Fluorescence
  dye lasers (R6G), 390
  Er doped fiber amplifiers, 374
  excimers, 415
  ruby, 357
  Ti:sapphire lasers, 393
Fluorescent lamp
  calculations, 740
  terminal characteristic, 731
Flux
  of electrons, 745
  of photons, 186
Four-level laser, 262
Fourier transform
  for band limited pulse, 109
  cosine pairs, 302, 306
  definition, 10
  in noise calculation, 709
Fox and Li results, 547–53
Franck–Condon principle, 692
Frantz–Nodvik equations, 313
Free electron laser, 417–23
Free spectral range (FSR), 147
Frequency, definition of, 114
Frequency chirp (*see* chirp)
Frequency pulling (*see* mode pulling)
Fresnel number
  definition, 550
  equivalent value, 561
Fresnel reflectivity, 508
Fringe contrast and visibility, 27

$g = \gamma l_g$, single pass, line integrated gain, 275
  $g_0 = \gamma_0 l_g$ (small signal value), 307, 520
$g_{1,2} = 1 - d/R_{1,2}$, cavity parameters, 43
$g_{1,2} = 2J_{1,2} + 1$ statistical weights of the states 1 and 2, 181
$g(\nu)$ = line shape such that $\int g(\nu)d\nu = 1$, 184–85
  $\bar{g}(\nu)$ such that $\bar{g}(\nu_0) = 1$, 220
GaAs (*see also* AlGaAs lasers)
  data, 444
  index of refraction of, 469
  spontaneous emission/absorption in, 461
Gain (output/input), 190
  absorption, 189
  bandwidth, 209
  relation to gain coefficient, 190
  in a semiconductor, 453
  small signal value, 207
gain coefficient (γ)
  complex value, 306
  definition, 189
  from density matrix, 642
  for P and R transitions, 689
  relation to pumping/lifetimes, 219
  relation to susceptibility, 593
  saturation of
    density matrix formulation, 642
    homogeneous medium, 215
    inhomogeneous medium, 223
  in terms of ϵ', ϵ'', 82
  in words, 190
Gain coupled DFB, 520
Gain guiding, 509–17
Gain switching, 240, 281
Gaussian beams, 63–79
  characteristics
    equiphase surfaces, 71
    radius of curvature, 68
    spot size, 67
  "dot" pattern of higher order modes, 75
  expansion law (*see* $ABCD$ law)
  in fibers, 98
  in simple resonators, 130
Gas lens, 52
Giant pulse (*see* Q switched lasers)
Graded barrier in AlGaAs system, 474
Grating-pair for pulse compression), 320

Green's theorem (inside front cover), 545
  function, 544
GRIN
  lens, 124
  wave guide, 96
Group index, 177
Group velocity
  definition, 107
  dispersion of, 111
Guided mode, 92

Hamiltonian
  for the energy in a cavity, 599
  for the energy operator, 618
  matrix elements of, 635
Hamilton's equation of motion, 599
Harmonic oscillator
  model of an atom, 591
  of a molecule, 687
Harpoon reaction, 413
Heavy (or normal) hole, 447
Helium/Neon laser, 397–403
  energy level diagram, 398
  transition data, 393
Helmholtz equation (*see* wave equation), 10, 92, 543
Hermite-Gaussian beam modes, 73
  differential equation, 98
  dot pattern from, 75
  Hermite polynomials, 73
Hermitian, definition, 630
Heterodyne principle, 722
  detection limits, 724
Heterostructures
  bandgap discontinuity in, 468
  lasers, 467
  waveguide model for, 87–90
Higher order modes (*see* Hermite-Gaussian beam modes)
High Q approximation
  for optimum coupling, 267
  for standing wave lasers, 278
Hole burning, spectral
  hole width, 231
  phenomena, 225
  relation to Rabi frequency, 640
Hole coupling, 156–57
Homogeneous broadening
  collision effects, 193
  linewidth in neon, 677
  natural linewidth, 196
  theory for, 191
Homojunction lasers, 464
Homonuclear molecules, 405
Hugyen's principle, 134
Hyperbolic secant pulse, 120, 674

Ideal laser, 223
  excimer approximation, 413
Image points, unstable resonators, 534
Impedance, wave, 12, 71
Index coupled DFB, 520
Index ellipsoid, 19
Index of refraction
  free carrier contribution to, 532
  of GaAs, AlGaAs semiconductors, 469
Indirect semiconductors, 443–44
Induced emission (*see* stimulated emission)
Inelastic collision, 744
  relation to superelastic ones, 773
Inhomogeneous broadening, causes of
  Doppler, 197
  isotope, 196, 274
  Stark, 196
Initial condition, ray tracing, 44–47
Injection lasers, 464
Integral equation, general formulation, 543–46
  applied to confocal resonator, 550–54
  in planar mirror cavities, 547–50
  in unstable cavities, 555
Integrated spectral absorption, 593
Intensity, definition (*see also* Poynting vector), 26
Interference, 23
Internal reflection (slab waveguide), 86
Inversion
  for an atomic system, 189
  lifetime (in density matrix formulation), 637
  pumping requirements for, 214–15
  in a semiconductor, 460–64
Ion lasers
  argon energy levels, 404
  cadmium, 425
Ionic bonding, 412
Iodine laser ($\lambda_0 = 1.315\ \mu$m), 683
Ionic channel (for excimer formation), 413
Ionization
  potentials of rare gas, 414
  rate for, 746
    relation to Townsend ionization coefficient, 753
Isotope effect
  in line broadening, 196, 274
  in mercury, 695

***J*, total angular momentum quantum number**, 181, 682
Jello, as a laser, 347

"Johnson" noise, 714
Joint density of states, 455
Junction detectors, 703
  avalanche photodiode, 707
  pin diodes, 706

k conservation rule (*see also* Franck–Condon principle), 443
$k = \omega n/c$, uniform plane wave phase constant, 10
kcal/mole; an energy unit, (inside cover)
Kerr effect, 115, 323
  broadening of a pulse, 320
  in fiber, 115
Kirchoff's radiation law, 235
Kramer–Kronig relation, 774
Kr^+Cl^-, Kr^+F^- lasers (*see also* excimer lasers), 414
k-space
  of allowed modes in a cavity, 176
  of Gaussian beam, 15
  of momentum states in a semiconductor, 446, 471

Lamb dip, 228
  inverted, in methane, 229
Lamp emission, 356
Laser amplifier
  power added relation, 222
  saturated gain, 221
  small signal gain, 219, 222
Laser arrays
  linear arrays, 570
  vertical cavity arrays, 482–86
Laser oscillation
  dynamic behavior, 274
    damped oscillations in, 282
    modulation response, 279
  ring configuration, 264
  standing wave, 268
  in 3, 4 level lasers, 261, 348
Lattice constant of common semiconductor, 444
Lens, thin
  $ABCD$ matrix for, 38
  f number and focal spot size for, 78
  field transmission coefficient for, 557
  transformation of a beam, 55
Lenslike media, 51
Lifetime
  broadening, 191
  of inversion, 637
  ratio, 220
  of states, 213
Line integrated gain (γl_g), 275
Line shape, physical interpretation, 183–85
  Doppler, 198
  Lorentzian, 195
  relation to stimulated emission cross-section, 186
  in a semiconductor, 480
Linewidth of oscillation, 241–42
Littman configuration, 396
Local time, 111, 312, 673
Longitudinal phase
  of a Gaussian beam, 69
  of a guided mode, 91
Lorentzian line shape, 195
  from the classical driven oscillator, 593
  from the density matrix, 641
Lorentz force equation, 421
Losses in fibers, 105
Loss coefficient, for cavities
  effective value, 208
  strip mirrors, 549
  tapered mirrors, 566
  unstable cavities, 537
L-S coupling, 682
LTE (local thermodynamic equilibrium), 377

Magnetic dipole transition in iodine, 683
Magnetic moment
  of an antenna, 611
  relation to angular momentum, 669
Magnetization, 32
Magnification, **M**, of beam size, 536
Maximally flat output (*see also* tapered mirrors), 567
Maxwell–Boltzmann distribution
  for atomic states, 197
  of electrons, 742
Maxwell's equation
  algebraic form, 11
  coupled to atoms, 590
  time independent, 9
Mechanical periodicity
  of a distributed feedback, 520
  of a wiggler, 418
Metastable levels
  in N_2, A state, 693
  in rare gases, 414
Michelson interferometer, 26
Microscopic reversibility (*see* detailed balancing, Appendix II), 770
Minimum spot size, 67
Minority carriers
  in photodiodes, 703
  in semiconductor lasers, 464
Mobility
  of (e, h) in common semiconductors, 444

Mobility (*cont.*)
of electrons in gases, 735
Modal gain
(*see* confinement factor), 482
(*see* fill-factor), 602
relation to material gain, 483
Mode
definition in a cavity, 133
matching, 142
in a transmission system, definition
via the $ABCD$ law, 100
phase constant of, 573
using a ray approach, 87
from the wave equation, 90
Mode density, electromagnetic
in a cubical cavity, 177
in semiconductor laser cavity, 489
Mode discrimination (in unstable resonators), 560
Mode locking
additive pulse, 321
colliding pulse (CPM), 317
equal amplitude modes, 298
forced
amplitude modulator, 305
phase modulation, 310
Gaussian, 301
modelocking force, 323
Mode pulling, 595
Mode volume, 137
Modulation of lasers
for modelocking, 305
resonance in, 280, 491
of a semiconductor laser, 489–91
variation of the pump, 277
Molecular constants, 684–91
Momentum
of electron/holes in a semiconductor, 442
of a photon, 443
Multimode oscillation, 223, 229–30
Multiple quantum well, 474
Multiphoton effects, 643
Multiplication, current
in APD, 721
in photomultiplier, 700

N_p, number of photons in a cavity, 152, 288
N_2 (nitrogen)
electronic transition laser, 692
energy storage in, 400
molecular data, 408
role in CO_2 laser
vibrational levels, 406
N_{tr}, density at optical transparency, 487
Natural broadening, 193
Nd:Glass, 369–71
energy levels, 370
physical data, 371
Nd:YAG
analysis, 363–69
data, 361–62
energy levels, 359
Negative branch, unstable resonator, 539
Negative glow, 732
Neon
model for, 401–3
transitions in, 399
Noise
background, 717
in electronic amplifiers, 716
Johnson, Nyquist, Planck, 715
shot, 731
spontaneous emission, 188
Noise figure, 716
Nonlinear Schrödinger equation (NLS), 118
Normal mode (*see* characteristic mode)
Normalization condition
on electron distribution function, 741
on line shape, 183
Norton equivalent circuit, 698, 718
Nuclear potential in excimer, 412
Numerical aperture (NA), 89
Nyquist noise, 714

Occupation probability, 619
Optic axis
in ruby (c axis), 353
in uniaxial crystal, 18
Optical confinement factor, 482, 486
Optical diode, 264
Optical pumping, 351
Optical resonators
fields in stable cavities, 130–39
integral equation approach, 544
losses in
coupling, 148–53
diffraction, 155, 549
resonance
general, 146
of a Hermite-Gaussian mode, 154
with tapered mirrors, 562
unstable
geometric optic approach, 535
field analysis, 555
Optically "thin" spectral lines, 235

Optical transparency
in ruby
in a semiconductor, 487
Optimum coupling, 267
Optimum pressure-diameter product, 398
Ordinary wave, 18
Orthogonality
of atomic wavefunctions, 618
of Hermite-Gaussian functions, 81
of Slater modes, 598
Oscillation
effect on line shape (*see also* hole burning)
homogeneous line, 211
inhomogeneous line, 223
frequency, 209
threshold for, 5, 208
Oscillator strength, 594
Oscillatory behavior, 284

P-branch (*see also* VR transitions), 688
Paraxial
ray, 36
wave equation, 65
Partial inversion, 691
Paschen notation (for neon), 399
Passband of a quarter-wave filter, 526
Periodic
acceleration in FEL's, 418
lens waveguide, 40
reflections in DFB's, 517
Perturbation
caused by optical Kerr effect, 117
theory, for atomic response, 619
theory, for guided waves, 101
Phase constant
for guided modes in a slab fiber, 505
Hermite-Gaussian beam, 69
of the supermodes of an array, 572
Phase interrupting collisions (*see also* T_2)
effect on coherence, 25
effect on induced polarization, 631
line broadening, 193
Phase modulation (for modelocking), 310
Phase velocity
in fibers, 107
of a Gaussian beam, 71
Phasor diagram
for a modelocked laser, 299
for waves in a cavity, 146
Phonon terminated laser, 382
Photodetectors, 698–707
Photoelectric effect, 697
Photomultipliers, 699
current gain in, 700
Photon lifetime, 151
cavity linewidth, 153
relation to Q, F, 153
relation to survival factor S, 152
of Slater modes, 603
Photon dynamics
approach to oscillation, 275–84
in a passive cavity, 152
in Q-switched laser, 288–93
Photo response of cathodes, 701
Photovoltaic operation, 705
pin diode, 704
Planck's radiation law, 175
noise (per mode), 714
for an optically thick transition, 235
Polarization current
classical derivation, 591
from density matrix, 639
in Maxwell's equation, 9, 590
in terms of dipole moment, 628
Polarization of the fields
Brewster's angle, 22
of modes in a heterostructure, 508
in uniaxial crystal, 18
Positive branch, unstable resonators, 539
Positive column, 732
Power balance in a gas discharge, 737
Power broadening, 626
Power gain, 207
Power output, optimum coupling, 267
Poynting vector, 12
Pressure broadening, 193
of CO_2 lines, 195
Pulse propagation (*see also* self-induced transparency)
distortion of, 315
in fibers (*see also* solitons), 109–15
in saturable amplifier/absorbers, 311
Pulse width
see also mode locked laser, 298–323
see also Q-switched lasers, 293
Pumping
effective, 223
efficiency, 261
methods, general, 347
Pythagorean relation, 91

***Q* branch of VR transitions**, 688
q complex beam parameter
in the field expression, 66
in terms of R, w, 76

q complex beam parameter (*cont.*)
transformation law ($ABCD$), 76
in a cavity, 135
Q, quality factor for cavity, 148–50
relation to F, $\Delta\omega$, τ_p, 150–53
relation to survival factor, 152
Q-switching, 284
Quadratic gain profile (*see* gain guiding), 509
Quadratic index
fibers and gases, 51–52
ray matrix for, 53
wave solution in, 96
Quantum detectors, 697
Quantum efficiency
of detectors, 697
of a laser, 261
of photocathodes, 701
Quantum hypothesis, 173
Quantum size effect (QSE), 469
band theory of, 476–81
quantum well lasers, 476
Quarter-wave bandpass filters, 525
Quasi-Fermi levels, 450
Quenching effects in rate equations, 192

R-branch of (VR transitions), 688
$R\#$ (R number)
relation to the V number, 96
slab fiber, 95
Rabi flopping frequency, 621
Radius of a curvature
of gain guided modes, 513
for a Hermite-Gaussian mode, 68
of modes in stable cavities, 131
in planar cavities, 548
in unstable cavities, 534
Radiation damping, 610
Radiation, from a supermode, 572, 585
Radiative processes, lifetime, 179
Raman effect, 643
classical analysis, 653
density matrix analysis, 659
gain coefficient, 655, 663
scattering in air, 658
scattering cross-section (exp), 657–58
Rare earth (RE) laser, 414
Rare-gas halide (RGH) laser, 414
Rate equations, 213
Ray matrix
for a gas lens, 55
for a GRIN lens or fiber, 53
for simple components, 35–37
Ray tracing, 35
in inhomogeneous medium, 52
in optical cavities, 39
Ray path, repetitive, 47
Rayleigh–Jeans distribution, 177
Recombination
of electrons and holes, 442, 463
rate in GaAs, 463
spectra in a semiconductor, 461
of (e, ions) in CO_2 lasers, 753
Reflection
band of a $\lambda/4$ filter, 526
of a Bragg reflector, 524
coefficient of a Fabry–Perot cavity, 149
of a distributed feedback structure, 523
Relativistic factor, $\gamma = [1 - (v/c)^2]^{-1/2}$, 419
Relaxation
oscillations in lasers, 281
terms in density matrix, 637
Resonance, 144
basic definition, 146
of Hermite-Gaussian beams, 154
Rhodamine 6G, absorption/fluorescence, 390
Rigrod analysis, 271
Ring laser, 264
Rotating wave approximation (RWA)
for atomic interactions, 620
for DFB lasers, 521
Rotational transitions, 685
Ruby laser, 351
absorption in, 354
data (R_1, R_2 lines), 353
energy level diagram, 352
Russel–Saunders coupling, 692
$R(\nu)$, spontaneous recombination spectra in a semiconductor, 460

S, photon survival factor, 5, 152
SI (or MKSA) system of units, 8
Sampled grating, 533
Saturated amplifier, power added, 222
Saturation
derivation from density matrix, 640–41
derivation from rate equations, 218–20
of inhomogeneous broadening, 230
intensity
gain saturation definition, 220
lifetime definition, 216
saturated absorber, 317
saturated absorption in CH_4, 229
Scaling laws for a discharge, 399
Scattering cross-section
electrons in argon, 739

Raman, 657
Scattering matrix (Appendix I)
beam splitter, 769
Bragg structure, 521
definition, 765
of two mirrors, 526
Schrödinger equation
derivation of Einstein coefficients, 617
for a finite barrier, 477
Secondary emission amplifier, 700
Seeding of a laser, 264
Selection rules
argon ion laser lines, 403
helium neon lines, 399
LS coupling rules, 682
Self focussing, 125
Self-induced transparency, 664
Self-reversed line, 236
Semi-classical quantum theory, 600
Semiconductors data
$Al_xGa_{1-x}As$ data, 468
cavities, 503
detectors, 701–7
general data, 444
Shot noise, 711–14
Signal-to-noise ratio, 718–24
heterodyne, 724
video detection, 718–20
Singlet states in dyes, 389–90
Slater modes, 597
Small signal gain coefficient γ_0, 189
Snell's law, 20
Solitons, 115
fundamental solution, 118
hyperbolic secant pulse, 120
period, 117
Space charge
limitation (Child–Langmuir law), 761
neutrality in a discharge, 733
Spatial coherence, 79
Specific refractivity of a gas, 52
Spectral distribution of power
in an active cavity, 157–58
due to ASE, 236
in a line (*see also* line shape), 10
of noise, 710
width of a mode-locked laser, 297
Speed (of an electron), 741
Spectroscopic rotation
for atoms (LS notation), 682
for molecules, 691
in rare earth, 358
Spherical wave
relation to Gaussian beam, 72
in unstable resonators, 534
Spontaneous emission
into a cavity mode, 239
Einstein coefficient for, 179
in a semiconductor, 459
Spontaneous power
of excimer laser, 415
relation to critical fluorescence power, 357
Spot-size
definition, 67–69
at the focal plane of a lens, 78
Spreading of a beam, 70
Stability
criterion for cavities, 39
diagram for, 42–43
Standing wave lasers, 268
Stark effect
splitting in rare earths, 358
shift (from density matrix), 639
Statistical weights ($g = 2J + 1$), 181
of rotation states, 686
Stimulated emission
characteristics of, 181
cross-section, 213
Einstein coefficient for, 179
rate in a semiconductor, 456
in the rate equation, 215
Stokes lines (*see* Raman effect), 652
Superelastic collisions, 771
Supermodes of an array, 570–71
Superstructure grating (*see also* "chirped" grating), 533
Susceptibility, complex, 9
classical derivation, 592
from density matrix, 641–42
relation to gain coefficient, 642

T_2, collisions (in density matrix), 633–41
effect on density matrix, 635
on line width, 641
Tapered mirrors
$ABCD$ matrix for, 563
in cavities, 562–70
Gaussian reflectivity, 556, 559
TEA (Transverse Electric Atmospheric) discharges, 748
TEM_{00}, TEM_{mp} modes, 63, *66*, 73
TE, TM modes
in blackbody cavity, 175
s and p polarization, 22
in slab fiber, 91–95
Temporal growth of photons, 239

Thermal distribution
of rotational levels, 686
of states in YAG, 362
Thermal noise, 714
Thermionic emission, 732
Thermodynamic equilibrium
blackbody radiation, 175–82
see detailed balancing, 770
population ratio, 362
Thevenin equivalent current, 710
Thin lens
ray matrix, 38
transformation of a wave, 55–56, 557
Three, four level lasers, 261
Three photon absorption, 645
Threshold
homojunction laser, 466
for a laser, 208
temperature dependence, 464
Ti:sapphire laser, 391
Townsend ionization coefficient, 753
Transition rates, 179–80
Transmission coefficient (*see* $ABCD$ matrix)
of a DFB structure, 523
of a Fabry–Perot cavity, 149
of a $\lambda/4$ structure, 526
Transverse operator ∇_t, or $\nabla_t \cdot$, 64
Tridiagonal matrix, 571
Triplet system (in a dye), 389
Tunable lasers
cavities for, 395–96
semiconductor lasers, 531

Ultraviolet catastrophe, 177
Uncertainty relation, 13–14
Uniaxial crystal, 17
Unit cell, 40
Unstable cavities
analysis of, 534–39, 555–67
definition, 44
losses in, 560

***V*# of a round fiber**, 95
Variational formula (for β), 103
Vacuum photodiode, 698
VCSEL (vertical cavity surface emitting laser), 482–86
Velocity distribution, 741
VR, vibration-rotational transitions, 684–88
Vibrational transitions
formula for, 687
N_2 data, 408
Virtual levels
Raman effect, 653
role in multiphoton effects, 643
Visibility of fringes, 27
Voight function (line shape), 198
VSWR, 24
Volume of a mode, 447, 471

***w* value**, 416
Wave equation, 11
with active atoms, 592
Wave impedance, 12
Waveguide
dielectric slab waveguide, 87
heterostructure, 503
Wave transformation by a lens, 55, 77, 557
Wiggler
configuration, 419
parameter, 422
Work function, photoelectric, 698

XeF, XeI, XeBr, XeCl, XeI (*see* **excimer lasers**), 414

YAG ($Y_3Al_5O_{12}$), 358
Er^{3+} levels in, 373
Nd^{3+} levels in, 360
physical transition data, 361

Table of Constants, Conversion Factors, and Vector Relations

Common Constants

Name	*Symbol*	*Value*
Speed of light	c	2.9979458×10^8 m/s $\quad \sim 3 \times 10^8$ m/s
Permittivity of free space	ϵ_0	8.8542×10^{-12} F/m
Permeability of free space	μ_0	$4\pi \times 10^{-7}$ H/m
Planck's constant	h	6.626076×10^{-34} J-sec
Electronic charge	e	1.60218×10^{-19} C
Electronic mass	m	9.1094×10^{-31} kg
Proton mass	M_p	1.67262×10^{-27} kg
Boltzmann constant	k	1.38066×10^{-23} J/°K
Avogadro's constant	N_A	6.02213×10^{23} atoms/mole
Loschmidt's number	L_0	2.687×10^{19} molecules/cm^3 at 0° C

Useful Conversion Factors

For Energy: E (joules)

If $h\nu/e = 1$ eV, then $\lambda_0 = 1.23985$ μm, $\bar{\nu} = 1/\lambda_0 = 8065.50$ cm^{-1}
and $E = 23.063$ kcal/mol

For Wavelength: λ (nanometers)

If $\lambda_0 = 1000$ nm, then $\lambda_0 = 1$ μm or 10,000 Å, $\bar{\nu} = 10^4$ cm^{-1}
and $h\nu/e = 1.23985$ eV

For Frequency: ν (Hertz)

If ν is given in Hz, then $\lambda_0 = c/\nu$ and $\bar{\nu}$ (cm^{-1}) $= 1/\lambda_0$ (cm)

Useful Vector Relations

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})$$

$$\nabla \cdot (\nabla \times \mathbf{A}) \equiv 0; \quad \nabla \times (\nabla V) \equiv 0$$

Laplacian of a scalar:
$$\nabla \cdot \nabla V \triangleq \nabla^2 V$$

Curl of a curl:
$$\nabla \times (\nabla \times \mathbf{A}) \equiv \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$$

Laplacian of a vector:
$$\nabla^2 \mathbf{A} \triangleq \nabla(\nabla \cdot \mathbf{A}) - \nabla \times \nabla \times \mathbf{A}$$

$$\nabla \cdot (a\mathbf{A}) \equiv \mathbf{A} \cdot \nabla a + a \nabla \cdot \mathbf{A}$$

Stokes theorem:
$$\iint \nabla \times \mathbf{A} \cdot \mathbf{n}\, dA \equiv \oint \mathbf{A} \cdot d\mathbf{l}$$

Divergence theorem:
$$\iiint \nabla \cdot \mathbf{D}\, dV \equiv \oiint \mathbf{D} \cdot \mathbf{n}\, dA$$

Green's theorems:
$$\int \left\{ \phi \nabla^2 \psi + \nabla \phi \cdot \nabla \psi \right\} dV \equiv \int \phi \nabla \psi \cdot \mathbf{n}\, dA;$$

$$\int \left\{ \phi \nabla^2 \psi - \psi \nabla^2 \phi \right\} dV \equiv \int [\phi \nabla \psi - \psi \nabla \phi] \cdot \mathbf{n}\, dA$$